

NAACL HLT2016 SAN DIEGO

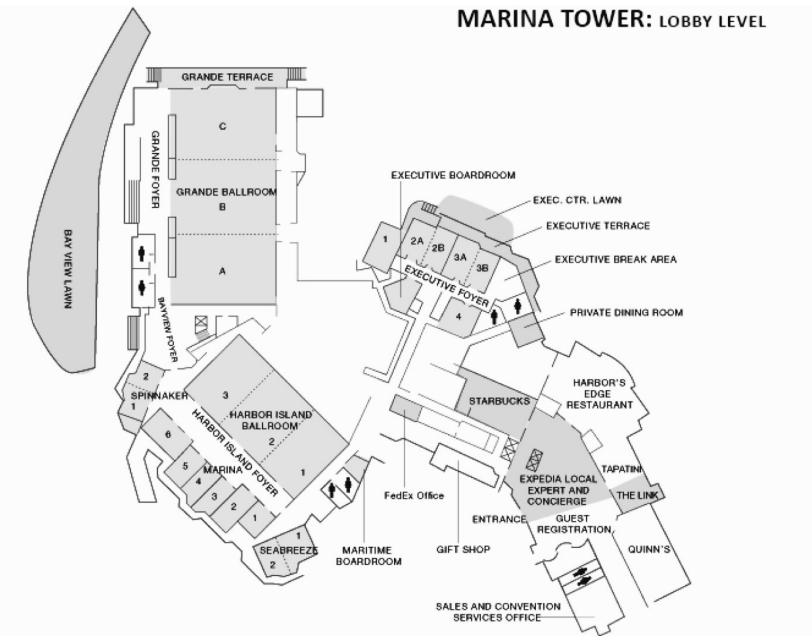
THE 15TH ANNUAL CONFERENCE
OF THE NORTH AMERICAN CHAPTER
OF THE ASSOCIATION FOR
COMPUTATIONAL LINGUISTICS:
HUMAN LANGUAGE TECHNOLOGIES

JUNE 12–17, 2016



CONFERENCE HANDBOOK

MARINA TOWER: LOBBY LEVEL



Contents

Table of Contents	i
1 Conference Information	1
Message from the General Chair	1
Message from the Program Committee Co-Chairs	3
Organizing Committee	6
2 Tutorials: Sunday, June 12	9
T1: English Resource Semantics	10
T2: Multilingual Multimodal Language Processing Using Neural Networks	12
T3: Question Answering with Knowledge Base, Web and Beyond	13
T4: Recent Progress in Deep Learning for NLP	14
TS: Scalable Statistical Relational Learning for NLP	16
T6: Statistical Machine Translation between Related Languages	18
Welcome Reception	20
3 Main Conference: Monday, June 13	21
Keynote Address: Regina Barzilay	22
Session 1	23
Session 2	30
Session 3	36
One Minute Madness	40
Poster Session 1	40
System Demonstrations 1	46
Student Workshop Posters 1	48
4 Main Conference: Tuesday, June 14	51
Session 4	52
Session 5	58
Session 6	64
Session 7	68
One Minute Madness	72

Poster Session 2	72
System Demonstrations 2	78
Student Workshop Posters 2	80
Social Event	83
5 Main Conference: Wednesday, June 15	85
Keynote Address: Ehud Reitter	86
Session 8	87
NAACL Business Meeting	94
Session 9	95
Best paper awards	101
6 Workshops: Thursday–Friday, June 16–17	103
W1: SemEval: 10th International Workshop on Semantic Evaluation	105
W2: Human-Computer Question Answering	119
W3: Fifth Workshop on Computational Linguistics for Literature (clfl2016)	121
W4: Computational Linguistics and Clinical Psychology — From Linguistic Signal to Clinical Reality (CLPsych)	122
W5: 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA 2016)	125
W6: 11th Workshop on Innovative Use of NLP for Building Educational Applications (BEA11)	127
W7: 2nd Workshop on Semantics-Driven Machine Translation (SedMT2016)	130
W8: Coreference Resolution Beyond OntoNotes	131
Optional Excursion	133
W9: Workshop on Computational Approaches to Deception Detection	134
W10: Workshop on Discontinuous Structures in Natural Language Processing	135
W11: The 4th Workshop on EVENTS: Definition, Detection, Coreference, and Representation	136
W12: The 4th Workshop on Metaphor in NLP (Meta4NLP)	137
W13: Workshop on Multilingual and Crosslingual Methods in NLP	138
W14: AKBC: 5th Workshop on Automated Knowledge Base Construction	140
W15: TextGraphs: 10th Workshop on Graph-based Methods for Natural Language Processing	142
7 Anti-harassment policy	143
Author Index	145
Sponsors	167

Conference Information

Message from the General Chair

Welcome to NAACL HLT 2016! This year's conference is held in San Diego, California, where we have assembled an exciting program of computational linguistics research.

The main program features a wide array of topics, and it includes excellent invited talks by Prof. Regina Barzilay and Prof. Ehud Reiter. In addition, we have six tutorials on the day before the main program, plus fifteen workshops on the following two days. Some of these workshops are back for their 10th or 11th incarnation, while others are brand-new. In parallel, we have a live demonstration track, and a Student Research Workshop that showcases work by the junior members of our research community.

This NAACL HLT meeting takes place only through the hard work of many people who deserve our gratitude.

Thanks to Priscilla Rasmussen for making local arrangements, handling registration, setting up social events, writing visa invitation letters, and solving a myriad of issues. Priscilla, your experience is a great asset to any conference!

The NAACL HLT organizing committee took all the steps to bring you a great conference. Many thanks to Ani Nenkova and Owen Rambow (Program Co-chairs), Mohit Bansal and Alexander M. Rush (Tutorial Co-chairs), Radu Soricut and Adrià de Gispert (Workshop Co-chairs), Jacob Andreas, Eunsol Choi, and Angeliki Lazaridou (Student Research Workshop Co-Chairs) and their faculty advisors Jacob Eisenstein and Nianwen Xue, Aliya Deri (Student Volunteer Coordinator), Julie Medero (Local Sponsorship Chair), Mark Finlayson, Sravana Reddy, and John DeNero (Demonstration Co-chairs), Adam Lopez and Margaret Mitchell (Publications Co-chairs), Jason Riesa (Website Chair), Wei Xu (Publicity Chair), and Jonathan May (Social Media Chair).

Thanks also to the NAACL Board for providing excellent advice, and thanks to previous chairs for their suggestions and timelines.

Sponsors of NAACL HLT 2016 include Baidu and Google (Platinum Sponsors), Amazon, Bloomberg, eBay, Microsoft Research, and UnitedHealth Group (Gold Sponsors), Huawei (Silver Sponsors), Civis Analytics, Facebook, @newsela, and Nuance (Bronze Sponsors), and the University of Washington (Supporter). Thanks for your extremely valuable contributions!

Message from the General Chair

Finally, thanks to the scientists, engineers, authors, and attendees who come to share and learn at this leading venue for computational linguistics research!

Kevin Knight
Information Sciences Institute, University of Southern California
NAACL HLT 2016 General Chair

Message from the Program Committee Co-Chairs

Welcome to San Diego for the 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies!

The conference has grown remarkably in the past five years: we had 698 submissions this year, despite our deadline right after the end-of-the-year holidays. As we worked on organizing the conference program, we made many changes to reflect the growth of the NAACL community, the increasing diversity of topics covered by the field, and the acceleration of the pace of the publication cycle.

We had a record short time between paper submission and author notification—less than two months. We settled on such compressed timeline in order to avoid spreading the reviewing period over the winter holidays, to ensure that papers spend only a short time under submission, and to coordinate submission deadlines with ACL. Our incredible team of area chairs and reviewers ensured that the planned schedule went smoothly.

As the computational linguistics field has expanded, it has become increasingly difficult to recruit a sufficient number of knowledgeable reviewers. We decided to reach out to the largest possible pool of computational linguists and provide convenient ways for the area chairs to control which reviewers they end up working with: we invited all researchers actively working in the area of computational linguistics/language processing to review for the conference. We defined “active researchers” to be those who have published at least five papers in the last ten years in the ACL, NAACL, EMNLP, EACL or COLING conferences. In order to be inclusive of the amazing young researchers who became active in the field only more recently, we also included everyone who had published at least three papers in the same venues for the last five years. This yielded a list of over 1,400 researchers that we invited to serve as reviewers for the conference. Of these, 685 agreed and participated in the review process. This is another record for NAACL HLT 2016, no previous NAACL has had such a large program committee. Among these, the area chairs recognized 120 as best reviewers.

Working with the reviewers were the 42 area chairs. We asked the area chairs to work in pairs, so they can have a back-up in case other obligations need their attention during the review period and to ensure that all decisions about reviewer assignment and paper recommendation are discussed in detail. All area chairs and reviewers submitted a list of keywords that describe their area of expertise (the full list appears in the conference call for papers). The area chairs were paired based on the keyword overlap.

To match reviewers to area chairs, we used a bidding system. For bidding, each area received a list of the 140 reviewers with best matching keyword profiles. If the area chairs did not know the work of a potential reviewer on their bidding list, they looked him or her up on DBLP or Google Scholar before making their final bid. Areas were assigned only reviewers for which the area chairs bid positively. Area chairs were free as usual to recruit additional reviewers they wished to work with.

Submissions were assigned to areas by taking into account the match between the paper keywords and the area chair keywords. Areas were capped at 40 submissions maximum (long and short combined). As in the past, reviewers bid on papers they wanted to review. 69% of the reviews were written by reviewers who had bid indicating that they want to review the paper; 29% of the reviews were written by reviewers who had bid indicating they are ok with reviewing the paper. The remaining 2% of reviews were written by reviewers who did not bid on the paper but were asked by an area chair to review it. Three reviewers were assigned a paper that they did not want to review according to their bid. The average reviewer load was 3 papers, which included a mix of long and short submissions. Only 43 reviewers had more than four papers to review.

Area chairs wrote meta-reviews, for use only by us, justifying their accept/reject recommendation. In making difficult decisions, we drew on these meta-reviews, the reviews themselves, the discussion among the reviewers, and the author response to the initial reviews.

We are happy with our changes to the review process: area chairs had control over the re-

viewers they worked with, reviewers were assigned papers they wanted to review and the overall reviewing load was low. Needless to say, there is room for further improvements. The reviewing process is crucial to the quality of this conference; only if the community has confidence in the quality of the reviewing process will this conference continue to be a leading conference in our field. Our goal has been to make sure that every single submission receives a complete and fair review and decision, and to make sure that the authors of every single submission understand why their paper was accepted or declined for the conference. We would like to thank our 685 reviewers, and we would especially like to thank our 42 area chairs, who were patient in allowing us to pursue some of the innovative aspects of this year's reviewing cycle.

Eighteen of the 698 initial submissions were withdrawn by the authors or rejected without review because of formatting violations. A total of 396 long and 284 short papers underwent review; 100 long and 82 short papers were accepted, for an acceptance rate of 25% and 29% respectively. In addition, ten TACL papers will be presented at the conference.

This year we decided to have shorter slots for oral presentations, in order to have more of the accepted papers presented as talks. In the program, long papers are allotted 20-minute slots (15 min presentation + 5 min questions). Short papers are allotted 10-minute slots (6 min presentation + 4 min questions).

The best paper award committee consisted of NAACL general and program chairs from the last three years. Not all past chairs could participate in the selection. The final best paper committee included Joyce Chai, Katrin Kirchhoff, Rada Mihalcea, Kristina Toutanova, Lucy Vanderwende and Hua Wu. They selected two best long papers and one best short paper, along with two runner-ups in each category.

Best Short Paper

Improving sentence compression by learning to predict gaze
Sigrid Klerke, Yoav Goldberg and Anders Søgaard

Short Paper, Runners Up

Patterns of Wisdom: Discourse-Level Style in Multi-Sentence Quotations
Kyle Boooten and Marti A. Hearst

A Joint Model of Orthography and Morphological Segmentation
Ryan Cotterell, Tim Vieira and Hinrich Schütze

Best Long Papers

Feuding Families and Former Friends: Unsupervised Learning for Dynamic Fictional Relationships
Mohit Iyyer, Anupam Guha, Snigdha Chaturvedi, Jordan Boyd-Graber and Hal Daumé III

Learning to Compose Neural Networks for Question Answering
Jacob Andreas, Marcus Rohrbach, Trevor Darrell and Dan Klein

Long Paper, Runners Up

Multi-way, Multilingual Neural Machine Translation with a Shared Attention Mechanism
Orhan Firat, Kyunghyun Cho and Yoshua Bengio

Black Holes and White Rabbits: Metaphor Identification with Visual Features
Ekaterina Shutova, Douwe Kiela and Jean Maillard

The conference program includes two inspiring invited talks by Regina Barzilay and Ehud Reiter. Both push the boundaries of the field, discussing the potential for real-world impact of language technologies.

Finally we would like to thank all other people who supported us in the past year in our work for NAACL HLT 2016. Last year's program chairs, Anoop Sarkar and Joyce Chai shared their valuable advice and promptly answered the many questions we had throughout the process. The NAACL board chair for 2015 (Hal Daumé III) and 2016 (Emily Bender) were our effective link with the NAACL board. The conference general chair, Kevin Knight, was always available to us when we needed to consult about decisions we were making. The conference business manager, Priscilla Rasmussen, gave us details about the venue and coordinated with us at the final stages of making the conference schedule. The ACL treasurer, Greame Hirst, answered questions about

the venue. The conference webmaster, Jason Riesa, put content on the conference webpage as soon as we made it available to him. The publication chairs, Meg Mitchell and Adam Lopez, answered all lingering author questions about formatting for submission and final versions. Many talks to all of them!

We look forward to an exciting conference!

NAACL HLT 2016 Program Co-Chairs
Ani Nenkova, University of Pennsylvania
Owen Rambow, Columbia University

Organizing Committee

General Chair

Kevin Knight, USC Information Sciences Institute

Program Co-chairs

Ani Nenkova, University of Pennsylvania

Owen Rambow, Columbia University

Workshop Co-chairs

Radu Soricut, Google

Adrià de Gispert, SDL

Local Sponsorship Chair

Julie Medero, Harvey Mudd College

Tutorial Co-chairs

Mohit Bansal, TTI Chicago

Alexander M. Rush, Harvard University

Demonstration Co-chairs

Mark Finlayson, Florida International University

Sravana Reddy, Wellesley College

John DeNero, University of California, Berkeley

Publication Co-chairs

Adam Lopez, University of Edinburgh

Margaret Mitchell, Microsoft Research

Local Arrangements Chair

Priscilla Rasmussen, ACL Business Manager

Student Research Workshop Co-chairs*Student Co-chairs*

Jacob Andreas, University of California, Berkeley

Eunsol Choi, University of Washington

Angeliki Lazaridou, University of Trento

Faculty Advisors

Jacob Eisenstein, Georgia Tech

Nianwen Xue, Brandeis University

Publicity Chair

Wei Xu, University of Pennsylvania

Social Media Chair

Jonathan May, USC Information Sciences Institute

Website Chair

Jason Riesa, Google

Student Volunteer Coordinator

Aliya Deri, USC Information Sciences Institute

Program Committee

Program Co-chairs

Ani Nenkova, University of Pennsylvania
Owen Rambow, Columbia University

Area Chairs

Mohit Bansal, TTI-Chicago
Regina Barzilay, MIT
Eduardo Blanco, University of North Texas
Asli Celikyilmaz, Microsoft
Cristian Danescu-Niculescu-Mizil, Cornell University
Markus Dreyer, Amazon
Chris Dyer, Carnegie Mellon University
Jacob Eisenstein, Georgia Institute of Technology
Micha Elsner, The Ohio State University
Eric Fosler-Lussier, The Ohio State University
Alexander Fraser, University of Munich
Michel Galley, Microsoft Research
Kevin Gimpel, Toyota Technological Institute at Chicago
Dilek Hakkani-Tür, Microsoft Research
Helen Hastie, Heriot-Watt University
Yulan He, Aston University
Dirk Hovy, University of Copenhagen
Heng Ji, Rensselaer Polytechnic Institute
Jing Jiang, Singapore Management University
Annie Louis, University of Essex
Chin-Yew Lin, Microsoft Research
Daniel Marcu, Information Sciences Institute, University of Southern California
Margaret Mitchell, Microsoft Research
Alessandro Moschitti, Qatar Computing Research Institute, HBKU
Hwee Tou Ng, National University of Singapore
Viet-An Nguyen, Facebook
Mari Ostendorf, University of Washington
Marius Pasca, Google
Slav Petrov, Google
Dan Roth, University of Illinois
Alexander Rush, Harvard University
Kenji Sagae, KITT.AI
Giorgio Satta, University of Padua
Hinrich Schütze, LMU Munich
William Schuler, the Ohio State University
Mihai Surdeanu, University of Arizona
Kristina Toutanova, Microsoft Research
Byron Wallace, University of Texas at Austin
Xiaojun Wan, Peking University
Furu Wei, Microsoft Research
Dekai Wu, Hong Kong University of Science and Technology
Fei Xia, University of Washington

2

Tutorials: Sunday, June 12

Overview

7:30–6:00	Registration	<i>Grande Foyer</i>
9:00–12:30	Morning Tutorials	
	English Resource Semantics <i>Dan Flickinger, Emily M. Bender, and Woodley Packard</i>	<i>Executive 3AB</i>
	Multilingual Multimodal Language Processing Using Neural Networks <i>Spinnaker</i> <i>Mitesh M. Khapra and Sarath Chandar</i>	
	Question Answering with Knowledge Base, Web and Beyond <i>Wen-tau Yih and Hao Ma</i>	<i>Marina 3</i>
10:30–11:00	Coffee break	<i>Grande Foyer</i>
12:30–2:00	Lunch break	
2:00–5:30	Afternoon Tutorials	
	Recent Progress in Deep Learning for NLP <i>Zhengdong Lu and Hang Li</i>	<i>Spinnaker</i>
	Scalable Statistical Relational Learning for NLP <i>William Yang Wang and William Cohen</i>	<i>Marina 3</i>
	Statistical Machine Translation between Related Languages <i>Pushpak Bhattacharyya, Mitesh M. Khapra, and Anoop Kunchukuttan</i>	<i>Executive 3AB</i>
3:30–4:00	Coffee break	<i>Grande Foyer</i>
6:00–9:00	Welcome Reception	<i>Grande Foyer and Terrace</i>

Tutorial 1

English Resource Semantics

Dan Flickinger, Emily M. Bender, and Woodley Packard

Sunday, June 12, 2016, 9:00–12:30pm

Executive 3AB

Recent years have seen a dramatic increase in interest in semantically-informed natural language processing, including parsing into semantic representations, grounded language processing that connects linguistic structures to world representations, proposals to integrate compositional and distributional approaches to semantics, and approaches to semantically-sensitive tasks including sentiment analysis, summarization, generation, machine translation, and information extraction which take into account linguistic structure beyond n-grams. The semantic inputs to this work include a wide range of representations, from word embeddings, to syntactic dependencies used as a proxy for semantic dependencies, to sentence-level semantic representations either partial (e.g. semantic role labels) or fully articulated.

The purpose of this tutorial is to make accessible an important resource in this space, namely the semantic representations produced by the English Resource Grammar (ERG; Flickinger 2000, 2011). The ERG is a broad-coverage, linguistically motivated precision grammar for English, associating richly detailed semantic representations with input sentences. These representations, dubbed English Resource Semantics or ERS, are in the formalism of Minimal Recursion Semantics (MRS; Copstake et al 2005). They include not only semantic roles, but also information about the scope of quantifiers and scopal operators including negation, as well as semantic representations of linguistically complex phenomena such as time and date expressions, conditionals, and comparatives (Flickinger et al. 2014). ERS can be expressed in various ways, including a logic-based syntax using predicates and arguments, dependency graphs and dependency triples. In addition, the representations can be obtained either from existing manually produced annotations over texts from a variety of genres (the Redwoods Treebank, Oepen et al 2004) and DeepBank (Wall Street Journal corpus: Flickinger et al 2012) or by processing new text with the ERG and its associated parsing and parse selection algorithms.

With high parsing accuracy with rich semantic representations, English Resource Semantics is a valuable source of information for many semantically-sensitive NLP tasks. ERS-based systems have achieved state-of-the-art results in various tasks, including the identification of speculative or negated event mentions in biomedical text (MacKinlay et al 2011), question generation (Yao et al 2012), detecting the scope of negation (Packard et al 2014), relating natural language to robot control language (Packard 2014), and recognizing textual entailment (PETE task; Lien & Kouylekov 2015). ERS representations have also been beneficial in semantic transfer-based MT (Oepen et al 2007, Bond et al 2011), ontology acquisition (Herbelot 2006), extraction of glossary sentences (Reiplinger et al 2012), sentiment analysis (Kramer & Gordon 2014), and the ACL Anthology Searchbench (Schäfer et al 2011).

The goal of this tutorial is to make this resource more accessible to the ACL community. Specifically, we take as our learning goals that tutorial participants will learn how to: (1) set up the ERG-based parsing stack, including preprocessing; (2) access ERG Redwoods/DeepBank treebanks in the various export formats; and (3) interpret ERS representations.

Dan Flickinger is a Senior Research Associate at CSLI at Stanford, and manager of the Linguistic Grammars Online (LinGO) Laboratory, where he is the principal developer of the LinGO English Resource Grammar (ERG), a precise broad-coverage implementation of HPSG. His primary research interests are in wide-coverage grammar engineering for both parsing and generation, lexical representation, the syntax-semantics interface, methodology for evaluation of semantically precise grammars, and practical applications of 'deep' processing. Applied and industrial experience includes co-founding the software company YY Technologies, which from 2000-2002 sold automated consumer email response technology incorporating the ERG; and since 2009 developing online educational software using the ERG for teaching English writing skills, first as part of the Education Program for Gifted Youth (EPGY) at Stanford, and for the past three years as a senior researcher at the EPGY spin-off company Redbird Advanced Learning, based in Oakland, California.

Emily M. Bender is a Professor in the Department of Linguistics and Associate Professor in the Department of Computer Science & Engineering at the University of Washington. Her primary research interests lie in multilingual grammar engineering, semantic representations, and the incorporation of linguistic knowledge, especially from semantics and linguistic typology, in computational linguistics. She is the primary developer of the Grammar Matrix grammar customization system, which is developed in the context of the DELPH-IN Consortium (Deep Linguistic Processing with HPSG Initiative). Her book, *Linguistic Fundamentals for Natural Language Processing: 100 Essentials from Morphology and Syntax* grew out of a NAACL 2012 tutorial on that topic.

Woodley Packard is a student at the University of Washington, pursuing a PhD in Computational Linguistics with Emily M. Bender. His research interests include efficient algorithms for grammar-based parsing, generation, annotation, and learning; robustness mechanisms for precision grammars; methodologies for contextually-informed disambiguation; and applications of semantic representations. He wrote and maintains the ACE parser/generator and the FFTB annotation tool. Recent work includes designing and building the top-performing entry for SemEval 2014 Task 6 on interpreting natural language commands to robots.

Tutorial 2

Multilingual Multimodal Language Processing Using Neural Networks

Mitesh M. Khapra and Sarath Chandar

Sunday, June 12, 2016, 9:00–12:30pm

Spinnaker

We live in an increasingly multilingual multimodal world where it is common to find multiple views of the same entity across modalities and languages. For example, news articles which get published in multiple languages are essentially different views of the same entity. Similarly, video, audio and multilingual subtitles are multiple views of the same movie clip. Given the proliferation of such multilingual multimodal content it is no longer sufficient to process a single modality or language at a time. Specifically, there is an increasing demand for allowing transfer, conversion and access across such multiple views of the data. For example, users want to translate/convert news articles to their native language, automatically caption their travel photos and even ask natural language questions over videos and images. This has led to a lot of excitement around this interdisciplinary research area which requires ideas from Machine Learning, Natural Language Processing, Speech and Computer Vision among other fields.

In this tutorial we focus on neural network based models for addressing various problems in this space. We will first introduce the participants to some of the basic concepts and building blocks that such approaches rely on. We will then describe some of these approaches in detail. There are two important parts to the tutorial. In the first part, we will talk about approaches which aim to learn a common representation for entities across languages and modalities thereby enabling cross lingual and cross modal access and transfer. In the second part we will talk about multilingual multimodal generation. For example, we will discuss neural network based approaches which aim at (i) generating translations in multiple languages, (ii) generating images given a natural language description and (iii) generating captions in multiple languages.

Mitesh M Khapra obtained his Ph.D. from the Indian Institute of Technology, Bombay in the area of Natural Language Processing with a focus on reusing resources for multilingual computation. His areas of interest include Statistical Machine Translation, Text Analytics, Crowdsourcing, Argument Mining and Deep Learning. He is currently working as a Researcher at IBM Research India where he is focusing on mining arguments from large unstructured text. He is also interested in learning common representations across languages and modalities with the view of enabling cross language and cross modal access. He has co-authored papers in top NLP and ML conferences such as ACL, NAACL, EMNLP, AAAI and NIPS.

Sarath Chandar is currently a PhD student in University of Montreal where he works with Yoshua Bengio and Hugo Larochelle on Deep Learning for complex NLP tasks like question answering and dialog systems. His research interests includes Machine Learning, Natural Language Processing, Deep Learning, and Reinforcement Learning. Before joining University of Montreal, he was a Blue Scholar in IBM Research India for a year.

Tutorial 3

Question Answering with Knowledge Base, Web and Beyond

Wen-tau Yih and Hao Ma

Sunday, June 12, 2016, 9:00–12:30pm

Marina 3

Developing a Question Answering (QA) system to automatically answer natural-language questions has been a long-standing research problem since the dawn of AI, for its clear practical and scientific value. For instance, whether a system can answer questions correctly is a natural way to evaluate a machine's understanding of a domain. Providing succinct and precise answers to informational queries is also the direction pursued by the next generation of search engines that aim to incorporate more "semantics", as well as the basic function in digital assistants like Siri and Cortana.

In this tutorial, we aim to give the audience a coherent overview of the research of question answering. We will first introduce a variety of QA problems proposed by pioneer researchers and briefly describe the early efforts. By contrasting with the current research trend in this domain, the audience can easily comprehend what technical problems remain challenging and what the main breakthroughs and opportunities are during the past half century. For the rest of the tutorial, we select three categories of the QA problems that have recently attracted a great deal of attention in the research community, and will present the tasks with the latest technical survey.

The first two categories regard answering factoid questions, where the main difference of the problem settings is the information source used for extracting answers. QA with knowledge base aims to answer natural language questions using real-world facts stored in an existing, large-scale database. The representative approach for this task is to develop a semantic parser (of questions), which will be the main focus. Other approaches like text matching in the embedding space and those driven by information extraction will also be discussed. The other category, QA with the Web, targets answering questions mainly from the facts extracted from general text corpora derived from the Web. In addition to the common components and techniques used in this setting, including passage retrieval, entity recognition and question analysis, we will also introduce latest work on how to leverage and incorporate additional structured and semi-structured data to improve the performance. The third category of the QA problems that we will highlight is the non-factoid questions. Due to its broad coverage, we will briefly cover three exemplary topics: story comprehension, reasoning questions and paragraph QA. The tutorial will conclude by summarizing a whole area of exciting and dynamic research that is worthy of more detailed investigation for many years to come.

Scott Wen-tau Yih

Hao Ma

Tutorial 4

Recent Progress in Deep Learning for NLP

Zhengdong Lu and Hang Li

Sunday, June 12, 2016, 2:00–5:30pm

Spinnaker

Neural network-based methods have been viewed as one of the major driving force in the recent development of natural language processing (NLP). We all have witnessed with great excitement how this subfield advances: new ideas emerge at an unprecedented speed and old ideas resurge in unexpected ways. In a nutshell, there are two major trends:

- Ideas and techniques from other fields of machine learning and artificial intelligence (A.I.) have increasing impact on neural network-based NLP methods.
- With end-to-end models taking on more complex tasks, the design of architecture and mechanisms often needs more domain knowledge from linguists and other domain experts.

Both trends are important to researchers in the computational linguistics community. Fundamental ideas like external memory or reinforcement learning, although introduced to NLP only recently, have quickly lead to significant improvement on tasks like natural language generation and question answering. On the other hand, with complicated neural systems with many cooperating components, it calls for linguistic knowledge in designing the right mechanism, architecture, and sometimes training setting.

As a simple example, the introducing of automatic alignment in neural machine translation, has quickly led to the state-of-the-art performance in machine translation and triggered a large body of sequence-to-sequence models. It is therefore important to get the researchers in computational linguistics community acquainted with the recent progress in deep learning for NLP.

We will focus on the work and ideas strongly related to the core of natural language and yet not so familiar to the majority of the community, which can be roughly categorized into: 1) the differentiable data-structures, and 2) the learning paradigms for NLP.

Differentiable data-structures, starting with the memory equipped with continuous operations in Neural Turing Machine, have been the foundation of deep models with sophisticated operations. Some members of it, such as Memory Network, have become famous on tasks like question answering and machine translation, while other development in this direction, including those with clear and important application in NLP, are relatively new to this community.

Deep learning, with its promise on end-to-end learning, not only enables the training of complex NLP models from scratch, but also extends the training setting to include remote and indirect supervision. We will introduce not only the end-to-end learning in its general notion, but also newly emerged topics in formulating learning objective, taming the non-differentiable operations, and designing the learning system for getting supervision signal from real world.

Zhengdong Lu is a senior researcher at Noah's Ark Lab, Huawei Technologies. His research interests are neural network-based methods for natural language processing, including dialogue, machine translation, semantic parsing, and reasoning. Previously he was an associate researcher at Microsoft Research Asia and a postdoctoral researcher at University of Texas at Austin, after receiving his Ph.D.degree from Oregon Health and Science University in 2008 in computer science. He has published over 30 papers in prestigious journals and conferences, including NIPS, ICML, ACL, KDD, IJCAI and AAAI, including over 10 recent papers on deep learning methods for NLP and AI.

Hang Li is director of the Noah's Ark Lab of Huawei Technologies, adjunct professors of Peking University and Nanjing University. His research areas include information retrieval, natural language processing, statistical machine learning, and data mining. He is ACM Distinguished Scientist. Hang graduated from Kyoto University in 1988 and earned his PhD from the University of Tokyo in 1998. He worked at the NEC lab as researcher during 1991 and 2001, and Microsoft Research Asia as senior researcher and research manager during 2001 and 2012. He joined Huawei Technologies in 2012. Hang has published three technical books, and more than 100 technical papers at top international conferences including SIGIR, WWW, WSDM, ACL, EMNLP, ICML, NIPS, SIGKDD and top international journals including CL, NLE, JMLR, TOIS, IRJ, IPM, TKDE, TWEB, TIST. He and his colleague's papers received the SIGKDD-08 best application paper award, the SIGIR-08 best student paper award, the ACL-12 best student paper award. Hang worked on the development of several products such as Microsoft SQL Server 2005, Office 2007, Office 2010, Live Search 2008, Bing 2009, Bing 2010, Huawei AppStore, Huawei Phones. He has 40 granted US patents. Hang is also very active in the research communities and has served or is serving top international conferences as PC chair, Senior PC member, or PC member, including SIGIR, WWW, WSDM, ACL, EMNLP, NIPS, SIGKDD, ICDM, ACML, and top international journals as associate editor or editorial board member, including CL, IRJ, TIST, JASIST, JCST. Hang Li gave a number of tutorials on various topics about machine learning for natural language processing and information retrieval, including a tutorial on learning to rank at ACL 2009.

Tutorial 5

Scalable Statistical Relational Learning for NLP

William Yang Wang and William Cohen

Sunday, June 12, 2016, 2:00–5:30pm

Marina 3

Statistical Relational Learning (SRL) is an interdisciplinary research area that combines first-order logic and machine learning methods for probabilistic inference. Although many Natural Language Processing (NLP) tasks (including text classification, semantic parsing, information extraction, coreference resolution, and sentiment analysis) can be formulated as inference in a first-order logic, most probabilistic first-order logics are not efficient enough to be used for large-scale versions of these tasks. In this tutorial, we provide a gentle introduction to the theoretical foundation of probabilistic logics, as well as their applications in NLP. We describe recent advances in designing scalable probabilistic logics, with a special focus on ProPPR. Finally, we provide a hands-on demo about scalable probabilistic logic programming for solving practical NLP problems.

Part 1: Foundations and Applications of Probabilistic First-Order Logic We will provide a brief review of some first-order learning systems that have been developed in the past: Markov Logic Networks (Richardson and Domingos, 2006), Stochastic Logic Programs (Muggleton, 1996). In this part, we introduce the semantics of the above languages with their inference (and learning) approaches. We analyze and discuss the core ideas behind of such language. We show various applications of probabilistic logics in NLP.

Part 2: Scalable Probabilistic Logics: A Case Study of ProPPR. We will focus on the efficiency issue, and introduce recent advances of scalable probabilistic logics, including lifted inference techniques (Van den Broeck and Suciu, 2014) and probabilistic soft logic (Bach et al., 2015). In particular, we will take CMU’s ProPPR (Wang et al., 2013) as a case study. We describe the main contributions of ProPPR: including its approximate personalized PageRank inference scheme, parallel stochastic gradient descent learning method, and its flexibility in theory engineering. We then introduce the structure learning methods in ProPPR (Wang et al., CIKM 2014), including a structured regularization method as an alternative to predicate invention (Wang et al., IJCAI 2015). We will also cover our latest attempt of learning first-order logic formula embeddings, and discuss its relationship to (and possible connections between) even newer approaches to modeling knowledge bases, relationships, and inference using deep learning methods. To conclude this part, we show an interesting application of ProPPR (Wang et al., ACL-IJCNLP 2015): a joint information extraction and knowledge reasoning engine.

Demos and Practical Applications. We switch from the theoretical presentations to an interactive demonstration session: we aim at providing a hands-on lab session to transfer the theories of scalable probabilistic logics into practices. More specifically, we will provide a demo of several applications on synthetic and real-world datasets. Participants are encouraged to check out our repository on Github (<https://github.com/TeamCohen/ProPPR>) and bring laptops to the tutorial. The list of demo examples to be considered are text categorization, entity resolution, knowledge base completion (Wang et al., MLJ 2015), dependency parsing (Wang et al., EMNLP 2014), structure learning, and joint information extraction & reasoning.

William Yang Wang is a final-year PhD student at the School of Computer Science, Carnegie Mellon University. He works with William Cohen on designing scalable learning and inference algorithms for statistical relational learning, knowledge reasoning, and information extraction. He has published about 30 papers at leading conferences and journals including ACL, EMNLP, and NAACL. He has received best paper awards (or nominations) at ASRU, CIKM, and EMNLP, a best reviewer award at NAACL 2015, the Richard King Mellon Presidential Fellowship in 2011, and he is a Facebook Fellowship finalist. He is an alumnus of Columbia University, and a former research scientist intern at Yahoo! Labs, Microsoft Research Redmond, and University of Southern California.

William W. Cohen is a professor of machine learning at Carnegie Mellon University. Dr. Cohen's research interests include information integration and machine learning, particularly information extraction, text categorization and learning from large datasets. He has a long-standing interest in statistical relational learning and learning models, or learning from data, that display non-trivial structure. He holds seven patents related to learning, discovery, information retrieval, and data integration, and is the author of more than 200 publications. He was a past president of International Machine Learning Society. He is a AAAI fellow, and was a winner of SIGMOD Test of Time Award and SIGIR Test of Time Award.

Tutorial 6

Statistical Machine Translation between Related Languages

Pushpak Bhattacharyya, Mitesh M. Khapra, and Anoop Kunchukuttan

Sunday, June 12, 2016, 2:00–5:30pm

Executive 3AB

Language-independent Statistical Machine Translation (SMT) has proven to be very challenging. The diversity of languages makes high accuracy difficult and requires substantial parallel corpus as well as linguistic resources (parsers, morph analyzers, etc.). An interesting observation is that a large chunk of machine translation (MT) requirements involve related languages. They are either : (i) between related languages, or (ii) between a lingua franca (like English) and a set of related languages. For instance, India, the European Union and South-East Asia have such translation requirements due to government, business and socio-cultural communication needs.

Related languages share a lot of linguistic features and the divergences among them are at a lower level of the NLP pipeline. The objective of the tutorial is to discuss how the relatedness among languages can be leveraged to bridge this language divergence thereby achieving some/all of these goals: (i) improving translation quality, (ii) achieving better generalization, (iii) sharing linguistic resources, and (iv) reducing resource requirements.

We will look at the existing research in SMT from the perspective of related languages, with the goal to build a toolbox of methods that are useful for translation between related languages. This tutorial would be relevant to Machine Translation researchers and developers, especially those interested in translation between low-resource languages which have resource-rich related languages. It will also be relevant for researchers interested in multilingual computation.

We start with a motivation for looking at the SMT problem from the perspective of related languages. We introduce notions of language relatedness useful for MT. We explore how lexical, morphological and syntactic similarity among related languages can help MT. Lexical similarity will receive special attention since related languages share a significant vocabulary in terms of cognates, loanwords, etc.

Then, we look beyond bilingual MT and present how pivot-based and multi-source methods incorporate knowledge from multiple languages, and handle language pairs lacking parallel corpora. We present some studies concerning the implications of languages relatedness to pivot-based SMT, and ways of handling language divergence in the pivot-based SMT scenario. Recent advances in deep learning have made it possible to train multi-language neural MT systems, which we think would be relevant to training between related languages.

We will summarize the tutorial by pointing out how the toolbox addresses the following goals we set out: (i) improving translation quality, (ii) achieving better generalization, (iii) sharing linguistic resources, and (iv) reducing resource requirements. We will conclude by emphasizing how the toolbox can be used to design translation system architectures customized to a set of related languages.

Time permitting, we will briefly describe a toolkit for Indian language NLP, which can be used to leverage similarities between Indian languages (http://anoopkunchukuttan.github.io/indic_nlp_library).

Pushpak Bhattacharyya is Vijay and Sita Vashee Chair Professor in the Department of Computer Science and Engineering at the Indian Institute of Technology Bombay (IITB) where he heads the Center for Natural Language Processing. He is also the Director of Indian Institute of Technology Patna.

Dr. Bhattacharyya obtained his Ph.D from IIT Bombay. His areas of interest cover a broad spectrum of problems in Natural Language Processing like machine translation, cross-lingual search, sentiment analysis - specially with reference to Indian languages.

Dr. Bhattacharyya has published extensively in top quality conferences and journals (about 200). He has also written a textbook on machine translation. He has advised 12 PhDs in NLP and ML, and is currently supervising 10 PhD students. He has also advised close to 125 masters students and above 40 bachelor degree students for their research work.

Mitesh Khapra obtained his Ph.D. from the Indian Institute of Technology Bombay in the area of Natural Language Processing with a focus on reusing resources for multilingual computation. His areas of interest include Statistical Machine Translation, Text Analytics, Crowdsourcing, Argument Mining and Deep Learning. He is currently working as a researcher at IBM Research India where he is focusing on mining arguments from large unstructured text. He has co-authored papers in NLP and ML conferences such as ACL, NAACL, EMNLP, AAAI and NIPS.

Anoop Kunchukuttan is a senior Ph.D student at the Indian Institute of Technology Bombay. He is advised by Prof. Pushpak Bhattacharyya on his research work involving machine translation and transliteration among related languages. He has also investigated other NLP problems - multiword extraction, grammar correction, crowdsourcing and information extraction. He has co-authored papers in NLP conferences such as ACL, NAACL, CONLL, LREC, ICON. He has worked in the software industry for about 5 years, during which he led the development of large scale systems for information extraction and retrieval over medical text. He completed his M.Tech in Computer Science & Engineering from IIT Bombay.

Welcome Reception

Sunday, June 12, 2016, 6:00pm – 9:00pm

Sheraton San Diego Hotel & Marina (conference venue)
Grande Foyer and Terrace

Catch up with your colleagues at the **Welcome Reception!** It will be held immediately following the Tutorials on Sunday, June 12 at 6:00pm in the Grande Foyer and Terrace of the Sheraton San Diego Hotel & Marina (the conference venue). Refreshments and a light dinner will be provided, and a cash bar will be available.

Main Conference: Monday, June 13

Overview

7:30–8:45	Breakfast	Pavilion	
9:00–9:15	Welcome (Kevin Knight, Ani Nenkova, Owen Rambow)	Grande Ballroom	
9:15–10:30	Invited talk: "How can NLP help cure cancer?" (Regina Barzilay)	Grande Ballroom	
10:30–11:00	Coffee break	Grande Foyer	
	Session 1		
11:00–12:30	Machine translation <i>Grande Ballroom A</i>	Summarization <i>Grande Ballroom B</i>	Dialog <i>Grande Ballroom C</i>
12:30–2:00	Lunch break		
	Session 2		
2:00–3:30	Language and Vision <i>Grande Ballroom A</i>	Parsing <i>Grande Ballroom B</i>	Named Entity Recognition <i>Grande Ballroom C</i>
3:30–4:00	Coffee break	Grande Foyer	
	Session 3		
4:00–5:00	Event detection <i>Grande Ballroom A</i>	Language Models <i>Grande Ballroom B</i>	Non Literal Language <i>Grande Ballroom C</i>
5:00–5:15	Break	Grande Foyer	
5:15–6:00	One Minute Madness	Grande Ballroom	
6:00–8:00	Poster Session 1	Pavilion	
6:00–8:00	System Demonstrations 1	Pavilion	
6:00–8:00	Student Workshop Posters 1	Pavilion	

Keynote Address: Regina Barzilay

How can NLP help cure cancer?

Monday, June 13, 2016, 9:15–10:30

Grande Ballroom



Abstract

Cancer inflicts a heavy toll on our society. One out of seven women will be diagnosed with breast cancer during their lifetime, a fraction of them contributing to about 450,000 deaths annually worldwide. Despite billions of dollars invested in cancer research, our understanding of the disease, treatment, and prevention is still limited.

Majority of cancer research today takes place in biology and medicine. Computer science plays a minor supporting role in this process if at all. In this talk, I hope to convince you that NLP as a field has a chance to play a significant role in this battle. Indeed, free-form text remains the primary means by which physicians record their observations and clinical findings. Unfortunately, this rich source of textual information is severely underutilized by predictive models in oncology. Current models rely primarily only on structured data.

In the first part of my talk, I will describe a number of tasks where NLP-based models can make a difference in clinical practice. For example, these include improving models of disease progression, preventing over-treatment, and narrowing down to the cure. This part of the talk draws on active collaborations with oncologists from Massachusetts General Hospital (MGH).

In the second part of the talk, I will push beyond standard tools, introducing new functionalities and avoiding annotation-hungry training paradigms ill-suited for clinical practice. In particular, I will focus on interpretable neural models that provide rationales underlying their predictions, and semi-supervised methods for information extraction.

Biography

Regina Barzilay is a professor in the Department of Electrical Engineering and Computer Science and a member of the Computer Science and Artificial Intelligence Laboratory at the Massachusetts Institute of Technology. Her research interests are in natural language processing. She is a recipient of various awards including of the NSF Career Award, the MIT Technology Review TR-35 Award, Microsoft Faculty Fellowship and several Best Paper Awards at NAACL and ACL. She received her Ph.D. in Computer Science from Columbia University, and spent a year as a postdoc at Cornell University.

Session 1 Overview – Monday, June 13, 2016

Track A	Track B	Track C	
<i>Machine translation</i> Grande Ballroom A	<i>Summarization</i> Grande Ballroom B	<i>Dialog</i> Grande Ballroom C	
Achieving Accurate Conclusions in Evaluation of Automatic Machine Translation Metrics <i>Graham and Liu</i>	Neural Network-Based Abstract Generation for Opinions and Arguments <i>Wang and Ling</i>	Integer Linear Programming for Discourse Parsing <i>Perret, Afantenos, Asher, and Morey</i>	11:00
Flexible Non-Terminals for Dependency Tree-to-Tree Reordering <i>Richardson, Cromierès, Nakazawa, and Kurohashi</i>	A Low-Rank Approximation Approach to Learning Joint Embeddings of News Stories and Images for Timeline Summarization <i>Wang, Mehdad, Radev, and Stent</i>	A Diversity-Promoting Objective Function for Neural Conversation Models <i>Li, Galley, Brockett, Gao, and Dolan</i>	11:20
Selecting Syntactic, Non-redundant Segments in Active Learning for Machine Translation <i>Miura, Neubig, Paul, and Nakamura</i>	Entity-balanced Gaussian pLSA for Automated Comparison <i>Contractor, Singla, and Mausam</i>	Multi-domain Neural Network Language Generation for Spoken Dialogue Systems <i>Wen, Gašić, Mrkšić, Rojas-Barahona, Su, Vandyke, and Young</i>	11:40
Multi-Source Neural Translation <i>Zoph and Knight</i>	Automatic Summarization of Student Course Feedback <i>Luo, Liu, Liu, and Litman</i>	A Long Short-Term Memory Framework for Predicting Humor in Dialogues <i>Bertero and Fung</i>	12:00
Controlling Politeness in Neural Machine Translation via Side Constraints <i>Sennrich, Haddow, and Birch</i>	Knowledge-Guided Linguistic Rewrites for Inference Rule Verification <i>Jain and Mausam</i>	Conversational Flow in Oxford-style Debates <i>Zhang, Kumar, Ravi, and Danescu-Niculescu-Mizil</i>	12:10
An Empirical Evaluation of Noise Contrastive Estimation for the Neural Network Joint Model of Translation <i>Cherry</i>	Abstractive Sentence Summarization with Attentive Recurrent Neural Networks <i>Chopra, Auli, and Rush</i>	Counter-fitting Word Vectors to Linguistic Constraints <i>Mrkšić, Ó Séaghdha, Thomson, Gašić, Rojas-Barahona, Su, Vandyke, Wen, and Young</i>	12:20

Parallel Session 1

Parallel Session 1A: Machine translation

Grande Ballroom A

Chair: David Chiang

Achieving Accurate Conclusions in Evaluation of Automatic Machine Translation Metrics

Yvette Graham and Qun Liu

11:00–11:20

Automatic Machine Translation metrics, such as BLEU, are widely used in empirical evaluation as a substitute for human assessment. Subsequently, the performance of a given metric is measured by its strength of correlation with human judgment. When a newly proposed metric achieves a stronger correlation over that of a baseline, it is important to take into account the uncertainty inherent in correlation point estimates prior to concluding improvements in metric performance. Confidence intervals for correlations with human judgment are rarely reported in metric evaluations, however, and when they have been reported, appropriate methods have unfortunately not been applied. Several issues including incorrect assumptions about correlation sampling distributions greatly risk the over-estimation of significant differences in metric performance. In this paper, we provide analysis of each of the issues leading to inaccuracies in current evaluations before providing detail of an appropriate method that overcomes previous challenges. Additionally, we propose a new method of translation sampling that in contrast achieves genuine high conclusivity in evaluation of the relative performance of metrics.

Flexible Non-Terminals for Dependency Tree-to-Tree Reordering

John Richardson, Fabien Cromierès, Toshiaki Nakazawa, and Sadao Kurohashi

11:20–11:40

A major benefit of tree-to-tree over tree-to-string translation is that we can use target-side syntax to improve reordering. While this is relatively simple for binarized constituency parses, the reordering problem is considerably harder for dependency parses, in which parent nodes can have arbitrarily many children. Previous approaches have tackled this problem by restricting grammar rules, reducing the expressive power of the translation model. In this paper we propose a general model for dependency tree-to-tree reordering based on flexible non-terminals that can compactly encode multiple insertion positions. We explore how insertion positions can be selected even in cases where rules do not entirely cover the children of input sentence words. The proposed method greatly improves the flexibility of translation rules at the cost of only a 30% increase in decoding time, and we demonstrate a 1.2–1.9 BLEU improvement over a strong tree-to-tree baseline.

Selecting Syntactic, Non-redundant Segments in Active Learning for Machine Translation

Akiva Miura, Graham Neubig, Michael Paul, and Satoshi Nakamura

11:40–12:00

Active learning is a framework that makes it possible to efficiently train statistical models by selecting informative examples from a pool of unlabeled data. Previous work has found this framework effective for machine translation (MT), making it possible to train better translation models with less effort, particularly when annotators translate short phrases instead of full sentences. However, previous methods for phrase-based active learning in MT fail to consider whether the selected units are coherent and easy for human translators to translate, and also have problems with selecting redundant phrases with similar content. In this paper, we tackle this problem by proposing two new methods for selecting more syntactically coherent and less redundant segments in active learning for MT. Experiments using both simulation and extensive manual translation by professional translators find the proposed method effective, achieving both greater gain of BLEU score for the same number of translated words, and allowing translators to be more confident in their translations.

Multi-Source Neural Translation

Barret Zoph and Kevin Knight

12:00–12:10

We build a multi-source machine translation model and train it to maximize the probability of a target English string given French and German sources. Using the neural encoder-decoder framework, we explore several combination methods and report up to +4.8 Bleu increases on top of a very strong attention-based neural translation model.

Controlling Politeness in Neural Machine Translation via Side Constraints

Rico Sennrich, Barry Haddow, and Alexandra Birch

12:10–12:20

Many languages use honorifics to express politeness, social distance, or the relative social status between the speaker and their addressee(s). In machine translation from a language without honorifics such as English, it is difficult to predict the appropriate honorific, but users may want to control the level of politeness in the

output. In this paper, we perform a pilot study to control honorifics in neural machine translation (NMT) via side constraints, focusing on English-German. We show that by marking up the (English) source side of the training data with a feature that encodes the use of honorifics on the (German) target side, we can control the honorifics produced at test time. Experiments show that the choice of honorifics has a big impact on translation quality as measured by BLEU, and oracle experiments show that substantial improvements are possible by constraining the translation to the desired level of politeness.

An Empirical Evaluation of Noise Contrastive Estimation for the Neural Network Joint Model of Translation

Colin Cherry

12:20–12:30

The neural network joint model of translation or NNJM (Devlin et al., 2014) combines source and target context to produce accurate language model probabilities and a powerful translation feature. However, its softmax layer necessitates a sum over the entire output vocabulary, which results in very slow maximum likelihood (MLE) training. This has led some groups to train using Noise Contrastive Estimation (NCE), which side-steps this sum. We carry out the first direct comparison of MLE and NCE training objectives for the NNJM, showing that NCE is significantly outperformed by MLE on large-scale Arabic-English and Chinese-English translation tasks. We also show that this drop can be avoided by using a recently proposed translation noise distribution.

Parallel Session 1B: Summarization

Grande Ballroom B

Chair: Fei Liu

Neural Network-Based Abstract Generation for Opinions and Arguments

Liu Wang and Wang Ling

11:00–11:20

We study the problem of generating abstractive summaries for opinionated text. We propose an attention-based neural network model that is able to absorb information from multiple text units to construct informative, concise, and fluent summaries. An importance-based sampling method is designed to allow the encoder to integrate information from an important subset of input. Automatic evaluation indicates that our system outperforms state-of-the-art abstractive and extractive summarization systems on two newly collected datasets of movie reviews and arguments. Our system summaries are also rated as more informative and grammatical in human evaluation.

A Low-Rank Approximation Approach to Learning Joint Embeddings of News Stories and Images for Timeline Summarization

William Yang Wang, Yashar Mehdad, Dragomir R. Radev, and Amanda Stent

11:20–11:40

A key challenge for timeline summarization is to generate a concise, yet complete storyline from large collections of news stories. Previous studies in extractive timeline generation are limited in two aspects: first, most prior work focuses on fully-observable ranking models or clustering models with hand-designed features that may not generalize well. Second, most summarization corpora are text-only, which means that text is the sole source of information considered in timeline summarization, and thus, the rich visual content from news images is ignored. To solve these issues, we leverage the success of matrix factorization techniques from recommender systems, and cast the generation problem as a sentence recommendation task, using a representation learning approach. To augment text-only corpora, for each candidate sentence in a news article, we take advantage of top-ranked relevant images from the Web and model the image using a convolutional neural network architecture. Finally, we propose a scalable low-rank approximation approach for learning joint embeddings of news stories and images. In experiments, we compare our model to various competitive baselines, and demonstrate the state-of-the-art performance of the proposed text-based and multimodal approaches.

Entity-balanced Gaussian pLSA for Automated Comparison

Danish Contractor, Parag Singla, and Mausam

11:40–12:00

Community created content (e.g., product descriptions, reviews) typically discusses one entity at a time and it can be hard as well as time consuming for a user to compare two or more entities. In response, we define a novel task of automatically generating entity comparisons from text. Our output is a table that semantically clusters descriptive phrases about entities. Our clustering algorithm is a Gaussian extension of probabilistic latent semantic analysis (pLSA), in which each phrase is represented in word vector embedding space. In addition, our algorithm attempts to balance information about entities in each cluster to generate meaningful comparison tables, where possible. We test our system's effectiveness on two domains, travel articles and movie reviews, and find that entity-balanced clusters are strongly preferred by users.

Automatic Summarization of Student Course Feedback

Wencan Luo, Fei Liu, Zitao Liu, and Diane Litman

12:00–12:10

Student course feedback is generated daily in both classrooms and online course discussion forums. Traditionally, instructors manually analyze these responses in a costly manner. In this work, we propose a new approach to summarizing student course feedback based on the integer linear programming (ILP) framework. Our approach allows different student responses to share co-occurrence statistics and alleviates sparsity issue. Experimental results on a real-world student feedback corpus show that our approach outperforms a range of baselines in terms of both ROUGE scores and human evaluation.

Knowledge-Guided Linguistic Rewrites for Inference Rule Verification

Prachi Jain and Mausam

12:10–12:20

A corpus of inference rules between a pair of relation phrases is typically generated using the statistical overlap of argument-pairs associated with the relations (e.g., PATTY, CLEAN). We investigate knowledge-guided linguistic rewrites as a secondary source of evidence and find that they can vastly improve the quality of inference rule corpora, obtaining 27 to 33 point precision improvement while retaining substantial recall. The facts inferred using cleaned inference rules are 29-32 points more accurate.

Abstractive Sentence Summarization with Attentive Recurrent Neural Networks

Sumit Chopra, Michael Auli, and Alexander M. Rush

12:20–12:30

Abstractive sentence summarization generates a shorter version of a given sentence while attempting to preserve its meaning. We introduce a conditional recurrent neural network (RNN) which generates a summary of an input sentence. The conditioning is provided by a novel convolutional attention-based encoder which ensures that the decoder focuses on the appropriate input words at each step of generation. Our model relies only on learned features and is easy to train in an end-to-end fashion on very large data sets. Our experiments show that the model significantly outperforms the recently proposed state-of-the-art method on the Gigaword corpus while performing competitively on the DUC-2004 shared task.

Parallel Session 1C: Dialog

Grande Ballroom C

*Chair: Mari Ostendorf***Integer Linear Programming for Discourse Parsing***Jérémie Perret, Stergos Afantinos, Nicholas Asher, and Mathieu Morey*

11:00–11:20

In this paper we present the first, to the best of our knowledge, discourse parser that is able to predict non-tree DAG structures. We use Integer Linear Programming (ILP) to encode both the objective function and the constraints as global decoding over local scores. Our underlying data come from multi-party chat dialogues, which require the prediction of DAGs. We use the dependency parsing paradigm, as has been done in the past, but we use the underlying formal framework of SDRT and exploit SDRT's notions of left and right distributive relations. We achieve an F-measure of 0.531 for fully labeled structures which beats the previous state of the art.

A Diversity-Promoting Objective Function for Neural Conversation Models*Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan*

11:20–11:40

Sequence-to-sequence neural network models for generation of conversational responses tend to generate safe, commonplace responses (e.g., I don't know) regardless of the input. We suggest that the traditional objective function, i.e., the likelihood of output (response) given input (message) is unsuited to response generation tasks. Instead we propose using Maximum Mutual Information (MMI) as the objective function in neural models. Experimental results demonstrate that the proposed MMI models produce more diverse, interesting, and appropriate responses, yielding substantive gains in BLEU scores on two conversational datasets and in human evaluations.

Multi-domain Neural Network Language Generation for Spoken Dialogue Systems*Tsung-Hsien Wen, Milica Gašić, Nikola Mrkšić, Lina M. Rojas-Barahona, Pei-Hao Su, David Vandyke, and Steve Young*

11:40–12:00

Moving from limited-domain natural language generation (NLG) to open domain is difficult because the number of semantic input combinations grows exponentially with the number of domains. Therefore, it is important to leverage existing resources and exploit similarities between domains to facilitate domain adaptation. In this paper, we propose a procedure to train multi-domain, Recurrent Neural Network-based (RNN) language generators via multiple adaptation steps. In this procedure, a model is first trained on counterfeited data synthesised from an out-of-domain dataset, and then fine tuned on a small set of in-domain utterances with a discriminative objective function. Corpus-based evaluation results show that the proposed procedure can achieve competitive performance in terms of BLEU score and slot error rate while significantly reducing the data needed to train generators in new, unseen domains. In subjective testing, human judges confirm that the procedure greatly improves generator performance when only a small amount of data is available in the domain.

A Long Short-Term Memory Framework for Predicting Humor in Dialogues*Dario Bertero and Pascale Fung*

12:00–12:10

We propose a first-ever attempt to employ a Long Short-Term memory based framework to predict humor in dialogues. We analyze data from a popular TV-sitcom, whose canned laughter give an indication of when the audience would react. We model the setup-punchline relation of conversational humor with a Long Short-Term Memory, with utterance encodings obtained from a Convolutional Neural Network. Our neural network framework is able to improve the F-score of 8% over a Conditional Random Field baseline. We show how the LSTM effectively models the setup-punchline relation reducing the number of false positives and increasing the recall. We aim to employ our humor prediction model to build effective empathetic machine able to understand jokes.

Conversational Flow in Oxford-style Debates*Justine Zhang, Ravi Kumar, Sujith Ravi, and Cristian Danescu-Niculescu-Mizil*

12:10–12:20

Public debates are a common platform for presenting and juxtaposing diverging views on important issues. In this work we propose a methodology for tracking how ideas flow between participants throughout a debate. We use this approach in a case study of Oxford-style debates—a competitive format where the winner is determined by audience votes—and show how the outcome of a debate depends on aspects of conversational flow. In particular, we find that winners tend to make better use of a debate's interactive component than losers, by actively pursuing their opponents' points rather than promoting their own ideas over the course of the conversation.

Counter-fitting Word Vectors to Linguistic Constraints

Nikola Mrkšić, Diarmuid Ó Séaghdha, Blaise Thomson, Milica Gašić, Lina M. Rojas-Barahona, Pei-Hao Su, David Vandyke, Tsung-Hsien Wen, and Steve Young 12:20–12:30

In this work, we present a novel counter-fitting method which injects antonymy and synonymy constraints into vector space representations in order to improve the vectors' capability for judging semantic similarity. Applying this method to publicly available pre-trained word vectors leads to a new state of the art performance on the SimLex-999 dataset. We also show how the method can be used to tailor the word vector space for the downstream task of dialogue state tracking, resulting in robust improvements across different dialogue domains.

Session 2 Overview – Monday, June 13, 2016

	Track A	Track B	Track C
	<i>Language and Vision</i> Grande Ballroom A	<i>Parsing</i> Grande Ballroom B	<i>Named Entity Recognition</i> Grande Ballroom C
2:00	Grounded Semantic Role Labeling <i>Yang, Gao, Liu, Xiong, Zhu, and Chai</i>	[TACL] Efficient Structured Inference for Transition-Based Parsing with Neural Networks and Error States <i>Vaswani and Sagae</i>	An Empirical Study of Automatic Chinese Word Segmentation for Spoken Language Understanding and Named Entity Recognition <i>Luo and Yang</i>
2:20	Black Holes and White Rabbits: Metaphor Identification with Visual Features [BEST LONG PAPER RUNNER-UP] <i>Shutova, Kiel, and Maillard</i>	Recurrent Neural Network Grammars <i>Dyer, Kuncoro, Ballesteros, and Smith</i>	Name Tagging for Low-resource Incident Languages based on Expectation-driven Learning <i>Zhang, Pan, Wang, Vaswani, Ji, Knight, and Marcu</i>
2:40	Bridge Correlational Neural Networks for Multilingual Multimodal Representation Learning <i>Rajendran, Khapra, Chandar, and Ravindran</i>	Expected F-Measure Training for Shift-Reduce Parsing with Recurrent Neural Networks <i>Xu, Auli, and Clark</i>	Neural Architectures for Named Entity Recognition <i>Lample, Ballesteros, Subramanian, Kawakami, and Dyer</i>
3:00	Unsupervised Visual Sense Disambiguation for Verbs using Multimodal Embeddings <i>Gella, Lapata, and Keller</i>	LSTM CCG Parsing <i>Lewis, Lee, and Zettlemoyer</i>	Dynamic Feature Induction: The Last Gist to the State-of-the-Art <i>Choi</i>
3:20	Stating the Obvious: Extracting Visual Common Sense Knowledge <i>Yatskar, Ordonez, and Farhadi</i>	Supertagging With LSTMs <i>Vaswani, Bisk, Sagae, and Musa</i>	Drop-out Conditional Random Fields for Twitter with Huge Mined Gazetteer <i>Yang, Kim, Sarikaya, and Kim</i>

Parallel Session 2

Parallel Session 2A: Language and Vision

Grande Ballroom A

Chair: Meg Mitchell

Grounded Semantic Role Labeling

Shaohua Yang, Qiaozi Gao, Changsong Liu, Caiming Xiong, Song-Chun Zhu, and Joyce Y. Chai
14:00–14:20

Semantic Role Labeling (SRL) captures semantic roles (or participants) such as agent, patient, and theme associated with verbs from the text. While it provides important intermediate semantic representations for many traditional NLP tasks (such as information extraction and question answering), it does not capture grounded semantics so that an artificial agent can reason, learn, and perform the actions with respect to the physical environment. To address this problem, this paper extends traditional SRL to grounded SRL where arguments of verbs are grounded to participants of actions in the physical world. By integrating language and vision processing through joint inference, our approach not only grounds explicit roles, but also grounds implicit roles that are not explicitly mentioned in language descriptions. This paper describes our empirical results and discusses challenges and future directions.

Black Holes and White Rabbits: Metaphor Identification with Visual Features [BEST LONG PAPER RUNNER-UP]

Ekaterina Shutova, Douwe Kiela, and Jean Maillard

14:20–14:40

Metaphor is pervasive in our communication, which makes it an important problem for natural language processing (NLP). Numerous approaches to metaphor processing have thus been proposed, all of which relied on linguistic features and textual data to construct their models. Human metaphor comprehension is, however, known to rely on both our linguistic and perceptual experience, and vision can play a particularly important role when metaphorically projecting imagery across domains. In this paper, we present the first metaphor identification method that simultaneously draws knowledge from linguistic and visual data. Our results demonstrate that it outperforms linguistic and visual models in isolation, as well as being competitive with the best-performing metaphor identification methods, that rely on hand-crafted knowledge about domains and perception.

Bridge Correlational Neural Networks for Multilingual Multimodal Representation Learning

Janarthanan Rajendran, Mitesh M. Khapra, Sarath Chandar, and Balaraman Ravindran 14:40–15:00

Recently there has been a lot of interest in learning common representations for multiple views of data. Typically, such common representations are learned using a parallel corpus between the two views (say, 1M images and their English captions). In this work, we address a real-world scenario where no direct parallel data is available between two views of interest (say, V1 and V2) but parallel data is available between each of these views and a pivot view (V3). We propose a model for learning a common representation for V1, V2 and V3 using only the parallel data available between V1V3 and V2V3. The proposed model is generic and even works when there are n views of interest and only one pivot view which acts as a bridge between them. There are two specific downstream applications that we focus on (i) transfer learning between languages L1,L2,...,Ln using a pivot language L and (ii) cross modal access between images and a language L1 using a pivot language L2. Our model achieves state-of-the-art performance in multilingual document classification on the publicly available multilingual TED corpus and promising results in multilingual multimodal retrieval on a new dataset created and released as a part of this work.

Unsupervised Visual Sense Disambiguation for Verbs using Multimodal Embeddings

Spandana Gella, Mirella Lapata, and Frank Keller

15:00–15:20

We introduce a new task, visual sense disambiguation for verbs: given an image and a verb, assign the correct sense of the verb, i.e., the one that describes the action depicted in the image. Just as textual word sense disambiguation is useful for a wide range of NLP tasks, visual sense disambiguation can be useful for multimodal tasks such as image retrieval, image description, and text illustration. We introduce VerSe, a new dataset that augments existing multimodal datasets (COCO and TUHOI) with sense labels. We propose an unsupervised algorithm based on Lesk which performs visual sense disambiguation using textual, visual, or multimodal embeddings. We find that textual embeddings perform well when gold-standard textual annotations (object labels and image descriptions) are available, while multimodal embeddings perform well on unannotated im-

ages. We also verify our findings by using the textual and multimodal embeddings as features in a supervised setting and analyse the performance of visual sense disambiguation task. VerSe is made publicly available and can be downloaded at: <https://github.com/spandanagella/verse>.

Stating the Obvious: Extracting Visual Common Sense Knowledge

Mark Yatskar, Vicente Ordonez, and Ali Farhadi

15:20–15:30

Obtaining common sense knowledge using current information extraction techniques is extremely challenging. In this work, we instead propose to derive simple common sense statements from fully annotated object detection corpora such as the Microsoft Common Objects in Context dataset. We show that many thousands of common sense facts can be extracted from such corpora at high quality. Furthermore, using WordNet and a novel sub-modular k-coverage formulation, we are able to generalize our initial set of common sense assertions to unseen objects and uncover over 400k potentially useful facts.

Parallel Session 2B: Parsing

Grande Ballroom B

Chair: Alexander M. Rush

[TACL] Efficient Structured Inference for Transition-Based Parsing with Neural Networks and Error States*Ashish Vaswani and Kenji Sagae*

14:00–14:20

Transition-based approaches based on local classification are attractive for dependency parsing due to their simplicity and speed, despite producing results slightly below the state-of-the-art. In this paper, we propose a new approach for approximate structured inference for transition-based parsing that produces scores suitable for global scoring using local models. This is accomplished with the introduction of error states in local training, which add information about incorrect derivation paths typically left out completely in locally-trained models. Using neural networks for our local classifiers, our approach achieves 93.61% accuracy for transition-based dependency parsing in English.

Recurrent Neural Network Grammars*Chris Dyer, Adhiguna Kuncoro, Miguel Ballesteros, and Noah A. Smith*

14:20–14:40

We introduce recurrent neural network grammars, probabilistic models of sentences with explicit phrase structure. We explain efficient inference procedures that allow application to both parsing and language modeling. Experiments show that they provide better parsing than any single previously published supervised generative model and better language modeling than state-of-the-art sequential RNNs.

Expected F-Measure Training for Shift-Reduce Parsing with Recurrent Neural Networks*Wenduan Xu, Michael Auli, and Stephen Clark*

14:40–15:00

We present expected F-measure training for shift-reduce parsing with RNNs, which enables the learning of a global parsing model optimized for sentence-level F1. We apply the model to CCG parsing, where it improves over a strong greedy RNN baseline, by 1.47% F1, yielding state-of-the-art results for shift-reduce CCG parsing.

LSTM CCG Parsing*Mike Lewis, Kenton Lee, and Luke Zettlemoyer*

15:00–15:20

We demonstrate that a state-of-the-art parser can be built using only a lexical tagging model and a deterministic grammar, with no explicit model of bi-lexical dependencies. Instead, all dependencies are implicitly encoded in an LSTM supertagger that assigns CCG lexical categories. The parser significantly outperforms all previously published CCG results, supports efficient and optimal A* decoding, and benefits substantially from semi-supervised tri-training. We give a detailed analysis, demonstrating that the parser can recover long-range dependencies with high accuracy and that the semi-supervised learning enables significant accuracy gains. By running the LSTM on a GPU, we are able to parse over 2600 sentences per second while improving state-of-the-art accuracy by 1.1 F1 in domain and up to 4.5 F1 out of domain.

Supertagging With LSTMs*Ashish Vaswani, Yonatan Bisk, Kenji Sagae, and Ryan Musa*

15:20–15:30

In this paper we present new state-of-the-art performance on CCG supertagging and parsing. Our model outperforms existing approaches by an absolute gain of 1.5%. We analyze the performance of several neural models and demonstrate that while feed-forward architectures can compete with bidirectional LSTMs on POS tagging, models that encode the complete sentence are necessary for the long range syntactic information encoded in supertags.

Parallel Session 2C: Named Entity Recognition

Grande Ballroom C

Chair: Alessandro Moschitti

An Empirical Study of Automatic Chinese Word Segmentation for Spoken Language Understanding and Named Entity Recognition

Wencan Luo and Fan Yang

14:00–14:20

Word segmentation is usually recognized as the first step for many Chinese natural language processing tasks, yet its impact on these subsequent tasks is relatively under-studied. For example, how to solve the *mismatch* problem when applying an existing word segmenter to new data? Does a better word segmenter yield a better subsequent NLP task performance? In this work, we conduct an initial attempt to answer these questions on two related subsequent tasks: semantic slot filling in spoken language understanding and named entity recognition. We propose three techniques to solve the mismatch problem: using word segmentation outputs as additional features, adaptation with partial-learning and taking advantage of n-best word segmentation list. Experimental results demonstrate the effectiveness of these techniques for both tasks and we achieve an error reduction of about 11% for spoken language understanding and 24% for named entity recognition over the baseline systems.

Name Tagging for Low-resource Incident Languages based on Expectation-driven Learning

Boliang Zhang, Xiaoman Pan, Tianlu Wang, Ashish Vaswani, Heng Ji, Kevin Knight, and Daniel Marcu

14:20–14:40

In this paper we tackle a challenging name tagging problem in an emergent setting - the tagger needs to be complete within a few hours for a new incident language (IL) using very few resources. Inspired by observing how human annotators attack this challenge, we propose a new expectation-driven learning framework. In this framework we rapidly acquire, categorize, structure and zoom in on IL-specific expectations (rules, features, patterns, gazetteers, etc.) from various non-traditional sources: consulting and encoding linguistic knowledge from native speakers, mining and projecting patterns from both mono-lingual and cross-lingual corpora, and typing based on cross-lingual entity linking. We also propose a cost-aware combination approach to compose expectations. Experiments on seven low-resource languages demonstrate the effectiveness and generality of this framework: we are able to setup a name tagger for a new IL within two hours, and achieve 33.8%-65.1% F-score.

Neural Architectures for Named Entity Recognition

Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer

14:40–15:00

State-of-the-art named entity recognition systems rely heavily on hand-crafted features and domain-specific knowledge in order to learn effectively from the small, supervised training corpora that are available. In this paper, we introduce two new neural architectures—one based on bidirectional LSTMs and conditional random fields, and the other that constructs and labels segments using a transition-based approach inspired by shift-reduce parsers. Our models rely on two sources of information about words: character-based word representations learned from the supervised corpus and unsupervised word representations learned from unannotated corpora. Our models obtain state-of-the-art performance in NER in four languages without resorting to any language-specific knowledge or resources such as gazetteers.

Dynamic Feature Induction: The Last Gist to the State-of-the-Art

Jinho D. Choi

15:00–15:20

We introduce a novel technique called dynamic feature induction that keeps inducing high dimensional features automatically until the feature space becomes ‘more’ linearly separable. Dynamic feature induction searches for the feature combinations that give strong clues for distinguishing certain label pairs, and generates joint features from these combinations. These induced features are trained along with the primitive low dimensional features. Our approach was evaluated on two core NLP tasks, part-of-speech tagging and named entity recognition, and showed the state-of-the-art results for both tasks, achieving the accuracy of 97.64 and the F1-score of 91.00 respectively, with about a 25% increase in the feature space.

Drop-out Conditional Random Fields for Twitter with Huge Mined Gazetteer

Eunsuk Yang, Young-Bum Kim, Ruhî Sarıkaya, and Yu-Seop Kim

15:20–15:30

In named entity recognition task especially for massive data like Twitter, having a large amount of high quality gazetteers can alleviate the problem of training data scarcity. One could collect large gazetteers from knowl-

edge graph and phrase embeddings to obtain high coverage of gazetteers. However, large gazetteers cause a side-effect called “feature under-training”, where the gazetteer features overwhelm the context features. To resolve this problem, we propose the dropout conditional random fields, which decrease the influence of gazetteer features with a high weight. Our experiments on named entity recognition with Twitter data lead to higher F1 score of 69.38%, about 4% better than the strong baseline presented in Smith and Osborne (2006).

Session 3 Overview – Monday, June 13, 2016

	Track A	Track B	Track C
	<i>Event detection</i> Grande Ballroom A	<i>Language Models</i> Grande Ballroom B	<i>Non Literal Language</i> Grande Ballroom C
4:00	Joint Extraction of Events and Entities within a Document Context <i>Yang and Mitchell</i>	Top-down Tree Long Short-Term Memory Networks <i>Zhang, Lu, and Lapata</i>	Questioning Arbitrariness in Language: a Data-Driven Study of Conventional Iconicity <i>Abramova and Fernández</i>
4:20	[TACL] A Hierarchical Distance-dependent Bayesian Model for Event Coreference Resolution <i>Yang, Cardie, and Fazier</i>	Recurrent Memory Networks for Language Modeling <i>Tran, Bisazza, and Monz</i>	Distinguishing Literal and Non-Literal Usage of German Particle Verbs <i>Köper and Schulte im Walde</i>
4:40	Joint Event Extraction via Recurrent Neural Networks <i>Nguyen, Cho, and Grishman</i>	A Latent Variable Recurrent Neural Network for Discourse-Driven Language Models <i>Ji, Haffari, and Eisenstein</i>	Phrasal Substitution of Idiomatic Expressions <i>Liu and Hwa</i>

Parallel Session 3

Parallel Session 3A: Event detection

Grande Ballroom A

Chair: Heng Ji

Joint Extraction of Events and Entities within a Document Context

Bishan Yang and Tom M. Mitchell

16:00–16:20

Events and entities are closely related; entities are often actors or participants in events and events without entities are uncommon. The interpretation of events and entities is highly contextually dependent. Existing work in information extraction typically models events separately from entities, and performs inference at the sentence level, ignoring the rest of the document. In this paper, we propose a novel approach that models the dependencies among variables of events, entities, and their relations, and performs joint inference of these variables across a document. The goal is to enable access to document-level contextual information and facilitate context-aware predictions. We demonstrate that our approach substantially outperforms the state-of-the-art methods for event extraction as well as a strong baseline for entity extraction.

[TACL] A Hierarchical Distance-dependent Bayesian Model for Event Coreference Resolution

Bishan Yang, Claire Cardie, and Peter Frazier

16:20–16:40

We present a novel hierarchical distance-dependent Bayesian model for event coreference resolution. While existing generative models for event coreference resolution are completely unsupervised, our model allows for the incorporation of pairwise distances between event mentions — information that is widely used in supervised coreference models to guide the generative clustering processing for better event clustering both within and across documents. We model the distances between event mentions using a feature-rich learnable distance function and encode them as Bayesian priors for nonparametric clustering. Experiments on the ECB+ corpus show that our model outperforms state-of-the-art methods for both within- and cross-document event coreference resolution.

Joint Event Extraction via Recurrent Neural Networks

Thien Huu Nguyen, Kyunghyun Cho, and Ralph Grishman

16:40–17:00

Event extraction is a particularly challenging problem in information extraction. The state-of-the-art models for this problem have either applied convolutional neural networks in a pipelined framework (Chen et al., 2015) or followed the joint architecture via structured prediction with rich local and global features (Li et al., 2013). The former is able to learn hidden feature representations automatically from data based on the continuous and generalized representations of words. The latter, on the other hand, is capable of mitigating the error propagation problem of the pipelined approach and exploiting the inter-dependencies between event triggers and argument roles via discrete structures. In this work, we propose to do event extraction in a joint framework with bidirectional recurrent neural networks, thereby benefiting from the advantages of the two models as well as addressing issues inherent in the existing approaches. We systematically investigate different memory features for the joint model and demonstrate that the proposed model achieves the state-of-the-art performance on the ACE 2005 dataset.

Parallel Session 3B: Language Models

Grande Ballroom B

Chair: Chris Dyer

Top-down Tree Long Short-Term Memory Networks

Xingxing Zhang, Liang Lu, and Mirella Lapata

16:00–16:20

Long Short-Term Memory (LSTM) networks, a type of recurrent neural network with a more complex computational unit, have been successfully applied to a variety of sequence modeling tasks. In this paper we develop Tree Long Short-Term Memory (TreeLSTM), a neural network model based on LSTM, which is designed to predict a tree rather than a linear sequence. TreeLSTM defines the probability of a sentence by estimating the generation probability of its dependency tree. At each time step, a node is generated based on the representation of the generated sub-tree. We further enhance the modeling power of TreeLSTM by explicitly representing the correlations between left and right dependents. Application of our model to the MSR sentence completion challenge achieves results beyond the current state of the art. We also report results on dependency parsing reranking achieving competitive performance.

Recurrent Memory Networks for Language Modeling

Ke Tran, Arianna Bisazza, and Christof Monz

16:20–16:40

Recurrent Neural Networks (RNN) have obtained excellent result in many natural language processing (NLP) tasks. However, understanding and interpreting the source of this success remains a challenge. In this paper, we propose Recurrent Memory Network (RMN), a novel RNN architecture, that not only amplifies the power of RNN but also facilitates our understanding of its internal functioning and allows us to discover underlying patterns in data. We demonstrate the power of RMN on language modeling and sentence completion tasks. On language modeling, RMN outperforms Long Short-Term Memory (LSTM) network on three large German, Italian, and English dataset. Additionally we perform in-depth analysis of various linguistic dimensions that RMN captures. On Sentence Completion Challenge, for which it is essential to capture sentence coherence, our RMN obtains 69.5% accuracy, surpassing the previous state-of-the-art by a large margin.

A Latent Variable Recurrent Neural Network for Discourse-Driven Language Models

Yangfeng Ji, Gholamreza Haffari, and Jacob Eisenstein

16:40–17:00

This paper presents a novel latent variable recurrent neural network architecture for jointly modeling sequences of words and (possibly latent) discourse relations between adjacent sentences. A recurrent neural network generates individual words, thus reaping the benefits of discriminatively-trained vector representations. The discourse relations are represented with a latent variable, which can be predicted or marginalized, depending on the task. The resulting model can therefore employ a training objective that includes not only discourse relation classification, but also word prediction. As a result, it outperforms state-of-the-art alternatives for two tasks: implicit discourse relation classification in the Penn Discourse Treebank, and dialog act classification in the Switchboard corpus. Furthermore, by marginalizing over latent discourse relations at test time, we obtain a discourse informed language model, which improves over a strong LSTM baseline.

Parallel Session 3C: Non Literal Language

Grande Ballroom C

Chair: Marie-Catherine de Marneffe

Questioning Arbitrariness in Language: a Data-Driven Study of Conventional Iconicity

Ekaterina Abramova and Raquel Fernández

16:00–16:20

This paper presents a data-driven investigation of “phonesthemes”, phonetic units said to carry meaning associations, thus challenging the traditionally assumed arbitrariness of language. Phonesthemes have received a substantial amount of attention within the cognitive science literature on sound iconicity, but nevertheless remain a controversial and understudied phenomenon. Here we employ NLP techniques to address two main questions: How can the existence of phonesthemes be tested at a large scale with quantitative methods? And how can the meaning arguably carried by a phonestheme be induced automatically from word embeddings? We develop novel methods to make progress on these fronts and compare our results to previous work, obtaining substantial improvements.

Distinguishing Literal and Non-Literal Usage of German Particle Verbs

Maximilian Köper and Sabine Schulze im Walde

16:20–16:40

This paper provides a binary, token-based classification of German particle verbs (PVs) into literal vs. non-literal usage. A random forest improving standard features (e.g., bag-of-words; affective ratings) with PV-specific information and abstraction over common nouns significantly outperforms the majority baseline. In addition, PV-specific classification experiments demonstrate the role of shared particle semantics and semantically related base verbs in PV meaning shifts.

Phrasal Substitution of Idiomatic Expressions

Changsheng Liu and Rebecca Hwa

16:40–17:00

Idioms pose a great challenge to natural language understanding. A system that can automatically paraphrase idioms in context has applications in many NLP tasks. This paper proposes a phrasal substitution method to replace idioms with their figurative meanings in literal English. Our approach identifies relevant replacement phrases from an idiom’s dictionary definition and performs appropriate grammatical and referential transformations to ensure that the idiom substitution fits seamlessly into the original context. The proposed method has been evaluated both by automatic metrics and human judgments. Results suggest that high quality paraphrases of idiomatic expressions can be achieved.

One Minute Madness

Time: 5:15–6:00

Location: Grande Ballroom

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

Chair: Joel Tetreault

Poster Session 1

Time: 6:00–8:00

Location: Pavilion

Leverage Financial News to Predict Stock Price Movements Using Word Embeddings and Deep Neural Networks

Yangtuo Peng and Hui Jiang

Financial news contains useful information on public companies and the market. In this paper we apply the popular word embedding methods and deep neural networks to leverage financial news to predict stock price movements in the market. Experimental results have shown that our proposed methods are simple but very effective, which can significantly improve the stock prediction accuracy on a standard financial database over the baseline system using only the historical price information.

Grammatical error correction using neural machine translation

Zheng Yuan and Ted Briscoe

This paper presents the first study using neural machine translation (NMT) for grammatical error correction (GEC). We propose a two-step approach to handle the rare word problem in NMT, which has been proved to be useful and effective for the GEC task. Our best NMT-based system trained on the CLC outperforms our SMT-based system when testing on the publicly available FCE test set. The same system achieves an F-0.5 score of 39.90% on the CoNLL-2014 shared task test set, outperforming the state-of-the-art and demonstrating that the NMT-based GEC system generalises effectively.

Multimodal Semantic Learning from Child-Directed Input

Angeliki Lazaridou, Grzegorz Chrupała, Raquel Fernández, and Marco Baroni

Children learn the meaning of words by being exposed to perceptually rich situations (linguistic discourse, visual scenes, etc). Current computational learning models typically simulate these rich situations through impoverished symbolic approximations. In this work, we present a distributed word learning model that operates on child-directed speech paired with realistic visual scenes. The model integrates linguistic and extra-linguistic information (visual and social cues), handles referential uncertainty, and correctly learns to associate words with objects, even in cases of limited linguistic exposure.

Recurrent Support Vector Machines For Slot Tagging In Spoken Language Understanding

Yangyang Shi, Kaisheng Yao, Hu Chen, Dong Yu, Yi-Cheng Pan, and Mei-Yuh Hwang

We propose recurrent support vector machine(RSVM) for slot tagging. This model is a combination of the recurrent neural network (RNN) and the structured support vector machine. RNN extracts features from the input sequence. The structured support vector machine uses a sequence-level discriminative objective function. The proposed model therefore combines the sequence representation capability of an RNN with the sequence-level discriminative objective. We have observed new state-of-the-art results on two benchmark datasets and one private dataset. RSVM obtained statistical significant 4% and 2% relative average F1 score improvement on ATIS dataset and Chunking dataset, respectively. Out of eight domains in Cortana live log dataset, RSVM achieved F1 score improvement on seven domains. Experiments also show that RSVM significantly speeds up the model training by skipping the weight updating for non-support vector training samples, compared against training using RNN with CRF or minimum cross-entropy objectives.

Expectation-Regulated Neural Model for Event Mention Extraction*Ching-Yun Chang, Zhiyang Teng, and Yue Zhang*

We tackle the task of extracting tweets that mention a specific event from all tweets that contain relevant keywords, for which the main challenges include unbalanced positive and negative cases, and the unavailability of manually labeled training data. Existing methods leverage a few manually given seed events and large unlabeled tweets to train a classifier, by using expectation regularization training with discrete ngram features. We propose a LSTM-based neural model that learns tweet-level features automatically. Compared with discrete ngram features, the neural model can potentially capture non-local dependencies and deep semantic information, which are more effective for disambiguating subtle semantic differences between true event mentions and false cases that use similar wording patterns. Results on both tweets and forum posts show that our neural model is more effective compared with a state-of-the-art discrete baseline.

Agreement on Target-bidirectional Neural Machine Translation*Lemao Liu, Masao Utiyama, Andrew Finch, and Eiichiro Sumita*

Neural machine translation (NMT) with recurrent neural networks, has proven to be an effective technique for end-to-end machine translation. However, in spite of its promising advances over traditional translation methods, it typically suffers from an issue of unbalanced outputs, that arise from both the nature of recurrent neural networks themselves, and the challenges inherent in machine translation. To overcome this issue, we propose an agreement model for neural machine translation and show its effectiveness on large-scale Japanese-to-English and Chinese-to-English translation tasks. Our results show the model can achieve improvements of up to 1.4 BLEU over the strongest baseline NMT system. Furthermore, this new end-to-end NMT approach substantially outperformed a phrase-based Moses baseline by up to 5.6 BLEU points.

Psycholinguistic Features for Deceptive Role Detection in Werewolf*Codruta Girlea, Roxana Girju, and Eyal Amir*

We tackle the problem of identifying deceptive agents in highly-motivated high-conflict dialogues. We consider the case where we only have textual information. We show the usefulness of psycho-linguistic deception and persuasion features on a small dataset for the game of Werewolf. We analyse the role of syntax and we identify some characteristics of players in deceptive roles.

Individual Variation in the Choice of Referential Form*Thiago Castro Ferreira, Emiel Krahmer, and Sander Wubben*

This study aims to measure the variation between writers in their choices of referential form by collecting and analysing a new and publicly available corpus of referring expressions. The corpus is composed of referring expressions produced by different participants in identical situations. Results, measured in terms of normalized entropy, reveal substantial individual variation. We discuss the problems and prospects of this finding for automatic text generation applications.

Joint Learning Templates and Slots for Event Schema Induction*Lei Sha, Sujian Li, Baobao Chang, and Zhifang Sui*

Automatic event schema induction (AESI) means to extract meta-event from raw text, in other words, to find out what types (templates) of event may exist in the raw text and what roles (slots) may exist in each event type. In this paper, we propose a joint entity-driven model to learn templates and slots simultaneously based on the constraints of templates and slots in the same sentence. In addition, the entities' semantic information is also considered for the inner connectivity of the entities. We borrow the normalized cut criteria in image segmentation to divide the entities into more accurate template clusters and slot clusters. The experiment shows that our model gains a relatively higher result than previous work.

Inferred Psycholinguistic Properties of Words*Gustavo Paetzold and Lucia Specia*

We introduce a bootstrapping algorithm for regression that exploits word embedding models. We use it to infer four psycholinguistic properties of words: Familiarity, Age of Acquisition, Concreteness and Imagery and further populate the MRC Psycholinguistic Database with these properties. The approach achieves 0.88 correlation with human-produced values and the inferred psycholinguistic features lead to state-of-the-art results when used in a Lexical Simplification task.

Intra-Topic Variability Normalization based on Linear Projection for Topic Classification*Quan Liu, Wu Guo, Zhen-Hua Ling, Hui Jiang, and Yu Hu*

This paper proposes a variability normalization algorithm to reduce the variability between intra-topic documents for topic classification. Firstly, an optimization problem is constructed based on linear variability removable assumption. Secondly, a new feature space for document representation is found by solving the optimization problem with kernel principle component analysis (KPCA). Finally, effective feature transformation is taken through linear projection. As for experiments, state-of-the-art SVM and KNN algorithm are adopted for topic classification respectively. Experimental results on a free-style conversational corpus show that the proposed variability normalization algorithm for topic classification achieves 3.8% absolute improvement for micro-F1 measure.

Shift-Reduce CCG Parsing using Neural Network Models

Bharat Ram Ambati, Tejaswini Deoskar, and Mark Steedman

We present a neural network based shift-reduce CCG parser, the first neural-network based parser for CCG. We also study the impact of neural network based tagging models, and greedy versus beam-search parsing, by using a structured neural network model. Our greedy parser obtains a labeled F-score of 83.27%, the best reported result for greedy CCG parsing in the literature (an improvement of 2.5% over a perceptron based greedy parser) and is more than three times faster. With a beam, our structured neural network model gives a labeled F-score of 85.57% which is 0.6% better than the perceptron based counterpart.

Online Multilingual Topic Models with Multi-Level Hyperpriors

Kriste Krstovski, David Smith, and Michael J. Kurtz

For topic models, such as LDA, that use a bag-of-words assumption, it becomes especially important to break the corpus into appropriately-sized “documents”. Since the models are estimated solely from the term cooccurrences, extensive documents such as books or long journal articles lead to diffuse statistics, and short documents such as forum posts or product reviews can lead to sparsity. This paper describes practical inference procedures for hierarchical models that smooth topic estimates for smaller sections with hyperpriors over larger documents. Importantly for large collections, these online variational Bayes inference methods perform a single pass over a corpus and achieve better perplexity than “flat” topic models on monolingual and multilingual data. Furthermore, on the task of detecting document translation pairs in large multilingual collections, polylingual topic models with multi-level hyperpriors (mlhPLTM) achieve significantly better performance than existing online PLTM models while retaining computational efficiency.

STransE: a novel embedding model of entities and relationships in knowledge bases

Dat Quoc Nguyen, Kairit Sirts, Lizhen Qu, and Mark Johnson

Knowledge bases of real-world facts about entities and their relationships are useful resources for a variety of natural language processing tasks. However, because knowledge bases are typically incomplete, it is useful to be able to perform link prediction, i.e., predict whether a relationship not in the knowledge base is likely to be true. This paper combines insights from several previous link prediction models into a new embedding model STransE that represents each entity as a low-dimensional vector, and each relation by two matrices and a translation vector. STransE is a simple combination of the SE and TransE models, but it obtains better link prediction performance on two benchmark datasets than previous embedding models. Thus, STransE can serve as a new baseline for the more complex models in the link prediction task.

An Unsupervised Model of Orthographic Variation for Historical Document Transcription

Dan Garrette and Hannah Alpert-Abrams

Historical documents frequently exhibit extensive orthographic variation, including archaic spellings and obsolete shorthand. OCR tools typically seek to produce so-called diplomatic transcriptions that preserve these variants, but many end tasks require transcriptions with normalized orthography. In this paper, we present a novel joint transcription model that learns, unsupervised, a probabilistic mapping between modern orthography and that used in the document. Our system thus produces dual diplomatic and normalized transcriptions simultaneously, and achieves a 35% relative error reduction over a state-of-the-art OCR model on diplomatic transcription, and a 46% reduction on normalized transcription.

Bidirectional RNN for Medical Event Detection in Electronic Health Records

Abhyuday N Jagannatha and Hong Yu

Sequence labeling for extraction of medical events and their attributes from unstructured text in Electronic Health Record (EHR) notes is a key step towards semantic understanding of EHRs. It has important applications in health informatics including pharmacovigilance and drug surveillance. The state of the art supervised

machine learning models in this domain are based on Conditional Random Fields (CRFs) with features calculated from fixed context windows. However, different types of medical events have dependencies in different context window sizes. In this application, we explored recurrent neural network frameworks and show that they significantly outperformed the CRF models. We analyze the performance of the RNN models and show that they can more efficiently discriminate between medical entities, using the same context size as the CRF models.

The Sensitivity of Topic Coherence Evaluation to Topic Cardinality

Jey Han Lau and Timothy Baldwin

When evaluating the quality of topics generated by a topic model, the convention is to score topic coherence — either manually or automatically — using the top-N topic words. This hyper-parameter N , or the cardinality of the topic, is often overlooked and selected arbitrarily. In this paper, we investigate the impact of this cardinality hyper-parameter on topic coherence evaluation. For two automatic topic coherence methodologies, we observe that the correlation with human ratings decreases systematically as the cardinality increases. More interestingly, we find that performance can be improved if the system scores and human ratings are aggregated over several topic cardinalities before computing the correlation. In contrast to the standard practice of using a fixed value of N (e.g. N = 5 or N = 10), our results suggest that calculating topic coherence over several different cardinalities and averaging results in a substantially more stable and robust evaluation. We release the code and the datasets used in this research, for reproducibility.

Transition-Based Syntactic Linearization with Lookahead Features

Ratish Puduppully, Yue Zhang, and Manish Srivastava

It has been shown that transition-based methods can be used for syntactic word ordering and tree linearization, achieving significantly faster speed compared with traditional best-first methods. State-of-the-art transition-based models give competitive results on abstract word ordering and unlabeled tree linearization, but significantly worse results on labeled tree linearization. We demonstrate that the main cause for the performance bottleneck is the sparsity of SHIFT transition actions rather than heavy pruning. To address this issue, we propose a modification to the standard transition-based feature structure, which reduces feature sparsity and allows lookahead features at a small cost to decoding efficiency. Our model gives the best reported accuracies on all benchmarks, yet still being over 30 times faster compared with best-first-search.

A Recurrent Neural Networks Approach for Estimating the Quality of Machine Translation Output

Hyun Kim and Jong-Hyeok Lee

This paper presents a novel approach using recurrent neural networks for estimating the quality of machine translation output. A sequence of vectors made by the prediction method is used as the input of these recurrent neural networks. The prediction method uses bi-directional recurrent neural networks architecture both on source and target sentence to fully utilize the bi-directional quality information from source and target sentence. Our experiments show that the proposed recurrent neural networks approach achieves a performance comparable to the existing state-of-the-art models for estimating the sentence-level quality of English-to-Spanish translation.

Symmetric Patterns and Coordinations: Fast and Enhanced Representations of Verbs and Adjectives

Roy Schwartz, Roi Reichart, and Ari Rappoport

State-of-the-art word embeddings, which are often trained on bag-of-words (BOW) contexts, provide a high quality representation of aspects of the semantics of nouns. However, their quality decreases substantially for the task of verb similarity prediction. In this paper we show that using symmetric pattern contexts (SPs, e.g., “X and Y”) improves word2vec verb similarity performance by up to 15% and is also instrumental in adjective similarity prediction. The unsupervised SP contexts are even superior to a variety of dependency contexts extracted using a supervised parser. Moreover, we observe that SPs and dependency coordination contexts (Coor) capture a similar type of information, and demonstrate that Coor contexts are superior to other dependency contexts including the set of all dependency contexts, although they are still inferior to SPs. Finally, there are substantially fewer SP contexts compared to alternative representations, leading to a massive reduction in training time. On an 8G words corpus and a 32 core machine, the SP model trains in 11 minutes, compared to 5 and 11 hours with BOW and all dependency contexts, respectively.

Breaking the Closed World Assumption in Text Classification

Geli Fei and Bing Liu

Existing research on multiclass text classification mostly makes the closed world assumption, which focuses on designing accurate classifiers under the assumption that all test classes are known at training time. A more realistic scenario is to expect unseen classes during testing (open world). In this case, the goal is to design a learning system that classifies documents of the known classes into their respective classes and also to reject documents from unknown classes. This problem is called open (world) classification. This paper approaches the problem by reducing the open space risk while balancing the empirical risk. It proposes to use a new learning strategy, called center-based similarity (CBS) space learning (or CBS learning), to provide a novel solution to the problem. Extensive experiments across two datasets show that CBS learning gives promising results on multiclass open text classification compared to state-of-the-art baselines.

Sequential Short-Text Classification with Recurrent and Convolutional Neural Networks

Ji Young Lee and Franck Dernoncourt

Recent approaches based on artificial neural networks (ANNs) have shown promising results for short-text classification. However, many short texts occur in sequences (e.g., sentences in a document or utterances in a dialog), and most existing ANN-based systems do not leverage the preceding short texts when classifying a subsequent one. In this work, we present a model based on recurrent neural networks and convolutional neural networks that incorporates the preceding short texts. Our model achieves state-of-the-art results on three different datasets for dialog act prediction.

Improved Neural Network-based Multi-label Classification with Better Initialization Leveraging Label Co-occurrence

Gakuto Kurata, Bing Xiang, and Bowen Zhou

In a multi-label text classification task, in which multiple labels can be assigned to one text, label co-occurrence itself is informative. We propose a novel neural network initialization method to treat some of the neurons in the final hidden layer as dedicated neurons for each pattern of label co-occurrence. These dedicated neurons are initialized to connect to the corresponding co-occurring labels with stronger weights than to others. In experiments with a natural language query classification task, which requires multi-label classification, our initialization method improved classification accuracy without any computational overhead in training and evaluation.

Learning Distributed Word Representations For Bidirectional LSTM Recurrent Neural Network

Peilu Wang, Yao Qian, Frank K. Soong, Lei He, and Hai Zhao

Bidirectional long short-term memory (BLSTM) recurrent neural network (RNN) has been successfully applied in many tagging tasks and shows state-of-the-art performance. To involve word-level information, BLSTM-RNN uses the distributed representations of words. Thus training better distributed word representations is an alternative way to improve the performance of BLSTM-RNN on tagging tasks aside refining model. In this work, we propose a novel approach to learn distributed word representations with BLSTM-RNN. Experimental results show that our approach learns high qualified distributed word representations, as the trained representations significantly elevate the performance of BLSTM-RNN on three tagging tasks: part-of-speech tagging, chunking and named entity recognition, surpassing word representations trained by other published methods.

Combining Recurrent and Convolutional Neural Networks for Relation Classification

Ngoc Thang Vu, Heike Adel, Pankaj Gupta, and Hinrich Schütze

This paper investigates two different neural architectures for the task of relation classification: convolutional neural networks and recurrent neural networks. For both models, we demonstrate the effect of different architectural choices. We present a new context representation for convolutional neural networks for relation classification (extended middle context). Furthermore, we propose connectionist bi-directional recurrent neural networks and introduce ranking loss for their optimization. Finally, we show that combining convolutional and recurrent neural networks using a simple voting scheme is accurate enough to improve results. Our neural models achieve state-of-the-art results on the SemEval 2010 relation classification task.

Building Chinese Affective Resources in Valence-Arousal Dimensions

Liang-Chih Yu, Lung-Hao Lee, Shuai Hao, Jin Wang, Yunchao He, Jun Hu, K. Robert Lai, and Xuejie Zhang

An increasing amount of research has recently focused on representing affective states as continuous numerical values on multiple dimensions, such as the valence-arousal (VA) space. Compared to the categorical approach that represents affective states as several classes (e.g., positive and negative), the dimensional approach can provide more fine-grained sentiment analysis. However, affective resources with valence-arousal ratings are still very rare, especially for the Chinese language. Therefore, this study builds 1) an affective lexicon called Chinese valence-arousal words (CVAW) containing 1,653 words, and 2) an affective corpus called Chinese valence-arousal text (CVAT) containing 2,009 sentences extracted from web texts. To improve the annotation quality, a corpus cleanup procedure is used to remove outlier ratings and improper texts. Experiments using CVAW words to predict the VA ratings of the CVAT corpus show results comparable to those obtained using English affective resources.

Improving event prediction by representing script participants

Simon Ahrendt and Vera Demberg

Automatically learning script knowledge has proved difficult, with previous work not or just barely beating a most-frequent baseline. Script knowledge is a type of world knowledge which can however be useful for various task in NLP and psycholinguistic modelling. We here propose a model that includes participant information (i.e., knowledge about which participants are relevant for a script) and show, on the Dinners from Hell corpus as well as the InScript corpus, that this knowledge helps us to significantly improve prediction performance on the narrative cloze task.

Structured Prediction with Output Embeddings for Semantic Image Annotation

Ariadna Quattoni, Arnau Ramisa, Pranava Swaroop Madhyastha, Edgar Simo-Serra, and Francesc Moreno-Noguer

We address the task of annotating images with semantic tuples. Solving this problem requires an algorithm able to deal with hundreds of classes for each argument of the tuple. In such contexts, data sparsity becomes a key challenge. We propose handling this by incorporating feature representations of both the inputs (images) and outputs (argument classes) into a factorized log-linear model.

Large-scale Multitask Learning for Machine Translation Quality Estimation

Kashif Shah and Lucia Specia

Multitask learning has been proven a useful technique in a number of Natural Language Processing applications where data is scarce and naturally diverse. Examples include learning from data of different domains and learning from annotations on translation quality provided by multiple annotators. *Tasks* in these scenarios would be the domains or the annotators. When faced with limited data for each task, a framework for the learning of tasks in parallel while using a shared representation is clearly helpful: what is learned for a given task can be transferred to other tasks while the particularities of each task are still modelled independently. Focusing on machine translation quality estimation as application, in this paper we show that multitask learning is also useful in cases where data is abundant. Based on two large-scale datasets, we explore models with multiple annotators and multiple languages and show that state-of-the-art multitask learning algorithms lead to improved results in all settings.

Conversational Markers of Constructive Discussions

Vlad Niculae and Cristian Danescu-Niculescu-Mizil

Group discussions are essential for organizing every aspect of modern life, from faculty meetings to senate debates, from grant review panels to papal conclaves. While costly in terms of time and organization effort, group discussions are commonly seen as a way of reaching better decisions compared to solutions that do not require coordination between the individuals (e.g. voting)—through discussion, the sum becomes greater than the parts. However, this assumption is not irrefutable: anecdotal evidence of wasteful discussions abounds, and in our own experiments we find that over 30% of discussions are unproductive. We propose a framework for analyzing conversational dynamics in order to determine whether a given task-oriented discussion is worth having or not. We exploit conversational patterns reflecting the flow of ideas and the balance between the participants, as well as their linguistic choices. We apply this framework to conversations naturally occurring in an online collaborative world exploration game developed and deployed to support this research. Using this setting, we show that linguistic cues and conversational patterns extracted from the first 20 seconds of a team discussion are predictive of whether it will be a wasteful or a productive one.

Vision and Feature Norms: Improving automatic feature norm learning through cross-modal maps

Luana Bulat, Douwe Kiela, and Stephen Clark

Property norms have the potential to aid a wide range of semantic tasks, provided that they can be obtained for large numbers of concepts. Recent work has focused on text as the main source of information for automatic property extraction. In this paper we examine property norm prediction from visual, rather than textual, data, using cross-modal maps learnt between property norm and visual spaces. We also investigate the importance of having a complete feature norm dataset, for both training and testing. Finally, we evaluate how these datasets and cross-modal maps can be used in an image retrieval task.

Cross-lingual Wikification Using Multilingual Embeddings

Chen-Tse Tsai and Dan Roth

Cross-lingual Wikification is the task of grounding mentions written in non-English documents to entries in the English Wikipedia. This task involves the problem of comparing textual clues across languages, which requires developing a notion of similarity between text snippets across languages. In this paper, we address this problem by jointly training multilingual embeddings for words and Wikipedia titles. The proposed method can be applied to all languages represented in Wikipedia, including those for which no machine translation technology is available. We create a challenging dataset in 12 languages and show that our proposed approach outperforms various baselines. Moreover, our model compares favorably with the best systems on the TAC KBP2015 Entity Linking task including those that relied on the availability of translation from the target language to English.

Deconstructing Complex Search Tasks: a Bayesian Nonparametric Approach for Extracting Sub-tasks

Rishabh Mehrotra, Prasanta Bhattacharya, and Emine Yilmaz

Search tasks, comprising a series of search queries serving a common informational need, have steadily emerged as accurate units for developing the next generation of task-aware web search systems. Most prior research in this area has focused on segmenting chronologically ordered search queries into higher level tasks. A more naturalistic viewpoint would involve treating query logs as convoluted structures of tasks-subtasks, with complex search tasks being decomposed into more focused sub-tasks. In this work, we focus on extracting sub-tasks from a given collection of on-task search queries. We jointly leverage insights from Bayesian non-parametrics and word embeddings to identify and extract sub-tasks from a given collection of *on-task* queries. Our proposed model can inform the design of the next generation of task-based search systems that leverage user's task behavior for better support and personalization.

System Demonstrations 1

Time: 6:00–8:00

Location: Pavilion

rstWeb - A Browser-based Annotation Interface for Rhetorical Structure Theory and Discourse Relations

Amir Zeldes

This paper presents rstWeb, a new browser-based interface for Rhetorical Structure Theory and other discourse relation annotations. Expanding on previous tools for RST, rstWeb allows annotators to work online using only a browser. Project administrators can easily collect multiple annotations of the same documents on a central server, keep track of annotation processes and assign tasks and annotation schemes to users. A local version using an embedded web framework is also available, running offline on a desktop browser under the localhost.

Instant Feedback for Increasing the Presence of Solutions in Peer Reviews

Huy Nguyen, Wenting Xiong, and Diane Litman

We present the design and evaluation of a web-based peer review system that uses natural language processing to automatically evaluate and provide instant feedback regarding the presence of solutions in peer reviews. Student reviewers can then choose to either revise their reviews to address the system's feedback, or ignore the feedback and submit their original reviews. A system deployment in multiple high school classrooms shows that our solution prediction model triggers instant feedback with high precision, and that the feedback is successful in increasing the number of peer reviews with solutions.

Farasa: A Fast and Furious Segmenter for Arabic*Ahmed Abdelali, Kareem Darwish, Nadir Durrani, and Hamdy Mubarak*

In this paper, we present Farasa, a fast and accurate Arabic segmenter. Our approach is based on SVM-rank using linear kernels. We measure the performance of the segmenter in terms of accuracy and efficiency, in two NLP tasks, namely Machine Translation (MT) and Information Retrieval (IR). Farasa outperforms or equalizes state-of-the-art Arabic segmenters (Stanford and MADAMIRA), while being more than one order of magnitude faster.

iAppraise: A Manual Machine Translation Evaluation Environment Supporting Eye-tracking*Ahmed Abdelali, Nadir Durrani, and Francisco Guzmán*

We present an extension to Appraise, an open source toolkit for human evaluation, to integrate eye-tracking functionality. Our goal is to analyze the human evaluation process based on the information collected from the gaze data. We modified the user interface (UI) for Appraise to communicate with an eye-tracking device, and collect the gaze data. Gaze signals captured by the eyetracker during the evaluation process, augmented with the word coordinates in the UI, results in gaze data. Such data i) invites researchers to propose models to predict human scores, ii) can serve to analyze evaluation process. To this end, we provide an analysis module, based on data collected during the evaluation session and the additional eye-tracking features.

Linguistica 5: Unsupervised Learning of Linguistic Structure*Jackson Lee and John Goldsmith*

This paper introduces Linguistica 5, a software for unsupervised learning of linguistic structure. It is a descendant of Goldsmith's (2001, 2006) Linguistica. Open-source and written in Python, the new Linguistica 5 is both a graphical user interface software and a Python library. While Linguistica 5 inherits its predecessors' strength in unsupervised learning of natural language morphology, it incorporates significant improvements in multiple ways. Notable new features include tools for data visualization as well as straightforward extensions for both its components and embedding in other programs.

TransRead: Designing a Bilingual Reading Experience with Machine Translation Technologies*François Yvon, Yong Xu, Marianna Apidianaki, Clément Pillias, and Pierre Cubaud*

In this paper, we use multilingual Natural Language Processing (NLP) tools to improve the reading experience of parallel texts on mobile devices. Such enterprise poses multiple challenging issues both from the NLP and from the Human Computer Interaction (HCI) perspectives. We discuss these problems and report on our own solutions, now implemented in a full-fledge bilingual reading device.

New Dimensions in Testimony Demonstration*Ron Artstein, Alesia Gainer, Kallirroi Georgila, Anton Leuski, Ari Shapiro, and David Traum*

New Dimensions in Testimony is a prototype dialogue system that allows users to conduct a conversation with a real person who is not available for conversation in real time. Users talk to a persistent representation of Holocaust survivor Pinchas Gutter on a screen, while a dialogue agent selects appropriate responses to user utterances from a set of pre-recorded video statements, simulating a live conversation. The technology is similar to existing conversational agents, but to our knowledge this is the first system to portray a real person. The demonstration will show the system on a range of screens (from mobile phones to large TVs), and allow users to have individual conversations with Mr. Gutter.

ArgRewrite: A Web-based Revision Assistant for Argumentative Writings*Fan Zhang, Rebecca Hwa, Diane Litman, and Homa B. Hashemi*

While intelligent writing assistants have become more common, they typically have little support for revision behavior. We present ArgRewrite, a novel web-based revision assistant that focus on rewriting analysis. The system supports two major functionalities: 1) to assist students as they revise, the system automatically extracts and analyzes revisions; 2) to assist teachers, the system provides an overview of students' revisions and allow teachers to correct the automatically analyzed results, ensuring that students get the correct feedback.

Scaling Up Word Clustering*Jon Dehdari, Liling Tan, and Josef van Genabith*

Word clusters improve performance in many NLP tasks including training neural network language models, but current increases in datasets are outpacing the ability of word clusterers to handle them. In this paper we

present a novel bidirectional, interpolated, refining, and alternating (BIRA) predictive exchange algorithm and introduce ClusterCat, a clusterer based on this algorithm. We show that ClusterCat is 3-85 times faster than four other well-known clusterers, while also improving upon the predictive exchange algorithm's perplexity by up to 18%. Notably, ClusterCat clusters a 2.5 billion token English News Crawl corpus in 3 hours. We also evaluate in a machine translation setting, resulting in shorter training times achieving the same translation quality measured in BLEU scores. ClusterCat is portable and freely available.

Task Completion Platform: A self-serve multi-domain goal oriented dialogue platform

Paul Crook, Alex Marin, Vipul Agarwal, Khushboo Aggarwal, Tasos Anastasakos, Ravi Bikkula, Daniel Boies, Asli Celikyilmaz, Senthilkumar Chandramohan, Zhaleh Feizollahi, Roman Holenstein, Minwoo Jeong, Omar Khan, Young-Bum Kim, Elizabeth Krawczyk, Xiaohu Liu, Danko Panic, Vasiliy Radostev, Nikhil Ramesh, Jean-Phillipe Robichaud, Alexandre Rochette, Logan Stromberg, and Ruhi Sarikaya

We demonstrate the Task Completion Platform (TCP); a multi-domain dialogue platform that can host and execute large numbers of goal-orientated dialogue tasks. The platform features a task configuration language, TaskForm, that allows the definition of each individual task to be decoupled from the overarching dialogue policy used by the platform to complete those tasks. This separation allows for simple and rapid authoring of new tasks, while dialogue policy and platform functionality evolve independent of the tasks. The current platform includes machine learnt models that provide contextual slot carry-over, flexible item selection, and task selection/switching. Any new task immediately gains the benefit of these pieces of built-in platform functionality. The platform is used to power many of the multi-turn dialogues supported by the Cortana personal assistant.

Student Workshop Posters 1

Time: 6:00–8:00

Location: Pavilion

An End-to-end Approach to Learning Semantic Frames with Feedforward Neural Network

Yukun Feng, Yipei Xu, and Dong Yu

We present an end-to-end method for learning verb-specific semantic frames with feedforward neural network (FNN). Previous works in this area mainly adopt a multi-step procedure including part-of-speech tagging, dependency parsing and so on. On the contrary, our method uses a FNN model that maps verb-specific sentences directly to semantic frames. The simple model gets good results on annotated data and has a good generalization ability. Finally we get 0.82 F-score on 63 verbs and 0.73 F-score on 407 verbs.

Analogy-based detection of morphological and semantic relations with word embeddings: what works and what doesn't.

Anna Gladkova, Aleksandr Drozd, and Satoshi Matsuoka

Following up on numerous reports of analogy-based identification of “linguistic regularities” in word embeddings, this study applies the widely used vector offset method to 4 types of linguistic relations: inflectional and derivational morphology, and lexicographic and encyclopedic semantics. We present a balanced test set with 99,200 questions in 40 categories, and we systematically examine how accuracy for different categories is affected by window size and dimensionality of the SVD-based word embeddings. We also show that GloVe and SVD yield similar patterns of results for different categories, offering further evidence for conceptual similarity between count-based and neural-net based models.

Argument Identification in Chinese Editorials

Marisa Chow

In this paper, we develop and evaluate several techniques for identifying argumentative paragraphs in Chinese editorials. We first use three methods of evaluation to score a paragraph’s argumentative nature: a relative word frequency approach; a method which targets known argumentative words in our corpus; and a combined approach which uses elements from the previous two. Then, we determine the best score thresholds for separating argumentative and non-argumentative paragraphs. The results of our experimentation show that our relative word frequency approach provides a reliable way to identify argumentative paragraphs, though the inaccuracies in scoring invite improvement through context-aware means.

Automatic tagging and retrieval of E-Commerce products based on visual features*Vasu Sharma and Harish Karnick*

This paper describes an automatic tag assignment algorithm for various e-commerce products where tag allotment is done solely based on the visual features in the image. It then builds a tag-based product retrieval system using the allotted tags. The explosive growth of e-commerce products being sold online has made manual annotation infeasible. Without such tagging it is very hard for customers to search and find these products. Hence a scalable approach catering to a large number of product images and allocating meaningful tags is essential. This will also eliminate the need for the laborious process of manually tagging such products. It can be the basis of an efficient tag-based product retrieval system. In this paper we propose one such approach based on feature extraction using Deep Convolutional Neural Networks to learn descriptive semantic features from product images. Then we use inverse distance weighted K-nearest neighbours classifiers together with several other multi-label classification approaches to assign appropriate tags to the images. We demonstrate the functioning of our algorithm for the Amazon product dataset for various categories of products like clothing and apparel, electronics, sports equipment etc.

Combining syntactic patterns and Wikipedia's hierarchy of hyperlinks to extract meronym relations*Debela Tesfaye Gemechu, Michael Zock, and Solomon Teferra*

We present here two methods for extraction of meronymic relation : (a) the first one relies solely on syntactic information. Unlike other approaches based on simple patterns, we determine their optimal combination to extract word pairs linked via a given semantic relation; (b) the second approach consists in combining syntactic patterns with the semantic information extracted from the Wikipedia hyperlink hierarchy (WHH) of the constituent words. By comparing our work with SemEval 2007 (Task 4 test set) and WordNet (WN) we found that our system clearly outperforms its competitors

Data-driven Paraphrasing and Stylistic Harmonization*Gerold Hintz*

Abstract This thesis proposal outlines the use of unsupervised data-driven methods for paraphrasing tasks. We motivate the development of knowledge-free methods at the guiding use case of multi-document summarization, which requires a domain-adaptable system for both the detection and generation of sentential paraphrases. First, we define a number of guiding research questions that will be addressed in the scope of this thesis. We continue to present ongoing work in unsupervised lexical substitution. An existing supervised approach is first adapted to a new language and dataset. We observe that supervised lexical substitution relies heavily on lexical semantic resources, and present an approach to overcome this dependency. We describe a method for unsupervised relation extraction, which we aim to leverage in lexical substitution as a replacement for knowledge-based resources.

Detecting “Smart” Spammers on Social Network: A Topic Model Approach*Linqing Liu, Yao Lu, Ye Luo, Renxian Zhang, Laurent Itti, and Jianwei Lu*

Spammer detection on social network is a challenging problem. The rigid anti-spam rules have resulted in emergence of “smart” spammers. They resemble legitimate users who are difficult to identify. In this paper, we present a novel spammer classification approach based on Latent Dirichlet Allocation (LDA), a topic model. Our approach extracts both the local and the global information of topic distribution patterns, which capture the essence of spamming. Tested on one benchmark dataset and one self-collected dataset, our proposed method outperforms other state-of-the-art methods in terms of averaged F1-score.

Developing language technology tools and resources for a resource-poor language: Sindhi
Raveesh Motlani

Sindhi is an Indo-Aryan language spoken by more than 75 million native speakers in the world. Nevertheless, Sindhi is a resource-poor language in terms of the availability of language technology tools and resources. In this thesis, we discuss the challenges faced and steps taken for creating raw and annotated datasets, constructing NLP tools such as a POS tagger, a morphological analyser, creating a transliteration system without parallel data in an unsupervised fashion and developing a SMT system for Sindhi-Urdu and adopting techniques for improving it.

4

Main Conference: Tuesday, June 14

Overview

7:30–8:45	Breakfast Session 4	Pavilion
9:00–10:30	Semantic Parsing <i>Grande Ballroom A</i>	Morphology and Phonology <i>Grande Ballroom B</i>
		Various <i>Grande Ballroom C</i>
10:30–11:00	Coffee break Session 5	Grande Foyer
11:00–12:30	Generation <i>Grande Ballroom A</i>	Sentiment <i>Grande Ballroom B</i>
		Knowledge Acquisition <i>Grande Ballroom C</i>
12:30–1:15	Lunch	
1:15–2:15	Panel Discussion: How Will Deep Learning Change Computational Linguistics? Session 6	Grande Ballroom
2:30–3:30	Machine Translation II <i>Grande Ballroom A</i>	Relation Extraction <i>Grande Ballroom B</i>
		Semantic Similarity <i>Grande Ballroom C</i>
3:30–4:00	Break Session 7	Grande Foyer
4:00–5:00	Machine Translation III <i>Grande Ballroom A</i>	Anaphora Resolution <i>Grande Ballroom B</i>
		Word Embeddings I <i>Grande Ballroom C</i>
5:15–6:00	One Minute Madness	Grande Ballroom
6:00–8:00	Poster Session 2	Pavilion
6:00–8:00	System Demonstrations 2	Pavilion
6:00–8:00	Student Workshop Posters 2	Pavilion
8:00–11:00	Social Event	Bay View Lawn

Session 4 Overview – Tuesday, June 14, 2016

	Track A	Track B	Track C
	<i>Semantic Parsing</i> Grande Ballroom A	<i>Morphology and Phonology</i> Grande Ballroom B	<i>Various</i> Grande Ballroom C
9:00	[TACL] Transforming Dependency Structures to Logical Forms for Semantic Parsing <i>Reddy, Täckström, Collins, Kwiatkowski, Das, Steedman, and Lapata</i>	Weighting Finite-State Transductions With Neural Context <i>Rastogi, Cotterell, and Eisner</i>	Syntactic Parsing of Web Queries with Question Intent <i>Pinter, Reichart, and Szpektor</i>
9:20	[TACL] Imitation Learning of Agenda-based Semantic Parsers <i>Berant and Liang</i>	Morphological Inflection Generation Using Character Sequence to Sequence Learning <i>Faruqui, Tsvetkov, Neubig, and Dyer</i>	Visualizing and Understanding Neural Models in NLP <i>Li, Chen, Hovy, and Jurafsky</i>
9:40	Probabilistic Models for Learning a Semantic Parser Lexicon <i>Krishnamurthy</i>	Towards Unsupervised and Language-independent Compound Splitting using Inflectional Morphological Transformations <i>Ziering and Plas</i>	Bilingual Word Embeddings from Parallel and Non-parallel Corpora for Cross-Language Text Classification <i>Mogadala and Rettinger</i>
10:00	[TACL] Semantic Parsing of Ambiguous Input through Paraphrasing and Verification <i>Arthur, Neubig, Sakti, Toda, and Nakamura</i>	Phonological Understanding <i>Jaech, Koncel-Kedziorski, and Ostendorf</i>	Joint Learning with Global Inference for Comment Classification in Community Question Answering <i>Joty, Márquez, and Nakov</i>
10:20	Unsupervised Compound Splitting With Distributional Semantics Rivals Supervised Methods <i>Riedl and Biemann</i>	A Joint Model of Orthography and Morphological Segmentation [BEST SHORT PAPER RUNNER-UP] <i>Cotterell, Vieira, and Schütze</i>	Weak Semi-Markov CRFs for Noun Phrase Chunking in Informal Text <i>Muis and Lu</i>

Parallel Session 4

Parallel Session 4A: Semantic Parsing

Grande Ballroom A

Chair: Mike Lewis

[TACL] Transforming Dependency Structures to Logical Forms for Semantic Parsing

Siva Reddy, Oscar Täckström, Michael Collins, Tom Kwiatkowski, Dipanjan Das, Mark Steedman, and Mirella Lapata 09:00–09:20

The strongly typed syntax of grammar formalisms such as CCG, TAG, LFG and HPSG offers a synchronous framework for deriving syntactic structures and semantic logical forms. In contrast—partly due to the lack of a strong type system—dependency structures are easy to annotate and have become a widely used form of syntactic analysis for many languages. However, the lack of a type system makes a formal mechanism for deriving logical forms from dependency structures challenging. We address this by introducing a robust system based on the lambda calculus for deriving neo-Davidsonian logical forms from dependency trees. These logical forms are then used for semantic parsing of natural language to Freebase. Experiments on the Free917 and WebQuestions datasets show that our representation is superior to the original dependency trees and that it outperforms a CCG-based representation on this task. Compared to prior work, we obtain the strongest result to date on Free917 and competitive results on WebQuestions.

[TACL] Imitation Learning of Agenda-based Semantic Parsers

Jonathan Berant and Percy Liang 09:20–09:40

Semantic parsers conventionally construct logical forms bottom-up in a fixed order, resulting in the generation of many extraneous partial logical forms. In this paper, we combine ideas from imitation learning and agenda-based parsing to train a semantic parser that searches partial logical forms in a more strategic order. Empirically, our parser reduces the number of constructed partial logical forms by an order of magnitude, and obtains a 6x-9x speedup over fixed-order parsing, while maintaining comparable accuracy.

Probabilistic Models for Learning a Semantic Parser Lexicon

Jayant Krishnamurthy 09:40–10:00

We introduce several probabilistic models for learning the lexicon of a semantic parser. Lexicon learning is the first step of training a semantic parser for a new application domain and the quality of the learned lexicon significantly impacts both the accuracy and efficiency of the final semantic parser. Existing work on lexicon learning has focused on heuristic methods that lack convergence guarantees and require significant human input in the form of lexicon templates or annotated logical forms. In contrast, our probabilistic models are trained directly from question/answer pairs using EM and our simplest model has a concave objective that guarantees convergence to a global optimum. An experimental evaluation on a set of 4th grade science questions demonstrates that our models improve semantic parser accuracy (35–70% error reduction) and efficiency (4–25x more sentences per second) relative to prior work despite using less human input. Our models also obtain competitive results on Geo880 without any dataset-specific engineering.

[TACL] Semantic Parsing of Ambiguous Input through Paraphrasing and Verification

Philip Arthur, Graham Neubig, Sakriani Sakti, Tomoki Toda, and Satoshi Nakamura 10:00–10:20

We propose a new method for semantic parsing of ambiguous and ungrammatical input, such as search queries. We do so by building on an existing semantic parsing framework that uses synchronous context free grammars (SCFG) to jointly model the input sentence and output meaning representation. We generalize this SCFG framework to allow not one, but multiple outputs. Using this formalism, we construct a grammar that takes an ambiguous input string and jointly maps it into both a meaning representation and a natural language paraphrase that is less ambiguous than the original input. This paraphrase can be used to disambiguate the meaning representation via verification using a language model that calculates the probability of each paraphrase.

Unsupervised Compound Splitting With Distributional Semantics Rivals Supervised Methods

Martin Riedl and Chris Biemann 10:20–10:30

In this paper we present a word decompounding method that is based on distributional semantics. Our method does not require any linguistic knowledge and is initialized using a large monolingual corpus. The core idea of our approach is that parts of single-word compounds (like “candle” and “stick”) are semantically similar to the entire compound, which helps to exclude spurious splits (like “candles” “tick”). We report results for

German and Dutch: For German, our unsupervised method comes on par with the performance of a rule-based and a supervised method and outperforms two unsupervised baselines. For Dutch, our method performs only slightly below a rule-based optimized compound splitter.

Parallel Session 4B: Morphology and Phonology

Grande Ballroom B

Chair: Dilek Hakkani-Tur

Weighting Finite-State Transductions With Neural Context*Pushpendre Rastogi, Ryan Cotterell, and Jason Eisner*

09:00–09:20

How should one apply deep learning to tasks such as morphological reinflection, which stochastically edit one string to get another? A recent approach to such sequence-to-sequence tasks is to compress the input string into a vector that is then used to generate the output string, using recurrent neural networks. In contrast, we propose to keep the traditional architecture, which uses a finite-state transducer to score all possible output strings, but to augment the scoring function with the help of recurrent networks. A stack of bidirectional LSTMs reads the input string from left-to-right and right-to-left, in order to summarize the input context in which a transducer arc is applied. We combine these learned features with the transducer to define a probability distribution over aligned output strings, in the form of a weighted finite-state automaton. This reduces hand-engineering of features, allows learned features to examine unbounded context in the input string, and still permits exact inference through dynamic programming. We illustrate our method on the tasks of morphological reinflection and lemmatization.

Morphological Inflection Generation Using Character Sequence to Sequence Learning*Manaal Faruqui, Yulia Tsvetkov, Graham Neubig, and Chris Dyer*

09:20–09:40

Morphological inflection generation is the task of generating the inflected form of a given lemma corresponding to a particular linguistic transformation. We model the problem of inflection generation as a character sequence to sequence learning problem and present a variant of the neural encoder-decoder model for solving it. Our model is language independent and can be trained in both supervised and semi-supervised settings. We evaluate our system on seven datasets of morphologically rich languages and achieve either better or comparable results to existing state-of-the-art models of inflection generation.

Towards Unsupervised and Language-independent Compound Splitting using Inflectional Morphological Transformations*Patrick Ziering and Lonneke van der Plas*

09:40–10:00

In this paper, we address the task of language-independent, knowledge-lean and unsupervised compound splitting, which is an essential component for many natural language processing tasks such as machine translation. Previous methods on statistical compound splitting either include language-specific knowledge (e.g., linking elements) or rely on parallel data, which results in limited applicability. We aim to overcome these limitations by learning compounding morphology from inflectional information derived from lemmatized monolingual corpora. In experiments for Germanic languages, we show that our approach significantly outperforms language-dependent state-of-the-art methods in finding the correct split point and that word inflection is a good approximation for compounding morphology.

Phonological Pun-derstanding*Aaron Jaech, Rik Koncel-Kedziorski, and Mari Ostendorf*

10:00–10:20

Paronomastic puns create humor through the relationship between a pun and its phonologically similar target. For example, in “Don’t take geologists for granite” the word “granite” is a pun with the target “granted”. The recovery of the target in the mind of the listener is essential to the success of the pun. This work introduces a new model for automatic target recovery and provides the first empirical test for this task. The model draws upon techniques for automatic speech recognition using weighted finite-state transducers, and leverages automatically learned phone edit probabilities that give insight into how people perceive sounds and into what makes a good pun. The model is evaluated on a small corpus where it is able to automatically recover a large fraction of the pun targets.

A Joint Model of Orthography and Morphological Segmentation [BEST SHORT PAPER RUNNER-UP]*Ryan Cotterell, Tim Vieira, and Hinrich Schütze*

10:20–10:30

We present a model of morphological segmentation that jointly learns to segment and restore orthographic changes, e.g., funniest to fun-y-est. We term this form of analysis canonical segmentation and contrast it with the traditional surface segmentation, which segments a surface form into a sequence of substrings, e.g., funniest to funn-i-est. We derive an importance sampling algorithm for approximate inference in the model and report experimental results on English, German and Indonesian.

Parallel Session 4C: Various

Grande Ballroom C

Chair: Jinho D. Choi

Syntactic Parsing of Web Queries with Question Intent

Yuval Pinter, Roi Reichart, and Idan Szpektor

09:00–09:20

Accurate automatic processing of Web queries is important for high-quality information retrieval from the Web. While the syntactic structure of a large portion of these queries is trivial, the structure of queries with question intent is much richer. In this paper we therefore address the task of statistical syntactic parsing of such queries. We first show that the standard dependency grammar does not account for the full range of syntactic structures manifested by queries with question intent. To alleviate this issue we extend the dependency grammar to account for segments - independent syntactic units within a potentially larger syntactic structure. We then propose two distant supervision approaches for the task. Both algorithms do not require manually parsed queries for training. Instead, they are trained on millions of (query, page title) pairs from the Community Question Answering (CQA) domain, where the CQA page was clicked by the user who initiated the query in a search engine. Experiments on a new treebank consisting of 5,000 Web queries from the CQA domain, manually parsed using the proposed grammar, show that our algorithms outperform alternative approaches trained on various sources: tens of thousands of manually parsed OntoNotes sentences, millions of unlabeled CQA queries and thousands of manually segmented CQA queries.

Visualizing and Understanding Neural Models in NLP

Jiwei Li, Xinlei Chen, Eduard Hovy, and Dan Jurafsky

09:20–09:40

While neural networks have been successfully applied to many NLP tasks the resulting vector-based models are very difficult to interpret. For example it's not clear how they achieve *compositionality*, building sentence meaning from the meanings of words and phrases. In this paper we describe strategies for visualizing compositionality in neural models for NLP, inspired by similar work in computer vision. We first plot unit values to visualize compositionality of negation, intensification, and concessive clauses, allowing us to see well-known markedness asymmetries in negation. We then introduce methods for visualizing a unit's *salience*, the amount that it contributes to the final composed meaning from first-order derivatives. Our general-purpose methods may have wide applications for understanding compositionality and other semantic properties of deep networks.

Bilingual Word Embeddings from Parallel and Non-parallel Corpora for Cross-Language Text Classification

Aditya Mogadala and Achim Rettinger

09:40–10:00

In many languages, sparse availability of resources causes numerous challenges for textual analysis tasks. Text classification is one of such standard tasks that is hindered due to limited availability of label information in low-resource languages. Transferring knowledge (i.e. label information) from high-resource to low-resource languages might improve text classification as compared to the other approaches like machine translation. We introduce BRAVE (Bilingual paRAGraph VEctors), a model to learn bilingual distributed representations (i.e. embeddings) of words without word alignments either from sentence-aligned parallel or label-aligned non-parallel document corpora to support cross-language text classification. Empirical analysis shows that classification models trained with our bilingual embeddings outperforms other state-of-the-art systems on three different cross-language text classification tasks.

Joint Learning with Global Inference for Comment Classification in Community Question Answering

Shafiq Joty, Lluís Márquez, and Preslav Nakov

10:00–10:20

This paper addresses the problem of comment classification in community Question Answering. Following the state of the art, we approach the task with a global inference process to exploit the information of all comments in the answer-thread in the form of a fully connected graph. Our contribution comprises two novel joint learning models that are online and integrate inference within learning. The first one jointly learns two node- and edge-level MaxEnt classifiers with stochastic gradient descent and integrates the inference step with loopy belief propagation. The second model is an instance of fully connected pairwise CRFs (FCCRF). The FCCRF model significantly outperforms all other approaches and yields the best results on the task to date. Crucial elements for its success are the global normalization and an Ising-like edge potential.

Weak Semi-Markov CRFs for Noun Phrase Chunking in Informal Text

Aldrian Obaja Muis and Wei Lu

10:20–10:30

This paper introduces a new annotated corpus based on an existing informal text corpus: the NUS SMS Corpus (Chen and Kan, 2013). The new corpus includes 76,490 noun phrases from 26,500 SMS messages, annotated by university students. We then explored several graphical models, including a novel variant of the semi-Markov conditional random fields (semi-CRF) for the task of noun phrase chunking. We demonstrated through empirical evaluations on the new dataset that the new variant yielded similar accuracy but ran in significantly lower running time compared to the conventional semi-CRF.

Session 5 Overview – Tuesday, June 14, 2016

	Track A	Track B	Track C
	<i>Generation</i> Grande Ballroom A	<i>Sentiment</i> Grande Ballroom B	<i>Knowledge Acquisition</i> Grande Ballroom C
11:00	What to talk about and how? Selective Generation using LSTMs with Coarse-to-Fine Alignment <i>Mei, Bansal, and Walter</i>	Ultradense Word Embed- dings by Orthogonal Trans- formation <i>Rothe, Ebert, and Schütze</i>	[TACL] Concept Ground- ing to Multiple Knowledge Bases via Indirect Supervision <i>Tsai and Roth</i>
11:20	Generation from Abstract Meaning Representation using Tree Transducers <i>Flanigan, Dyer, Smith, and Carbonell</i>	Separating Actor-View from Speaker-View Opinion Ex- pressions using Linguistic Features <i>Wiegand, Schulder, and Ruppenhofer</i>	Mapping Verbs in Different Languages to Knowledge Base Relations using Web Text as Interlingua <i>Wijaya and Mitchell</i>
11:40	A Corpus and Semantic Parser for Multilingual Nat- ural Language Querying of OpenStreetMap <i>Haas and Riezler</i>	Clustering for Simultaneous Extraction of Aspects and Features from Reviews <i>Chen, Martineau, Cheng, and Sheth</i>	Comparing Convolutional Neural Networks to Tradition- al Models for Slot Filling <i>Adel, Roth, and Schütze</i>
12:00	Natural Language Commu- nication with Robots <i>Bisk, Yuret, and Marcu</i>	Opinion Holder and Tar- get Extraction on Opinion Compounds – A Linguistic Approach <i>Wiegand, Bocionek, and Ruppenhofer</i>	A Corpus and Cloze Eval- uation for Deeper Under- standing of Commonsense Stories <i>Mostafazadeh, Chambers, He, Parikh, Batra, Vander- wende, Kohli, and Allen</i>
12:20	Inter-document Contextual Language model <i>Tran, Zukerman, and Haffari</i>	Capturing Reliable Fine- Grained Sentiment Associa- tions by Crowdsourcing and Best—Worst Scaling <i>Kiritchenko and Mohammad</i>	Dynamic Entity Represen- tation with Max-pooling Improves Machine Reading <i>Kobayashi, Tian, Okazaki, and Inui</i>

Parallel Session 5

Parallel Session 5A: Generation

Grande Ballroom A

Chair: Lu Wang

What to talk about and how? Selective Generation using LSTMs with Coarse-to-Fine Alignment

Hongyuan Mei, Mohit Bansal, and Matthew R. Walter

11:00–11:20

We propose an end-to-end, domain-independent neural encoder-aligner-decoder model for selective generation, i.e., the joint task of content selection and surface realization. Our model first encodes a full set of over-determined database event records via an LSTM-based recurrent neural network, then utilizes a novel coarse-to-fine aligner to identify the small subset of salient records to talk about, and finally employs a decoder to generate free-form descriptions of the aligned, selected records. Our model achieves the best selection and generation results reported to-date (with 59% relative improvement in generation) on the benchmark WeatherGov dataset, despite using no specialized features or linguistic resources. Using an improved k-nearest neighbor beam filter helps further. We also perform a series of ablations and visualizations to elucidate the contributions of our key model components. Lastly, we evaluate the generalizability of our model on the RoboCup dataset, and get results that are competitive with or better than the state-of-the-art, despite being severely data-starved.

Generation from Abstract Meaning Representation using Tree Transducers

Jeffrey Flanigan, Chris Dyer, Noah A. Smith, and Jaime Carbonell

11:20–11:40

Language generation from purely semantic representations is a challenging task. This paper addresses generating English from the Abstract Meaning Representation (AMR), consisting of re-entrant graphs whose nodes are concepts and edges are relations. The new method is trained statistically from AMR-annotated English and consists of two major steps: (i) generating an appropriate spanning tree for the AMR, and (ii) applying tree-to-string transducers to generate English. The method relies on discriminative learning and an argument realization model to overcome data sparsity. Initial tests on held-out data show good promise despite the complexity of the task. The system is available open-source as part of JAMR at: <http://github.com/jflanigan/jamr>

A Corpus and Semantic Parser for Multilingual Natural Language Querying of OpenStreetMap

Carolin Haas and Stefan Riezler

11:40–12:00

We present a corpus of 2,380 natural language queries paired with machine readable formulae that can be executed against world wide geographic data of the OpenStreetMap (OSM) database. We use the corpus to learn an accurate semantic parser that builds the basis of a natural language interface to OSM. Furthermore, we use response-based learning on parser feedback to adapt a statistical machine translation system for multilingual database access to OSM. Our framework allows to map fuzzy natural language expressions such as “nearby”, “north of”, or “in walking distance” to spatial polygons on an interactive map. Furthermore, it combines syntactic complexity and compositionality with a reasonable lexical variability of queries, making it an interesting new publicly available dataset for research on semantic parsing.

Natural Language Communication with Robots

Yonatan Bisk, Deniz Yuret, and Daniel Marcu

12:00–12:20

We propose a framework for devising empirically testable algorithms for bridging the communication gap between humans and robots. We instantiate our framework in the context of a problem setting in which humans give instructions to robots using unrestricted natural language commands, with instruction sequences being subservient to building complex goal configurations in a blocks world. We show how one can collect meaningful training data and we propose three neural architectures for interpreting contextually grounded natural language commands. The proposed architectures allow us to correctly understand/ground the blocks that the robot should move when instructed by a human who uses unrestricted language. The architectures have more difficulty in correctly understanding/grounding the spatial relations required to place blocks correctly, especially when the blocks are not easily identifiable.

Inter-document Contextual Language model

Quan Hung Tran, Ingrid Zukerman, and Gholamreza Haffari

12:20–12:30

In this paper, we examine the impact of employing contextual, structural information from a tree-structured document set to derive a language model. Our results show that this information significantly improves the accuracy of the resultant model.

Parallel Session 5B: Sentiment

Grande Ballroom B

Chair: Ellen Riloff

Ultradense Word Embeddings by Orthogonal Transformation*Sascha Rothe, Sebastian Ebert, and Hinrich Schütze*

11:00–11:20

Embeddings are generic representations that are useful for many NLP tasks. In this paper, we introduce DENSIFIER, a method that learns an orthogonal transformation of the embedding space that focuses the information relevant for a task in an ultradense subspace of a dimensionality that is smaller by a factor of 100 than the original space. We show that ultradense embeddings generated by DENSIFIER reach state of the art on a lexicon creation task in which words are annotated with three types of lexical information – sentiment, concreteness and frequency. On the SemEval2015 10B sentiment analysis task we show that no information is lost when the ultradense subspace is used, but training is an order of magnitude more efficient due to the compactness of the ultradense space.

Separating Actor-View from Speaker-View Opinion Expressions using Linguistic Features*Michael Wiegand, Marc Schulder, and Josef Ruppenhofer*

11:20–11:40

We examine different features and classifiers for the categorization of opinion words into actor and speaker view. To our knowledge, this is the first comprehensive work to address sentiment views on the word level taking into consideration opinion verbs, nouns and adjectives. We consider many high-level features requiring only few labeled training data. A detailed feature analysis produces linguistic insights into the nature of sentiment views. We also examine in how far global constraints between different opinion words help to increase classification performance. Finally, we show that our (prior) word-level annotation correlates with contextual sentiment views.

Clustering for Simultaneous Extraction of Aspects and Features from Reviews*Lu Chen, Justin Martineau, Doreen Cheng, and Amit Sheth*

11:40–12:00

This paper presents a clustering approach that simultaneously identifies product features and groups them into aspect categories from online reviews. Unlike prior approaches that first extract features and then group them into categories, the proposed approach combines feature and aspect discovery instead of chaining them. In addition, prior work on feature extraction tends to require seed terms and focus on identifying explicit features, while the proposed approach extracts both explicit and implicit features, and does not require seed terms. We evaluate this approach on reviews from three domains. The results show that it outperforms several state-of-the-art methods on both tasks across all three domains.

Opinion Holder and Target Extraction on Opinion Compounds – A Linguistic Approach*Michael Wiegand, Christine Bocionek, and Josef Ruppenhofer*

12:00–12:20

We present an approach to the new task of opinion holder and target extraction on opinion compounds. Opinion compounds (e.g. “user rating” or “victim support”) are noun compounds whose head is an opinion noun. We do not only examine features known to be effective for noun compound analysis, such as paraphrases and semantic classes of heads and modifiers, but also propose novel features tailored to this new task. Among them, we examine paraphrases that jointly consider holders and targets, a verb detour in which noun heads are replaced by related verbs, a global head constraint allowing inferencing between different compounds, and the categorization of the sentiment view that the head conveys.

Capturing Reliable Fine-Grained Sentiment Associations by Crowdsourcing and Best—Worst Scaling*Svetlana Kiritchenko and Saif M. Mohammad*

12:20–12:30

Access to word—sentiment associations is useful for many applications, including sentiment analysis, stance detection, and linguistic analysis. However, manually assigning fine-grained sentiment association scores to words has many challenges with respect to keeping annotations consistent. We apply the annotation technique of Best—Worst Scaling to obtain real-valued sentiment association scores for words and phrases in three different domains: general English, English Twitter, and Arabic Twitter. We show that on all three domains the ranking of words by sentiment remains remarkably consistent even when the annotation process is repeated with a different set of annotators. We also, for the first time, determine the minimum difference in sentiment association that is perceptible to native speakers of a language.

Parallel Session 5C: Knowledge Acquisition

Grande Ballroom C

Chair: Ray Mooney

[TACL] Concept Grounding to Multiple Knowledge Bases via Indirect Supervision

Chen-Tse Tsai and Dan Roth

11:00–11:20

We consider the problem of disambiguating concept mentions appearing in documents and grounding them in multiple knowledge bases, where each knowledge base addresses some aspects of the domain. This problem poses a few additional challenges beyond those addressed in the popular Wikification problem. Key among them is that most knowledge bases do not contain the rich textual and structural information Wikipedia does; consequently, the main supervision signal used to train Wikification rankers does not exist anymore. In this work we develop an algorithmic approach that, by carefully examining the relations between various related knowledge bases, generates an indirect supervision signal it uses to train a ranking model that accurately chooses knowledge base entries for a given mention; moreover, it also induces prior knowledge that can be used to support a global coherent mapping of all the concepts in a given document to the knowledge bases. Using the biomedical domain as our application, we show that our indirectly supervised ranking model outperforms other unsupervised baselines and that the quality of this indirect supervision scheme is very close to a supervised model. We also show that considering multiple knowledge bases together has an advantage over grounding concepts to each knowledge base individually.

Mapping Verbs in Different Languages to Knowledge Base Relations using Web Text as Interlingua

Derry Tanti Wijaya and Tom M. Mitchell

11:20–11:40

In recent years there have been many knowledge bases (KBs) constructed yet there is not yet a verb resource that maps to these growing KB resources. A resource that contains verbs in different languages to identify KB relations will be useful for extracting more facts to the KBs from any language text and to aid alignment and integration of knowledge across different KBs and languages. Such multi-lingual verb resource can also be useful for tasks such as machine reading and machine translation. In this paper, we present a scalable approach to automatically construct such verb resource using a very large web text corpus as a kind of interlingua to relate verb phrases to KB relations. Given a text corpus in any language and any KB, it can produce a mapping of that language's verb phrases to the KB relations. Experiments with English NELL KB and ClueWeb corpus show that the learned English verb-to-relation mapping is effective for extracting relation instances from English text. Using the same method on a Portuguese NELL KB and a Portuguese text corpus, we are able to construct automatically a verb resource in Portuguese that is effective for extracting relation instances from Portuguese text.

Comparing Convolutional Neural Networks to Traditional Models for Slot Filling

Heike Adel, Benjamin Roth, and Hinrich Schütze

11:40–12:00

We address relation classification in the context of slot filling, the task of finding and evaluating fillers like “Steve Jobs” for the slot X in “X founded Apple”. We propose a convolutional neural network which splits the input sentence into three parts according to the relation arguments and compare it to state-of-the-art and traditional approaches of relation classification. Finally, we combine different methods and show that the combination is better than individual approaches. We also analyze the effect of genre differences on performance.

A Corpus and Cloze Evaluation for Deeper Understanding of Commonsense Stories

Nasrin Mostafazadeh, Nathanael Chambers, Xiaodong He, Devi Parikh, Dhruv Batra, Lucy Vanderwende, Pushmeet Kohli, and James Allen

12:00–12:20

Representation and learning of commonsense knowledge is one of the foundational problems in the quest to enable deep language understanding. This issue is particularly challenging for understanding causal and correlational relationships between events. While this topic has received a lot of interest in the NLP community, research has been hindered by the lack of a proper evaluation framework. This paper attempts to address this problem with a new framework for evaluating story understanding and script learning: the ‘Story Cloze Test’. This test requires a system to choose the correct ending to a four-sentence story. We created a new corpus of ~50k five-sentence commonsense stories, ROC-Short-Stories, to enable this evaluation. This corpus is unique in two ways: (1) it captures a rich set of causal and temporal commonsense relations between daily events, and (2) it is a high quality collection of everyday life stories that can also be used for story generation. Experimental evaluation shows that a host of baselines and state-of-the-art models based on shallow language understanding struggle to achieve a high score on the Story Cloze Test. We discuss these implications for script and story

learning, and offer suggestions for deeper language understanding.

Dynamic Entity Representation with Max-pooling Improves Machine Reading

Sosuke Kobayashi, Ran Tian, Naoki Okazaki, and Kentaro Inui

12:20–12:30

We propose a novel neural network model for machine reading, DER Network, which explicitly implements a reader building dynamic meaning representations for entities by gathering and accumulating information around the entities as it reads a document. Evaluated on a recent large scale dataset (Hermann et al., 2015), our model exhibits better results than previous research, and we find that max-pooling is suited for modeling the accumulation of information on entities. Further analysis suggests that our model can put together multiple pieces of information encoded in different sentences to answer complicated questions.

Session 6 Overview – Tuesday, June 14, 2016

	Track A	Track B	Track C
	<i>Machine Translation II</i> Grande Ballroom A	<i>Relation Extraction</i> Grande Ballroom B	<i>Semantic Similarity</i> Grande Ballroom C
2:30	Speed-Constrained Tuning for Statistical Machine Translation Using Bayesian Optimization <i>Beck, Gispert, Iglesias, Waite, and Byrne</i>	Multilingual Relation Extraction using Compositional Universal Schema <i>Verga, Belanger, Strubell, Roth, and McCallum</i>	DAG-Structured Long Short-Term Memory for Semantic Compositionality <i>Zhu, Sobhani, and Guo</i>
2:50	Multi-Way, Multilingual Neural Machine Translation with a Shared Attention Mechanism [BEST LONG PAPER RUNNER-UP] <i>Firat, Cho, and Bengio</i>	Effective Crowd Annotation for Relation Extraction <i>Liu, Soderland, Bragg, Lin, Ling, and Weld</i>	Bayesian Supervised Domain Adaptation for Short Text Similarity <i>Sultan, Boyd-Graber, and Sumner</i>
3:10	Incorporating Structural Alignment Biases into an Attentional Neural Translation Model <i>Cohn, Hoang, Vylomova, Yao, Dyer, and Haffari</i>	A Translation-Based Knowledge Graph Embedding Preserving Logical Property of Relations <i>Yoon, Song, Park, and Park</i>	Pairwise Word Interaction Modeling with Deep Neural Networks for Semantic Similarity Measurement <i>He and Lin</i>

Parallel Session 6

Parallel Session 6A: Machine Translation II

Grande Ballroom A

Chair: Colin Cherry

Speed-Constrained Tuning for Statistical Machine Translation Using Bayesian Optimization

Daniel Beck, Adrià de Gispert, Gonzalo Iglesias, Aurelien Waite, and Bill Byrne 14:30–14:50

We address the problem of automatically finding the parameters of a statistical machine translation system that maximize BLEU scores while ensuring that decoding speed exceeds a minimum value. We propose the use of Bayesian Optimization to efficiently tune the speed-related decoding parameters by easily incorporating speed as a noisy constraint function. The obtained parameter values are guaranteed to satisfy the speed constraint with an associated confidence margin. Across three language pairs and two speed constraint values, we report overall optimization time reduction compared to grid and random search. We also show that Bayesian Optimization can decouple speed and BLEU measurements, resulting in a further reduction of overall optimization time as speed is measured over a small subset of sentences.

Multi-Way, Multilingual Neural Machine Translation with a Shared Attention Mechanism [BEST LÖNG PAPER RUNNER-UP]

Orhan Firat, Kyunghyun Cho, and Yoshua Bengio 14:50–15:10

We propose multi-way, multilingual neural machine translation. The proposed approach enables a single neural translation model to translate between multiple languages, with a number of parameters that grows only linearly with the number of languages. This is made possible by having a single attention mechanism that is shared across all language pairs. We train the proposed multi-way, multilingual model on ten language pairs from WMT'15 simultaneously and observe clear performance improvements over models trained on only one language pair. In particular, we observe that the proposed model significantly improves the translation quality of low-resource language pairs.

Incorporating Structural Alignment Biases into an Attentional Neural Translation Model

Trevor Cohn, Cong Duy Vu Hoang, Ekaterina Vylomova, Kaisheng Yao, Chris Dyer, and Gholaemreza Haffari 15:10–15:30

Neural encoder-decoder models of machine translation have achieved impressive results, rivalling traditional translation models. However their modelling formulation is overly simplistic, and omits several key inductive biases built into traditional models. In this paper we extend the attentional neural translation model to include structural biases from word based alignment models, including positional bias, Markov conditioning, fertility and agreement over translation directions. We show improvements over a baseline attentional model and standard phrase-based model over several language pairs, evaluating on difficult languages in a low resource setting.

Parallel Session 6B: Relation Extraction

Grande Ballroom B

Chair: Byron C. Wallace

Multilingual Relation Extraction using Compositional Universal Schema

Patrick Verga, David Belanger, Emma Strubell, Benjamin Roth, and Andrew McCallum 14:30–14:50

Universal schema builds a knowledge base (KB) of entities and relations by jointly embedding all relation types from input KBs as well as textual patterns observed in raw text. In most previous applications of universal schema, each textual pattern is represented as a single embedding, preventing generalization to unseen patterns. Recent work employs a neural network to capture patterns' compositional semantics, providing generalization to all possible input text. In response, this paper introduces significant further improvements to the coverage and flexibility of universal schema relation extraction: predictions for entities unseen in training and multilingual transfer learning to domains with no annotation. We evaluate our model through extensive experiments on the English and Spanish TAC KBP benchmark, outperforming the top system from TAC 2013 slot-filling using no handwritten patterns or additional annotation. We also consider a multilingual setting in which English training data entities overlap with the seed KB, but Spanish text does not. Despite having no annotation for Spanish data, we train an accurate predictor, with additional improvements obtained by tying word embeddings across languages. Furthermore, we find that multilingual training improves English relation extraction accuracy. Our approach is thus suited to broad-coverage automated knowledge base construction in a variety of languages and domains.

Effective Crowd Annotation for Relation Extraction

Angli Liu, Stephen Soderland, Jonathan Bragg, Christopher H. Lin, Xiao Ling, and Daniel S. Weld 14:50–15:10

Can crowdsourced annotation of training data boost performance for relation extraction over methods based solely on distant supervision? While crowdsourcing has been shown effective for many NLP tasks, previous researchers found only minimal improvement when applying the method to relation extraction. This paper demonstrates that a much larger boost is possible, e.g., raising F1 from 0.40 to 0.60. Furthermore, the gains are due to a simple, generalizable technique, Gated Instruction, which combines an interactive tutorial, feedback to correct errors during training, and improved screening.

A Translation-Based Knowledge Graph Embedding Preserving Logical Property of Relations

Hee-Geun Yoon, Hyun-Je Song, Seong-Bae Park, and Se-Young Park 15:10–15:30

This paper proposes a novel translation-based knowledge graph embedding that preserves the logical properties of relations such as transitivity and symmetricity. The embedding space generated by existing translation-based embeddings do not represent transitive and symmetric relations, because they ignore the role of entities in triples. Thus, we introduce a role-specific projection which maps an entity to distinct vectors according to its role in a triple. That is, a head entity is projected onto an embedding space by a head projection operator, and a tail entity is projected by a tail projection operator. This idea is applied to TransE, TransR, and TransD to produce lppTransE, lppTransR, and lppTransD, respectively. According to the experimental results on link prediction and triple classification, the proposed logical property preserving embeddings show the state-of-the-art performance at both tasks. These results prove that it is critical to preserve logical properties of relations while embedding knowledge graphs, and the proposed method does it effectively.

Parallel Session 6C: Semantic Similarity

Grande Ballroom C

Chair: Dipanjan Das

DAG-Structured Long Short-Term Memory for Semantic Compositionality

Xiaodan Zhu, Parinaz Sobhani, and Hongyu Guo

14:30–14:50

Recurrent neural networks, particularly Long Short-Term Memory (LSTM), have recently shown to be very effective in a wide range of sequence modeling problems, core to which is effective learning of distributed representation for sequences or subsequences. A strong assumption in almost all the previous models, however, posits that the learned representation (e.g., the meaning of a sentence), is fully composable from the atomic components (e.g., representation for words), while noncompositionality is a basic phenomenon in human languages. In this paper, we relieve this assumption by extending the chain-structured LSTM to directed acyclic graphs (DAGs), with the aim to endow linear-chain LSTMs with the capability of considering both compositionality and non-compositionality in a semantic composition framework. Our experiments on a sentiment composition benchmark dataset demonstrate that the proposed model achieves the state-of-the-art performance, significantly outperforming models that lack this ability.

Bayesian Supervised Domain Adaptation for Short Text Similarity

Md Arafa Sultan, Jordan Boyd-Graber, and Tamara Sumner

14:50–15:10

Identification of short text similarity (STS) is a high-utility NLP task with applications in a variety of domains. We explore adaptation of STS algorithms to different target domains and applications. A two-level hierarchical Bayesian model is employed for domain adaptation (DA) of a linear STS model to text from different sources (e.g., news, tweets). This model is then further extended for multitask learning (MTL) of three related tasks: STS, short answer scoring (SAS) and answer sentence ranking (ASR). In our experiments, the adaptive model demonstrates better overall cross-domain and cross-task performance over two non-adaptive baselines.

Pairwise Word Interaction Modeling with Deep Neural Networks for Semantic Similarity Measurement

Hua He and Jimmy Lin

15:10–15:30

Textual similarity measurement is a challenging problem, as it requires understanding the semantics of input sentences. Most previous neural network models use coarse-grained sentence modeling, which has difficulty capturing fine-grained word-level information for semantic comparisons. As an alternative, we propose to explicitly model pairwise word interactions and present a novel similarity focus mechanism to identify important correspondences for better similarity measurement. Our ideas are implemented in a novel neural network architecture that demonstrates state-of-the-art accuracy on three SemEval tasks and two answer selection tasks.

Session 7 Overview – Tuesday, June 14, 2016

	Track A	Track B	Track C
	<i>Machine Translation III</i> Grande Ballroom A	<i>Anaphora Resolution</i> Grande Ballroom B	<i>Word Embeddings I</i> Grande Ballroom C
4:00	An Attentional Model for Speech Translation Without Transcription <i>Duong, Anastasopoulos, Chiang, Bird, and Cohn</i>	A Novel Approach to Dropped Pronoun Translation <i>Wang, Tu, Zhang, Li, Way, and Liu</i>	Embedding Lexical Features via Low-Rank Tensors <i>Yu, Dredze, Arora, and Gormley</i>
4:20	Information Density and Quality Estimation Features as Translationese Indicators for Human Translation Classification <i>Rubino, Lapshinova-Koltunski, and Genabith</i>	Learning Global Features for Coreference Resolution <i>Wiseman, Rush, and Shieber</i>	The Role of Context Types and Dimensionality in Learning Word Embeddings <i>Melamud, McClosky, Patwardhan, and Bansal</i>
4:40	Interpretese vs. Translationese: The Uniqueness of Human Strategies in Simultaneous Interpretation <i>He, Boyd-Graber, and Daumé III</i>	Search Space Pruning: A Simple Solution for Better Coreference Resolvers <i>Moosavi and Strube</i>	Improve Chinese Word Embeddings by Exploiting Internal Structure <i>Xu, Liu, Zhang, Li, and Chen</i>
4:50	LSTM Neural Reordering Feature for Statistical Machine Translation <i>Cui, Wang, and Li</i>	Unsupervised Ranking Model for Entity Coreference Resolution <i>Ma, Liu, and Hovy</i>	

Parallel Session 7

Parallel Session 7A: Machine Translation III

Grande Ballroom A

Chair: Marine Carpuat

An Attentional Model for Speech Translation Without Transcription

Long Duong, Antonios Anastasopoulos, David Chiang, Steven Bird, and Trevor Cohn 16:00–16:20

For many low-resource languages, spoken language resources are more likely to be annotated with translations than with transcriptions. This bilingual speech data can be used for word-spotting, spoken document retrieval, and even for documentation of endangered languages. We experiment with extensions of the neural, attentional model of Bahdanau et al. for such data. On phone-to-word alignment and translation reranking tasks, we achieve large improvements relative to several baselines. On the more challenging speech-to-word alignment task, our model nearly matches GIZA++'s performance on gold transcriptions, but without any recourse to transcriptions or to a lexicon.

Information Density and Quality Estimation Features as Translationese Indicators for Human Translation Classification

Raphael Rubino, Ekaterina Lapshinova-Koltunski, and Josef van Genabith 16:20–16:40

This paper introduces information density and machine translation quality estimation inspired features to automatically detect and classify human translated texts. We investigate two settings: discriminating between translations and comparable originally authored texts, and distinguishing two levels of translation professionalism. Our framework is based on delexicalised sentence-level dense feature vector representations combined with a supervised machine learning approach. The results show state-of-the-art performance for mixed-domain translationese detection with information density and quality estimation based features, while results on translation expertise classification are mixed.

Interpretese vs. Translationese: The Uniqueness of Human Strategies in Simultaneous Interpretation

He He, Jordan Boyd-Graber, and Hal Daumé III 16:40–16:50

Computational approaches to simultaneous interpretation are held back by how little we know about the tactics human interpreters use. We produce a parallel corpus of translated and simultaneously interpreted text and study differences between them through a computational approach. Our analysis reveals that human interpreters regularly apply several effective tactics to reduce translation latency, including sentence segmentation and passivization. In addition to these unique, clever strategies, we show that limited human memory also causes other idiosyncratic properties of human interpretation such as generalization and omission of source content.

LSTM Neural Reordering Feature for Statistical Machine Translation

Yiming Cui, Shijin Wang, and Jianfeng Li 16:50–17:00

Artificial neural networks are powerful models, which have been widely applied into many aspects of machine translation, such as language modeling and translation modeling. Though notable improvements have been made in these areas, the reordering problem still remains a challenge in statistical machine translations. In this paper, we present a novel neural reordering model that directly models word pairs and their alignment. Further by utilizing LSTM recurrent neural networks, much longer context could be learned for reordering prediction. Experimental results on NIST OpenMT12 Arabic-English and Chinese-English 1000-best rescoring task show that our LSTM neural reordering feature is robust, and achieves significant improvements over various baseline systems.

Parallel Session 7B: Anaphora Resolution

Grande Ballroom B

Chair: Vincent Ng

A Novel Approach to Dropped Pronoun Translation

Longyue Wang, Zhaopeng Tu, Xiaojun Zhang, Hang Li, Andy Way, and Qun Liu 16:00–16:20

Dropped Pronouns (DP) in which pronouns are frequently dropped in the source language but should be retained in the target language are challenge in machine translation. In response to this problem, we propose a semi-supervised approach to recall possibly missing pronouns in the translation. Firstly, we build training data for DP generation in which the DPs are automatically labelled according to the alignment information from a parallel corpus. Secondly, we build a deep learning-based DP generator for input sentences in decoding when no corresponding references exist. More specifically, the generation is two-phase: (1) DP position detection, which is modeled as a sequential labelling task with recurrent neural networks; and (2) DP prediction, which employs a multilayer perceptron with rich features. Finally, we integrate the above outputs into our translation system to recall missing pronouns by both extracting rules from the DP-labelled training data and translating the DP-generated input sentences. Experimental results show that our approach achieves a significant improvement of 1.58 BLEU points in translation performance with 66% F-score for DP generation accuracy.

Learning Global Features for Coreference Resolution

Sam Wiseman, Alexander M. Rush, and Stuart M. Shieber 16:20–16:40

There is compelling evidence that coreference prediction would benefit from modeling global information about entity-clusters. Yet, state-of-the-art performance can be achieved with systems treating each mention prediction independently, which we attribute to the inherent difficulty of crafting informative cluster-level features. We instead propose to use recurrent neural networks (RNNs) to learn latent, global representations of entity clusters directly from their mentions. We show that such representations are especially useful for the prediction of pronominal mentions, and can be incorporated into an end-to-end coreference system that outperforms the state of the art without requiring any additional search.

Search Space Pruning: A Simple Solution for Better Coreference Resolvers

Nafise Sadat Moosavi and Michael Strube 16:40–16:50

There is a significant gap between the performance of a coreference resolution system on gold mentions and on system mentions. This gap is due to the large and unbalanced search space in coreference resolution when using system mentions. In this paper we show that search space pruning is a simple but efficient way of improving coreference resolvers. By incorporating our pruning method in one of the state-of-the-art coreference resolution systems, we achieve the best reported overall score on the CoNLL 2012 English test set. A version of our pruning method is available with the Cort coreference resolution source code.

Unsupervised Ranking Model for Entity Coreference Resolution

Xuezhe Ma, Zhengzhong Liu, and Eduard Hovy 16:50–17:00

Coreference resolution is one of the first stages in deep language understanding and its importance has been well recognized in the natural language processing community. In this paper, we propose a generative, unsupervised ranking model for entity coreference resolution. Our unsupervised system achieves 58.44% F1 score of the CoNLL metric on the English data from CoNLL-2012 shared task, outperforming the Stanford deterministic system, the winner of the CoNLL 2011 shared task, by 3.01% absolute.

Parallel Session 7C: Word Embeddings I

Grande Ballroom C

Chair: Manaal Faruqui

Embedding Lexical Features via Low-Rank Tensors*Mo Yu, Mark Dredze, Raman Arora, and Matthew R. Gormley*

16:00–16:20

Modern NLP models rely heavily on engineered features, which often combine word and contextual information into complex lexical features. Such combination results in large numbers of features, which can lead to over-fitting. We present a new model that represents complex lexical features, which contains parts for words, contextual information and labels, in a tensor that captures conjunction information among these parts. We apply low-rank tensor approximations to the corresponding parameter tensors to reduce the parameter space and improve prediction speed. Furthermore, we investigate two methods for handling features that include n-grams of mixed lengths. Our model achieves state-of-the-art results on tasks in relation extraction, PP-attachment, and preposition disambiguation.

The Role of Context Types and Dimensionality in Learning Word Embeddings*Oren Melamud, David McClosky, Siddharth Patwardhan, and Mohit Bansal*

16:20–16:40

We provide the first extensive evaluation of how using different types of context to learn skip-gram word embeddings affects performance on a wide range of intrinsic and extrinsic NLP tasks. Our results suggest that while intrinsic tasks tend to exhibit a clear preference to particular types of contexts and higher dimensionality, more careful tuning is required for finding the optimal settings for most of the extrinsic tasks that we considered. Furthermore, for these extrinsic tasks, we find that once the benefit from increasing the embedding dimensionality is mostly exhausted, simple concatenation of word embeddings, learned with different context types, can yield further performance gains. As an additional contribution, we propose a new variant of the skip-gram model that learns word embeddings from weighted contexts of substitute words.

Improve Chinese Word Embeddings by Exploiting Internal Structure*Jian Xu, Jiawei Liu, Liangang Zhang, Zhengyu Li, and Huanhuan Chen*

16:40–17:00

Recently, researchers have demonstrated that both Chinese word and its component characters provide rich semantic information when learning Chinese word embeddings. However, they ignored the semantic similarity across component characters in a word. In this paper, we learn the semantic contribution of characters to a word by exploiting the similarity between a word and its component characters with the semantic knowledge obtained from other languages. We propose a similarity-based method to learn Chinese word and character embeddings jointly. This method is also capable of disambiguating Chinese characters and distinguishing non-compositional Chinese words. Experiments on word similarity and text classification demonstrate the effectiveness of our method.

One Minute Madness

Time: 5:15–6:00

Location: Grande Ballroom

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

Chair: Joel Tetreault

Poster Session 2

Time: 6:00–8:00

Location: Pavilion

Assessing Relative Sentence Complexity using an Incremental CCG Parser

Bharat Ram Ambati, Siva Reddy, and Mark Steedman

Given a pair of sentences, we present computational models to assess if one sentence is simpler to read than the other. While existing models explored the usage of phrase structure features using a non-incremental parser, experimental evidence suggests that the human language processor works incrementally. We empirically evaluate if syntactic features from an incremental CCG parser are more useful than features from a non-incremental phrase structure parser. Our evaluation on Simple and Standard Wikipedia sentence pairs suggests that incremental CCG features are indeed more useful than phrase structure features achieving 0.44 points gain in performance. Incremental CCG parser also gives significant improvements in speed (12 times faster) in comparison to the phrase structure parser. Furthermore, with the addition of psycholinguistic features, we achieve the strongest result to date reported on this task. Our code and data can be downloaded from <https://github.com/bharatambati/sent-compl>.

Frustratingly Easy Cross-Lingual Transfer for Transition-Based Dependency Parsing

Ophélie Lacroix, Lauriane Aufrant, Guillaume Wisniewski, and François Yvon

In this paper, we present a straightforward strategy for transferring dependency parsers across languages. The proposed method learns a parser from partially annotated data obtained through the projection of annotations across unambiguous word alignments. It does not rely on any modeling of the reliability of dependency and/or alignment links and is therefore easy to implement and parameter free. Experiments on six languages show that our method is at par with recent algorithmically demanding methods, at a much cheaper computational cost. It can thus serve as a fair baseline for transferring dependencies across languages with the use of parallel corpora.

Geolocation for Twitter: Timing Matters

Mark Dredze, Miles Osborne, and Prabhayan Kambadur

Automated geolocation of social media messages can benefit a variety of downstream applications. However, these geolocation systems are typically evaluated without attention to how changes in time impact geolocation. Since different people, in different locations write messages at different times, these factors can significantly vary the performance of a geolocation system over time. We demonstrate cyclical temporal effects on geolocation accuracy in Twitter, as well as rapid drops as test data moves beyond the time period of training data. We show that temporal drift can effectively be countered with even modest online model updates.

Fast and Easy Short Answer Grading with High Accuracy

Md Arafat Sultan, Cristobal Salazar, and Tamara Sumner

We present a fast, simple and high-accuracy short answer grading system. Given a short-answer question and its correct answer, key measures of the correctness of a student response can be derived from its semantic similarity with the correct answer. Our supervised model (1) utilizes recent advances in the identification of short-text similarity, and (2) augments text similarity features with key grading-specific constructs. We present experimental results where our model demonstrates top performance on multiple benchmarks.

Interlocking Phrases in Phrase-based Statistical Machine Translation*Ye Kyaw Thu, Andrew Finch, and Eiichiro Sumita*

This paper presents an study of the use of interlocking phrases in phrase-based statistical machine translation. We examine the effect on translation quality when the translation units used in the translation hypotheses are allowed to overlap on the source side, on the target side and on both sides. A large-scale evaluation on 380 language pairs was conducted. Our results show that overall the use of overlapping phrases improved translation quality by 0.3 BLEU points on average. Further analysis revealed that language pairs requiring a larger amount of re-ordering benefited the most from our approach. When the evaluation was restricted to such pairs, the average improvement increased to up to 0.75 BLEU points with over 97% of the pairs improving. Our approach requires only a simple modification to the decoding algorithm and we believe it should be generally applicable to improve the performance of phrase-based decoders.

Eyes Don't Lie: Predicting Machine Translation Quality Using Eye Movement*Hassan Sajjad, Francisco Guzmán, Nadir Durrani, Ahmed Abdelali, Houda Bouamor, Irina Temnikova, and Stephan Vogel*

Poorly translated text is often disfluent and difficult to read. In contrast, well-formed translations require less time to process. In this paper, we model the differences in reading patterns of Machine Translation (MT) evaluators using novel features extracted from their gaze data, and we learn to predict the quality scores given by those evaluators. We test our predictions in a pairwise ranking scenario, measuring Kendall's tau correlation with the judgments. We show that our features provide information beyond fluency, and can be combined with BLEU for better predictions. Furthermore, our results show that reading patterns can be used to build semi-automatic metrics that anticipate the scores given by the evaluators.

Making Dependency Labeling Simple, Fast and Accurate*Tianxiao Shen, Tao Lei, and Regina Barzilay*

This work addresses the task of dependency labeling—assigning labels to an (unlabeled) dependency tree. We employ and extend a feature representation learning approach, optimizing it for both high speed and accuracy. We apply our labeling model on top of state-of-the-art parsers and evaluate its performance on standard benchmarks including the CoNLL-2009 and the English PTB datasets. Our model processes over 1,700 English sentences per second, which is 30 times faster than the sparse-feature method. It improves labeling accuracy over the outputs of top parsers, achieving the best LAS on 5 out of 7 datasets.

Deep Lexical Segmentation and Syntactic Parsing in the Easy-First Dependency Framework*Matthieu Constant, Joseph Le Roux, and Nadi Tomeh*

We explore the consequences of representing token segmentations as hierarchical structures (trees) for the task of Multiword Expression (MWE) recognition, in isolation or in combination with dependency parsing. We propose a novel representation of token segmentation as trees on tokens, resembling dependency trees. Given this new representation, we present and evaluate two different architectures to combine MWE recognition and dependency parsing in the easy-first framework: a pipeline and a joint system, both taking advantage of lexical and syntactic dimensions. We experimentally validate that MWE recognition significantly helps syntactic parsing.

Sentiment Composition of Words with Opposing Polarities*Svetlana Kiritchenko and Saif M. Mohammad*

In this paper, we explore sentiment composition in phrases that have at least one positive and at least one negative word-phrases like 'happy accident' and 'best winter break'. We compiled a dataset of such opposing polarity phrases and manually annotated them with real-valued scores of sentiment association. Using this dataset, we analyze the linguistic patterns present in opposing polarity phrases. Finally, we apply several unsupervised and supervised techniques of sentiment composition to determine their efficacy on this dataset. Our best system, which incorporates information from the phrase's constituents, their parts of speech, their sentiment association scores, and their embedding vectors, obtains an accuracy of over 80% on the opposing polarity phrases.

Learning to Recognize Ancillary Information for Automatic Paraphrase Identification*Simone Filice and Alessandro Moschitti*

Previous work on Automatic Paraphrase Identification (PI) is mainly based on modeling text similarity between two sentences. In contrast, we study methods for automatically detecting whether a text fragment only

appearing in a sentence of the evaluated sentence pair is important or ancillary information with respect to the paraphrase identification task. Engineering features for this new task is rather difficult, thus, we approach the problem by representing text with syntactic structures and applying tree kernels on them. The results show that the accuracy of our automatic Ancillary Text Classifier (ATC) is promising, i.e., 68.6%, and its output can be used to improve the state of the art in PI.

Learning a POS tagger for AAVE-like language

Anna Jørgensen, Dirk Hovy, and Anders Søgaard

Part-of-speech (POS) taggers trained on newswire perform much worse on domains such as subtitles, lyrics, or tweets. In addition, these domains are also heterogeneous, e.g., with respect to registers and dialects. In this paper, we consider the problem of learning a POS tagger for subtitles, lyrics, and tweets associated with African-American Vernacular English (AAVE). We learn from a mixture of randomly sampled and manually annotated Twitter data and unlabeled data, which we automatically and partially label using mined tag dictionaries. Our POS tagger obtains a tagging accuracy of 89% on subtitles, 85% on lyrics, and 83% on tweets, with up to 55% error reductions over a state-of-the-art newswire POS tagger, and 15-25% error reductions over a state-of-the-art Twitter POS tagger.

PIC a Different Word: A Simple Model for Lexical Substitution in Context

Stephen Roller and Katrin Erk

The Lexical Substitution task involves selecting and ranking lexical paraphrases for a target word in a given sentential context. We present PIC, a simple measure for estimating the appropriateness of substitutes in a given context. PIC outperforms another simple, comparable model proposed in recent work, especially when selecting substitutes from the entire vocabulary. Analysis shows that PIC improves over baselines by incorporating frequency biases into predictions.

Bootstrapping Translation Detection and Sentence Extraction from Comparable Corpora

Kriste Krstovski and David Smith

Most work on extracting parallel text from comparable corpora depends on linguistic resources such as seed parallel documents or translation dictionaries. This paper presents a simple baseline approach for bootstrapping a parallel collection. It starts by observing documents published on similar dates and the co-occurrence of a small number of identical tokens across languages. It then uses fast, online inference for a latent variable model to represent multilingual documents in a shared topic space where it can do efficient nearest-neighbor search. Starting from the Gigaword collections in English and Spanish, we train a translation system that outperforms one trained on the WMT'11 parallel training set.

Discriminative Reranking for Grammatical Error Correction with Statistical Machine Translation

Tomoya Mizumoto and Yuji Matsumoto

Research on grammatical error correction has received considerable attention. For dealing with all types of errors, grammatical error correction methods that employ statistical machine translation (SMT) have been proposed in recent years. An SMT system generates candidates with scores for all candidates and selects the sentence with the highest score as the correction result. However, the 1-best result of an SMT system is not always the best result. Thus, we propose a reranking approach for grammatical error correction. The reranking approach is used to re-score N-best results of the SMT and reorder the results. Our experiments show that our reranking system using parts of speech and syntactic features improves performance and achieves state-of-the-art quality, with an F0:5 score of 40.0.

Patterns of Wisdom: Discourse-Level Style in Multi-Sentence Quotations [BEST SHORT PAPER RUNNER-UP]

Kyle Booten and Marti A. Hearst

Quotations are kernels not just of wisdom but also of beautiful and striking language. While recent studies have characterized the stylistic features of quotations, we characterize the order of stylistic information within quotations. Analyzing a corpus of two-sentence quotations collected from the social network Tumblr, we explore the ways that both low-level features and high-level features tend to occur in either the first or second sentence. Through analysis of examples, we interpret these tendencies as manifestations of rhetorical patterns. Results from a prediction task suggest that stylistic patterns are more prominent in quotations than in a comparison corpus.

Right-truncatable Neural Word Embeddings*Jun Suzuki and Masao Nagata*

This paper proposes an incremental learning strategy of neural word embedding methods, which is a simple modification of conventional methods, such as SkipGrams and Global Vectors. Since our method iteratively constructs embedding vectors one dimension by one dimension, obtained vectors equip a unique property, namely, any right-truncated vector matches the solutions of the corresponding lower-dimensional embedding. Therefore, a single embedding vector can manage a wide range of dimensional requirements imposed by many different uses and applications. We demonstrate the effectiveness of our method by evaluating the obtained embedding vectors on a wide variety of linguistic benchmark data.

MAWPS: A Math Word Problem Repository*Rik Koncel-Kedziorski, Subhro Roy, Aida Amini, Nate Kushman, and Hannaneh Hajishirzi*

Recent work across several AI subdisciplines has focused on automatically solving math word problems. In this paper we introduce MAWPS, an online repository of Math Word Problems, to provide a unified testbed to evaluate different algorithms. MAWPS allows for the automatic construction of datasets with particular characteristics, providing tools for tuning the lexical and template overlap of a dataset as well as for filtering ungrammatical problems from web-sourced corpora. The online nature of this repository facilitates easy community contribution. At present, we have amassed 3,320 problems, including the full datasets used in several prominent works.

Cross-genre Event Extraction with Knowledge Enrichment*Hao Li and Heng Ji*

The goal of Event extraction is to extract structured information of events that are of interest from unstructured documents. Existing event extractors for social media suffer from two major problems: lack of context and informal nature. In this paper, instead of conducting event extraction solely on each social media message, we incorporate cross-genre knowledge to boost the event extractor performance. Experiment results demonstrate that without any additional annotations, our proposed approach is able to provide 15% absolute F-score improvement over the state-of-the-art.

Emergent: a novel data-set for stance classification*William Ferreira and Andreas Vlachos*

We present Emergent, a novel data-set derived from a digital journalism project for rumour debunking. The data-set contains 300 rumoured claims and 2,595 associated news articles, collected and labelled by journalists with an estimation of their veracity true, false or unverified. Each associated article is summarized into a headline and labelled to indicate whether its stance is for, against, or observing the claim, where observing indicates that the article merely repeats the claim. Thus, Emergent provides a real-world data source for a variety of natural language processing tasks in the context of fact-checking. Further to presenting the dataset, we address the task of determining the article headline stance with respect to the claim. For this purpose we use a logistic regression classifier and develop features that examine the headline and its agreement with the claim. The accuracy achieved was 73% which is 26% higher than the one achieved by the Excitements Open Platform (Magnini et al., 2014).

BIRA: Improved Predictive Exchange Word Clustering*Jon Dehdari, Liling Tan, and Josef van Genabith*

Word clusters are useful for many NLP tasks including training neural network language models, but current increases in datasets are outpacing the ability of word clusterers to handle them. Little attention has been paid thus far on inducing high-quality word clusters at a large scale. The predictive exchange algorithm is quite scalable, but sometimes does not provide as good perplexity as other slower clustering algorithms. We introduce the bidirectional, interpolated, refining, and alternating (BIRA) predictive exchange algorithm. It improves upon the predictive exchange algorithm's perplexity by up to 18%, giving it perplexities comparable to the slower two-sided exchange algorithm, and better perplexities than the slower Brown clustering algorithm. Our BIRA implementation is fast, clustering a 2.5 billion token English News Crawl corpus in 3 hours. It also reduces machine translation training time while preserving translation quality. Our implementation is portable and freely available.

Integrating Morphological Desegmentation into Phrase-based Decoding*Mohammad Salameh, Colin Cherry, and Grzegorz Kondrak*

When translating into a morphologically complex language, segmenting the target language can reduce data sparsity, while introducing the complication of desegmenting the system output. We present a method for decoder-integrated desegmentation, allowing features that consider the desegmented target, such as a word-level language model, to be considered throughout the entire search space. Our results on a large-scale, English to Arabic translation task show significant improvement over the 1-best desegmentation baseline.

The Instantiation Discourse Relation: A Corpus Analysis of Its Properties and Improved Detection

Junyi Jessy Li and Ani Nenkova

Instantiation is a fairly common discourse relation and past work has suggested that it plays special roles in local coherence, in sentiment expression and in content selection in summarization. In this paper we provide the first systematic corpus analysis of the relation and show that relation-specific features can improve considerably the detection of the relation. We show that sentences involved in Instantiation are set apart from other sentences by the use of gradable (subjective) adjectives, the occurrence of rare words and by different patterns in part of speech use. Words across arguments of Instantiation are connected through hypernym and meronym relations significantly more often than in other sentences and that they stand out in context by being significantly less similar to each other than other adjacent sentence pairs. These factors provide substantial predictive power that improves the identification of Instantiation by more than 5% F-measure.

Sparse Bilingual Word Representations for Cross-lingual Lexical Entailment

Yogarshi Vyas and Marine Carpuat

We introduce the task of cross-lingual lexical entailment, which aims to detect whether the meaning of a word in one language can be inferred from the meaning of a word in another language. We construct a gold standard for this task, and propose an unsupervised solution based on distributional word representations. As commonly done in the monolingual setting, we assume a word e entails a word f if the prominent context features of e are a subset of those of f. To address the challenge of comparing contexts across languages, we propose a novel method for inducing sparse bilingual word representations from monolingual and parallel texts. Our approach yields an F-score of 70%, and significantly outperforms strong baselines based on translation and on existing word representations.

Automatic Prediction of Linguistic Decline in Writings of Subjects with Degenerative Dementia

Davy Weissenbacher, Travis A. Johnson, Laura Wojtulewicz, Amylou Dueck, Dona Locke, Richard Caselli, and Graciela Gonzalez

Given the limited success of medication in reversing the effects of Alzheimer's and other dementias, a lot of the neuroscience research has been focused on early detection, in order to slow the progress of the disease through different interventions. We propose a Natural Language Processing approach applied to descriptive writing to attempt to discriminate decline due to normal aging from decline due to pre-dementia conditions. Within the context of a longitudinal study on Alzheimer's disease, we created a unique corpus of 201 descriptions of a control image written by subjects of the study. Our classifier, computing linguistic features, was able to discriminate normal from cognitively impaired patients to an accuracy of 86.1% using lexical and semantic irregularities found in their writing. This is a promising result towards elucidating the existence of a general pattern in linguistic deterioration caused by dementia that might be detectable from a subject's written descriptive language.

Consensus Maximization Fusion of Probabilistic Information Extractors

Miguel Rodríguez, Sean Goldberg, and Daisy Zhe Wang

Current approaches to Information Extraction (IE) are capable of extracting large amounts of facts with associated probabilities. Because no current IE system is perfect, complementary and conflicting facts are obtained when different systems are run over the same data. Knowledge Fusion (KF) is the problem of aggregating facts from different extractors. Existing methods approach KF using supervised learning or deep linguistic knowledge, which either lack sufficient data or are not robust enough. We propose a semi-supervised application of Consensus Maximization to the KF problem, using a combination of supervised and unsupervised models. Consensus Maximization Fusion (CM Fusion) is able to promote high quality facts and eliminate incorrect ones. We demonstrate the effectiveness of our system on the NIST Slot Filler Validation contest, which seeks to evaluate and aggregate multiple independent information extractors. Our system achieved the highest F1 score relative to other system submissions.

Simple, Fast Noise-Contrastive Estimation for Large RNN Vocabularies*Barret Zoph, Ashish Vaswani, Jonathan May, and Kevin Knight*

We present a simple algorithm to efficiently train language models with noise-contrastive estimation (NCE) on graphics processing units (GPUs). Our NCE-trained language models achieve significantly lower perplexity on the One Billion Word Benchmark language modeling challenge, and contain one sixth of the parameters of the best single model in Chelba et al. (2013). When incorporated into a strong Arabic-English machine translation system they give a strong boost in translation quality. We release a toolkit so that others may also train large-scale, large vocabulary LSTM language models with NCE, parallelizing computation across multiple GPUs.

Automatically Inferring Implicit Properties in Similes*Ashequl Qadir, Ellen Riloff, and Marilyn A. Walker*

A simile is a figure of speech comparing two fundamentally different things. Sometimes, a simile will explain the basis of a comparison by explicitly mentioning a shared property. For example, “my room is as cold as Antarctica” gives “cold” as the property shared by the room and Antarctica. But most similes do not give an explicit property (e.g., “my room feels like Antarctica”) leaving the reader to infer that the room is cold. We tackle the problem of automatically inferring implicit properties evoked by similes. Our approach involves three steps: (1) generating candidate properties from different sources, (2) evaluating properties based on the influence of multiple simile components, and (3) aggregated ranking of the properties. We also present an analysis showing that the difficulty of inferring an implicit property for a simile correlates with its interpretive diversity.

Visual Storytelling

Ting-Hao (Kenneth) Huang, Francis Ferraro, Nasrin Mostafazadeh, Ishan Misra, Aishwarya Agrawal, Jacob Devlin, Ross Girshick, Xiaodong He, Pushmeet Kohli, Dhruv Batra, C. Lawrence Zitnick, Devi Parikh, Lucy Vanderwende, Michel Galley, and Margaret Mitchell

We introduce the first dataset for sequential vision-to-language, and explore how this data may be used for the task of visual storytelling. The dataset includes 81,743 unique photos in 20,211 sequences, aligned to descriptive and story language. We establish several strong baselines for the storytelling task, and motivate an automatic metric to benchmark progress. We argue that modelling figurative and social language, as provided for in this data and the storytelling task, has the potential to move artificial intelligence from basic understandings of typical visual scenes towards more and more human-like understanding of grounded event structure and subjective expression.

PRIMT: A Pick-Revise Framework for Interactive Machine Translation*Shanbo Cheng, Shujian Huang, Huadong Chen, Xin-Yu Dai, and Jiajun Chen*

Interactive machine translation (IMT) is a method which uses human-computer interactions to improve the quality of MT. Traditional IMT methods employ a left-to-right order for the interactions, which is difficult to directly modify critical errors at the end of the sentence. In this paper, we propose an IMT framework in which the interaction is decomposed into two simple human actions: picking a critical translation error (Pick) and revising the translation (Revise). The picked phrase could be at any position of the sentence, which improves the efficiency of human computer interaction. We also propose automatic suggestion models for the two actions to further reduce the cost of human interaction. Experiment results demonstrate that by interactions through either one of the actions, the translation quality could be significantly improved. Greater gains could be achieved by iteratively performing both actions.

Incorporating Side Information into Recurrent Neural Network Language Models*Cong Duy Vu Hoang, Trevor Cohn, and Gholamreza Haffari*

Recurrent neural network language models (RNNLM) have recently demonstrated vast potential in modelling long-term dependencies for NLP problems, ranging from speech recognition to machine translation. In this work, we propose methods for conditioning RNNLMs on external side information, e.g., metadata such as keywords, description, document title or topic headline. Our experiments show consistent improvements of RNNLMs using side information over the baselines for two different datasets and genres in two languages. Interestingly, we found that side information in a foreign language can be highly beneficial in modelling texts in another language, serving as a form of cross-lingual language modelling.

Capturing Semantic Similarity for Entity Linking with Convolutional Neural Networks*Matthew Francis-Landau, Greg Durrett, and Dan Klein*

A key challenge in entity linking is making effective use of contextual information to disambiguate mentions that might refer to different entities in different contexts. We present a model that uses convolutional neural networks to capture semantic correspondence between a mention’s context and a proposed target entity. These convolutional networks operate at multiple granularities to exploit various kinds of topic information, and their rich parameterization gives them the capacity to learn which n-grams characterize different topics. We combine these networks with a sparse linear model to achieve state-of-the-art performance on multiple entity linking datasets, outperforming the prior systems of Durrett and Klein (2014) and Nguyen et al. (2014).

K-Embeddings: Learning Conceptual Embeddings for Words using Context

Thuy Vu and D. Stott Parker

We describe a technique for adding contextual distinctions to word embeddings by extending the usual embedding process — into two phases. The first phase resembles existing methods, but also constructs K classifications of concepts. The second phase uses these classifications in developing refined K embeddings for words, namely word K-embeddings. The technique is iterative, scalable, and can be combined with other methods (including Word2Vec) in achieving still more expressive representations. Experimental results show consistently large performance gains on a Semantic-Syntactic Word Relationship test set for different K settings. For example, an overall gain of 20% is recorded at $K = 5$. In addition, we demonstrate that an iterative process can further tune the embeddings and gain an extra 1% ($K = 10$ in 3 iterations) on the same benchmark. The examples also show that polysemous concepts are meaningfully embedded in our K different conceptual embeddings for words.

[TACL] Learning Composition Models for Phrase Embeddings

Mo Yu and Mark Dredze

Lexical embeddings can serve as useful representations for words for a variety of NLP tasks, but learning embeddings for phrases can be challenging. While separate embeddings are learned for each word, this is infeasible for every phrase. We construct phrase embeddings by learning how to compose word embeddings using features that capture phrase structure and context. We propose efficient unsupervised and task-specific learning objectives that scale our model to large datasets. We demonstrate improvements on both language modeling and several phrase semantic similarity tasks with various phrase lengths. We make the implementation of our model and the datasets available for general use.

System Demonstrations 2

Time: 6:00–8:00

Location: Pavilion

Illinois Math Solver: Math Reasoning on the Web

Subhra Roy and Dan Roth

There has been a recent interest in understanding text to perform mathematical reasoning. In particular, most of these efforts have focussed on automatically solving school level math word problems. In order to make advancements in this area accessible to people, as well as to facilitate this line of research, we release the Illinois Math Solver, a web based tool that supports performing mathematical reasoning. Illinois Math Solver can answer a wide range of mathematics questions, ranging from compositional operation questions like “What is the result when 6 is divided by the sum of 7 and 5 ?” to elementary school level math word problems, like “I bought 6 apples. I ate 3 of them. How many do I have left ?”. The web based demo can be used as a tutoring tool for elementary school students, since it not only outputs the final result, but also the mathematical expression to compute it. The tool will allow researchers to understand the capabilities and limitations of a state of the art arithmetic problem solver, and also enable crowd based data acquisition for mathematical reasoning. The system is currently online at http://cogcomp-xz.cs.illinois.edu/page/demo_view/Math.

LingoTurk: managing crowdsourced tasks for psycholinguistics

Florian Pusse, Asad Sayeed, and Vera Demberg

LingoTurk is an open-source, freely available crowdsourcing client/server system aimed primarily at psycholinguistic experimentation where custom and specialized user interfaces are required but not supported by popular crowdsourcing task management platforms. LingoTurk enables user-friendly local hosting of experiments as well as condition management and participant exclusion. It is compatible with Amazon Mechanical

Turk and Prolific Academic. New experiments can easily be set up via the Play Framework and the LingoTurk API, while multiple experiments can be managed from a single system.

Sentential Paraphrasing as Black-Box Machine Translation

Courtney Napaless, Chris Callison-Burch, and Matt Post

We present a simple, prepackaged solution to generating paraphrases of English sentences. We use the Paraphrase Database (PPDB) for monolingual sentence rewriting and provide machine translation language packs: prepackaged, tuned models that can be downloaded and used to generate paraphrases on a standard Unix environment. The language packs can be treated as a black box or customized to specific tasks. In this demonstration, we will explain how to use the included interactive web-based tool to generate sentential paraphrases.

A Tag-based English Math Word Problem Solver with Understanding, Reasoning and Explanation

Chao-Chun Liang, Kuang-Yi Hsu, Chien-Tsung Huang, Chung-Min Li, Shen-Yu Miao, and Keh-Yih Su

This paper presents a tag-based statistical math word problem solver with understanding, reasoning, and explanation. It analyzes the text and transforms both body and question parts into their tag-based logic forms, and then performs inference on them. The proposed tag-based approach provides the flexibility for annotating an extracted math quantity with its associated syntactic and semantic information, which can be used to identify the desired operand and filter out irrelevant quantities. The proposed approach is thus less sensitive to the irrelevant information and could provide the answer more precisely. Also, it can handle much more problem types other than addition and subtraction.

Cross-media Event Extraction and Recommendation

Di Lu, Clare Voss, Fangbo Tao, Xiang Ren, Rachel Guan, Rostyslav Korolov, Tongtao Zhang, Dongang Wang, Hongzhi Li, Taylor Cassidy, Heng Ji, Shih-fu Chang, Jiawei Han, William Wallace, James Hendler, Mei Si, and Lance Kaplan

The sheer volume of unstructured multimedia data (e.g., texts, images, videos) posted on the Web during events of general interest is overwhelming and difficult to distill if seeking information relevant to a particular concern. We have developed a comprehensive system that searches, identifies, organizes and summarizes complex events from multiple data modalities. It also recommends events related to the user's ongoing search based on previously selected attribute values and dimensions of events being viewed. In this paper we briefly present the algorithms of each component and demonstrate the system's capabilities.

SODA:Service Oriented Domain Adaptation Architecture for Microblog Categorization

Himanshu Sharad Bhatt, Sandipan Dandapat, Peddamuthu Balaji, Shourya Roy, Sharmistha Jat, and Deepali Semwal

We demonstrate SODA (Service Oriented Domain Adaptation) for efficient and scalable cross-domain microblog categorization which works on the principle of transfer learning. It is developed on a novel similarity-based iterative domain adaptation algorithm while extended with features such as active learning and interactive GUI to be used by business professionals. SODA demonstrates efficient classification accuracy on new collections while minimizing and sometimes eliminating the need for expensive data labeling efforts. SODA also implements an active learning (AL) technique to select informative instances from the new collection to seek annotations, if a small amount of labeled data is required by the adaptation algorithm.

Lecture Translator - Speech translation framework for simultaneous lecture translation

Markus Müller, Thai Son Nguyen, Jan Niehues, Eunah Cho, Bastian Krüger, Thanh-Le Ha, Kevin Kilgour, Matthias Sperber, Mohammed Mediani, Sebastian Stütter, and Alex Waibel

Foreign students at German universities often have difficulties following lectures as they are often held in German. Since human interpreters are too expensive for universities we are addressing this problem via speech translation technology deployed in KIT's lecture halls. Our simultaneous lecture translation system automatically translates lectures from German to English in real-time. Other supported language directions are English to Spanish, English to French, English to German and German to French. Automatic simultaneous translation is more than just the concatenation of automatic speech recognition and machine translation technology, as the input is an unsegmented, practically infinite stream of spontaneous speech. The lack of segmentation and the spontaneous nature of the speech makes it especially difficult to recognize and translate it with sufficient quality. In addition to quality, speed and latency are of the utmost importance in order for the system to enable students to follow lectures. In this paper we present our system that performs the task of simultaneous

speech translation of university lectures by performing speech translation on a stream of audio in real-time and with low latency. The system features several techniques beyond the basic speech translation task, that make it fit for real-world use. Examples of these features are a continuous stream speech recognition without any prior segmentation of the input audio, punctuation prediction, run-on decoding and run-on translation with continuously updating displays in order to keep the latency as low as possible.

Zara The Supergirl: An Empathetic Personality Recognition System

Pascale Fung, Anik Dey, Farhad Bin Siddique, Ruixi Lin, Yang Yang, Yan Wan, and Ho Yin Ricky Chan

Zara the Supergirl is an interactive system that, while having a conversation with a user, uses its built in sentiment analysis, emotion recognition, facial and speech recognition modules, to exhibit the human-like response of sharing emotions. In addition, at the end of a 5-10 minute conversation with the user, it can give a comprehensive personality analysis based on the user's interaction with Zara. This is a first prototype that has incorporated a full empathy module, the recognition and response of human emotions, into a spoken language interactive system that enhances human-robot understanding. Zara was shown at the World Economic Forum in Dalian in September 2015.

Kathaa: A Visual Programming Framework for NLP Applications

Sharada Prasanna Mohanty, Nehal J Wani, Manish Srivastava, and Dipti Misra Sharma

In this paper, we present Kathaa, an open source web based Visual Programming Framework for NLP applications. It supports design, execution and analysis of complex NLP systems by choosing and visually connecting NLP modules from an already available and easily extensible Module library. It models NLP systems as a Directed Acyclic Graph of optionally parallelized information flow, and lets the user choose and use available modules in their NLP applications irrespective of their technical proficiency. Kathaa exposes a precise Module definition API to allow easy integration of external NLP components (along with their associated services as docker containers), it allows everyone to publish their services in a standardized format for everyone else to use it out of the box.

“Why Should I Trust You?”: Explaining the Predictions of Any Classifier

Marco Ribeiro, Sameer Singh, and Carlos Guestrin

Despite widespread adoption in NLP, machine learning models remain mostly black boxes. Understanding the reasons behind predictions is, however, quite important in assessing trust in a model. Trust is fundamental if one plans to take action based on a prediction, or when choosing whether or not to deploy a new model. In this work, we describe LIME, a novel explanation technique that explains the predictions of any classifier in an interpretable and faithful manner. We further present a method to explain models by presenting representative individual predictions and their explanations in a non-redundant manner. We propose a demonstration of these ideas on different NLP tasks such as document classification, politeness detection, and twitter sentiment analysis, with classifiers like neural networks and SVMs. The user interactions include explanations of free-form text, challenging users to identify the better classifier from a pair, and perform basic feature engineering to improve the classifiers.

Student Workshop Posters 2

Time: 6:00–8:00

Location: Pavilion

Effects of Communicative Pressures on Novice L2 Learners' Use of Optional Formal Devices

Yoav Binoun, Francesca Delogu, Clayton Greenberg, Mindaugas Mozuraitis, and Matthew Crocker

We conducted an Artificial Language Learning experiment to examine the production behavior of language learners in a dynamic communicative setting. Participants were exposed to a miniature language with two optional formal devices and were then asked to use the acquired language to transfer information in a cooperative game. The results showed that language learners optimize their use of the optional formal devices to transfer information efficiently and that they avoid the production of ambiguous information. These results could be used within the context of a language model such that the model can more accurately reflect the production behavior of human language learners.

Explicit Argument Identification for Discourse Parsing In Hindi: A Hybrid Pipeline*Rohit Jain and Dipti Misra Sharma*

Shallow discourse parsing enables us to study discourse as a coherent piece of information rather than a sequence of clauses, sentences and paragraphs. In this paper, we identify arguments of explicit discourse relations in Hindi. This is the first such work carried out for Hindi. Building upon previous work carried out on discourse connective identification in Hindi, we propose a hybrid pipeline which makes use of both sub-tree extraction and linear tagging approaches. We report state-of-the-art performance for this task.

Exploring Fine-Grained Emotion Detection in Tweets*Jasy Suet Yan Liew and Howard R. Turtle*

We examine if common machine learning techniques known to perform well in coarse-grained emotion and sentiment classification can also be applied successfully on a set of fine-grained emotion categories. We first describe the grounded theory approach used to develop a corpus of 5,553 tweets manually annotated with 28 emotion categories. From our preliminary experiments, we have identified two machine learning algorithms that perform well in this emotion classification task and demonstrated that it is feasible to train classifiers to detect 28 emotion categories without a huge drop in performance compared to coarser-grained classification schemes.

Extraction of Bilingual Technical Terms for Chinese-Japanese Patent Translation*Wei Yang, Jinghui Yan, and Yves Lepage*

The translation of patents or scientific papers is a key issue that should be helped by the use of statistical machine translation (SMT). In this paper, we propose a method to improve Chinese-Japanese patent SMT by pre-marking the training corpus with aligned bilingual multi-word terms. We automatically extract multi-word terms from monolingual corpora by combining statistical and linguistic filtering methods. We use the sampling-based alignment method to identify aligned terms and set some threshold on translation probabilities to select the most promising bilingual multi-word terms. We pre-mark a Chinese-Japanese training corpus with such selected aligned bilingual multi-word terms. We obtain the performance of over 70% precision in bilingual term extraction and a significant improvement of BLEU scores in our experiments on a Chinese-Japanese patent parallel corpus.

Hateful Symbols or Hateful People? Predictive Features for Hate Speech Detection on Twitter*Zeerak Waseem and Dirk Hovy*

Hate speech in the form of racist and sexist remarks are a common occurrence on social media. For that reason many social media services address the problem of identifying hate speech, but the definition of hate speech varies markedly and is largely manual effort. We provide a list of criteria founded in critical race theory to annotate a publicly available corpus of more than 16k tweets, and analyze the impact of various features in conjunction with character n-grams, for detection of said hate speech. We also present the most indicative words, and present a dictionary based on our data.

Non-decreasing Sub-modular Function for Comprehensible Summarization*Liton J Kurisinkel, Pruthwik Mishra, Vigneshwaran Muralidaran, Vasudeva Varma, and Dipti Misra Sharma*

Extractive summarization techniques typically aim to maximize the information coverage of the summary with respect to the original corpus and report accuracies in ROUGE scores. Automated text summarization techniques should consider the dimensions of comprehensibility, coherence and readability. In the current work, we identify the discourse structure which provides the context for the creation of a sentence. We leverage the information from the structure to frame a mono-tone (non-decreasing) sub-modular scoring function for generating comprehensible summaries. Our approach improves the overall quality of comprehensibility of the summary in terms of human evaluation and gives sufficient content coverage with comparable ROUGE score. We also formulate a metric to measure summary comprehensibility in terms of Contextual Independence of a sentence. The metric is shown to be representative of human judgement of text comprehensibility.

Phylogenetic simulations over constraint-based grammar formalisms*Andrew Lamont and Jonathan Washington*

Computational phylogenetics has been shown to be effective over grammatical characteristics. Recent work suggests that constraint-based formalisms are compatible with such an approach (Eden, 2013). In this paper, we report on simulations to determine how useful constraint-based formalisms are in phylogenetic research and

under what conditions. Popular computational methods for phylogenetic research (estimating the evolutionary histories of languages) primarily involve comparisons over cognate sets (Nichols and Warnow, 2008). Recent works (Dunn et al., 2005; Longobardi and Guardiano, 2009) indicate that comparing sets of grammatical parameters can be effective as well. However, generating a large number of meaningful parameters remains a formal obstacle. In this paper we argue that constraint-based grammar formalisms may be exploited for parameter generation, and explore to what extent such research is feasible.

Question Answering over Knowledge Base using Factual Memory Networks

Sarthak Jain

In the task of question answering, Memory Networks have recently shown to be quite effective towards complex reasoning as well as scalability, in spite of limited range of topics covered in training data. In this paper, we introduce Factual Memory Network, which learns to answer questions by extracting and reasoning over relevant facts from a Knowledge Base. Our system generate distributed representation of questions and KB in same word vector space, extract a subset of initial candidate facts, then try to find a path to answer entity using multi-hop reasoning and refinement. Additionally, we also improve the run-time efficiency of our model using various computational heuristics.

Using Related Languages to Enhance Statistical Language Models

Anna Currey, Alina Karakanta, and Jon Dehdari

The success of many language modeling methods and applications relies heavily on the amount of data available. This problem is further exacerbated in statistical machine translation, where parallel data in the source and target languages is required. However, large amounts of data are only available for a small number of languages; as a result, many language modeling techniques are inadequate for the vast majority of languages. In this paper, we attempt to lessen the problem of a lack of training data for low-resource languages by adding data from related high-resource languages in three experiments. First, we interpolate language models trained on the target language and on the related language. In our second experiment, we select the sentences most similar to the target language and add them to our training corpus. Finally, we integrate data from the related language into a translation model for a statistical machine translation application. Although we do not see many significant improvements over baselines trained on a small amount of data in the target language, we discuss some further experiments that could be attempted in order to augment language models and translation models with data from related languages.

Social Event

Tuesday, June 14, 2016, 8pm–11pm

Bay View Lawn

The NAACL 2016 Social and Networking Event will be held immediately following the Tuesday Poster Session in the Bay View Lawn.

Main Conference: Wednesday, June 15

Overview

7:30 – 8:45	Breakfast	Pavilion	
9:00 – 10:15	Invited talk: "Human-based evaluations of language generation systems" (Ehud Reiter)	Grande Ballroom	
10:15 – 10:45	Coffee break	Grande Foyer	
	Session 8		
10:45 – 12:15	Question Answering <i>Grande Ballroom A</i>	Multilingual Processing <i>Grande Ballroom B</i>	Word Embeddings II <i>Grande Ballroom C</i>
12:15 – 1:00	Lunch		
1:00 – 2:00	NAACL business meeting, Grande Ballroom A	Grande Ballroom A	
	Session 9		
2:15 – 3:45	Argumentation and Discourse <i>Grande Ballroom A</i>	Misc Semantics <i>Grande Ballroom B</i>	Text Categorization <i>Grande Ballroom C</i>
3:45 – 4:15	Coffee break	Grande Foyer	
4:15 – 5:45	Best paper awards	Grande Ballroom	
5:35 – 5:45	Closing remarks	Grande Ballroom	

Keynote Address: Ehud Reitter

Evaluating Natural Language Generation Systems

Wednesday, June 15, 2016, 9:00–10:15am

Grande Ballroom



Abstract

Natural Language Generation (NLG) systems have different characteristics than other NLP systems, which effects how they are evaluated. In particular, it can be difficult to meaningfully evaluate NLG texts by comparing them against gold-standard reference texts, because (A) there are usually many possible texts which are acceptable to users and (B) some NLG systems produce texts which are better (as judged by human users) than human-written corpus texts. Partially because of these reasons, the NLG community places much more emphasis on human-based evaluations than most areas of NLP.

I will discuss the various ways in which NLG systems are evaluated, focusing on human-based evaluations. These typically either measure the success of generated texts at achieving a goal (eg, measuring how many people change their behaviour after reading behaviour-change texts produced by an NLG system); or ask human subjects to rate various aspects of generated texts (such as readability, accuracy, and appropriateness), often on Likert scales. I will use examples from evaluations I have carried out, and highlight some of the lessons I have learnt, including the importance of reporting negative results, the difference between laboratory and real-world evaluations, and the need to look at worse-case as well as average-case performance. I hope my talk will be interesting and relevant to anyone who is interested in the evaluation of NLP systems.

Biography

Ehud Reitter is a Professor of Computing Science at the University of Aberdeen and also Chief Scientist of Arria NLG. He has worked on natural language generation for the past 30 years, on methodology (including evaluation) and resources as well as algorithms, and is one of the most cited authors in NLG. His 2000 book *Building Natural Language Generation Systems* is widely used as an NLG textbook. Dr Reitter currently spends most of his time trying to commercialise NLG at Arria (one of the largest specialist NLG companies), which grew out of a startup he cofounded in 2009.

Session 8 Overview – Wednesday, June 15, 2016

Track A	Track B	Track C	
<i>Question Answering</i> Grande Ballroom A	<i>Multilingual Processing</i> Grande Ballroom B	<i>Word Embeddings II</i> Grande Ballroom C	
[TACL] A Joint Model for Answer Sentence Ranking and Answer Extraction <i>Sultan, Castelli, and Florian</i>	Multilingual Language Processing From Bytes <i>Gillick, Brunk, Vinyals, and Subramanya</i>	Bilingual Learning of Multi-sense Embeddings with Discrete Autoencoders <i>Šuster, Titov, and Noord</i>	10:45
Convolutional Neural Networks vs. Convolution Kernels: Feature Engineering for Answer Sentence Reranking <i>Tymoshenko, Bonadiman, and Moschitti</i>	Ten Pairs to Tag – Multilingual POS Tagging via Coarse Mapping between Embeddings <i>Zhang, Gaddy, Barzilay, and Jaakkola</i>	Polyglot Neural Language Models: A Case Study in Cross-Lingual Phonetic Representation Learning <i>Tsvetkov, Sitaram, Faruqui, Lample, Littell, Mortensen, Black, Levin, and Dyer</i>	11:05
Semi-supervised Question Retrieval with Gated Convolutions <i>Lei, Joshi, Barzilay, Jaakkola, Tymoshenko, Moschitti, and Márquez</i>	Part-of-Speech Tagging for Historical English <i>Yang and Eisenstein</i>	Learning Distributed Representations of Sentences from Unlabelled Data <i>Hill, Cho, and Korhonen</i>	11:25
[TACL] Parsing Algebraic Word Problems into Equations <i>Koncel-Kedziorski, Hajishirzi, Sabharwal, Etzioni, and Dumas Ang</i>	Statistical Modeling of Creole Genesis <i>Murawaki</i>	[TACL] Learning to Understand Phrases by Embedding the Dictionary <i>Hill, Cho, Korhonen, and Bengio</i>	11:45
This is how we do it: Answer Reranking for Open-domain How Questions with Paragraph Vectors and Minimal Feature Engineering <i>Bogdanova and Foster</i>	Shallow Parsing Pipeline - Hindi-English Code-Mixed Social Media Text <i>Sharma, Gupta, Motlani, Bansal, Shrivastava, Mami-di, and Sharma</i>	Retrofitting Sense-Specific Word Vectors Using Parallel Text <i>Ettinger, Resnik, and Carpuat</i>	12:05

Parallel Session 8

Parallel Session 8A: Question Answering

Grande Ballroom A

Chair: Eduard Hovy

[TACL] A Joint Model for Answer Sentence Ranking and Answer Extraction

Md Arafat Sultan, Vittorio Castelli, and Radu Florian

10:45–11:05

Answer sentence ranking and answer extraction are two key challenges in question answering that have traditionally been treated in isolation, i.e., as independent tasks. In this article, we (1) explain how both tasks are related at their core by a common quantity, and (2) propose a simple and intuitive joint probabilistic model that addresses both via joint computation but task-specific application of that quantity. In our experiments with two TREC datasets, our joint model substantially outperforms state-of-the-art systems in both tasks.

Convolutional Neural Networks vs. Convolution Kernels: Feature Engineering for Answer Sentence Reranking

Kateryna Tymoshenko, Daniele Bonadiman, and Alessandro Moschitti

11:05–11:25

In this paper, we study, compare and combine two state-of-the-art approaches to automatic feature engineering: Convolution Tree Kernels (CTKs) and Convolutional Neural Networks (CNNs) for learning to rank answer sentences in a Question Answering (QA) setting. When dealing with QA, the key aspect is to encode relational information between the constituents of question and answer in learning algorithms. For this purpose, we propose novel CNNs using relational information and combined them with relational CTKs. The results show that (i) both approaches achieve the state of the art on a question answering task, where CTKs produce higher accuracy and (ii) combining such methods leads to unprecedented high results.

Semi-supervised Question Retrieval with Gated Convolutions

Tao Lei, Hrishikesh Joshi, Regina Barzilay, Tommi Jaakkola, Kateryna Tymoshenko, Alessandro Moschitti, and Lluís Màrquez

11:25–11:45

Question answering forums are rapidly growing in size with no effective automated ability to refer to and reuse answers already available for previous posted questions. In this paper, we develop a methodology for finding semantically related questions. The task is difficult since 1) key pieces of information are often buried in extraneous details in the question body and 2) available annotations on similar questions are scarce and fragmented. We design a recurrent and convolutional model (gated convolution) to effectively map questions to their semantic representations. The models are pre-trained within an encoder-decoder framework (from body to title) on the basis of the entire raw corpus, and fine-tuned discriminatively from limited annotations. Our evaluation demonstrates that our model yields substantial gains over a standard IR baseline and various neural network architectures (including CNNs, LSTMs and GRUs).

[TACL] Parsing Algebraic Word Problems into Equations

Rik Koncel-Kedziorski, Hannaneh Hajishirzi, Ashish Sabharwal, Oren Etzioni, and Siena Dumas Ang

11:45–12:05

This paper formalizes the problem of solving multi-sentence algebraic word problems as that of generating and scoring equation trees. We use integer linear programming to generate equation trees and score their likelihood by learning local and global discriminative models. These models are trained on a small set of word problems and their answers, without any manual annotation, in order to choose the equation that best matches the problem text. We refer to the overall system as ALGES. We compare ALGES with previous work and show that it covers the full gamut of arithmetic operations whereas Hosseini et al. (2014) only handle addition and subtraction. In addition, ALGES overcomes the brittleness of the Kushman et al. (2014) approach on single-equation problems, yielding a 15% to 50% reduction in error.

This is how we do it: Answer Reranking for Open-domain How Questions with Paragraph Vectors and Minimal Feature Engineering

Dasha Bogdanova and Jennifer Foster

12:05–12:15

We present a simple yet powerful approach to non-factoid answer reranking whereby question-answer pairs are represented by concatenated distributed representation vectors and a multilayer perceptron is used to compute the score for an answer. Despite its simplicity, our approach achieves state-of-the-art performance on a public dataset of How questions, outperforming systems which employ sophisticated feature sets. We attribute this good performance to the use of paragraph instead of word vector representations and to the use of suitable data

for training these representations.

Parallel Session 8B: Multilingual Processing

Grande Ballroom B

Chair: Mohit Bansal

Multilingual Language Processing From Bytes

Dan Gillick, Cliff Brunk, Oriol Vinyals, and Amarnag Subramanya

10:45–11:05

We describe an LSTM-based model which we call Byte-to-Span (BTS) that reads text as bytes and outputs span annotations of the form [start, length, label] where start positions, lengths, and labels are separate entries in our vocabulary. Because we operate directly on unicode bytes rather than language-specific words or characters, we can analyze text in many languages with a single model. Due to the small vocabulary size, these multilingual models are very compact, but produce results similar to or better than the state-of-the-art in Part-of-Speech tagging and Named Entity Recognition that use only the provided training datasets (no external data sources). Our models are learning “from scratch” in that they do not rely on any elements of the standard pipeline in Natural Language Processing (including tokenization), and thus can run in standalone fashion on raw text.

Ten Pairs to Tag – Multilingual POS Tagging via Coarse Mapping between Embeddings

Yuan Zhang, David Gaddy, Regina Barzilay, and Tommi Jaakkola

11:05–11:25

In the absence of annotations in the target language, multilingual models typically draw on extensive parallel resources. In this paper, we demonstrate that accurate multilingual part-of-speech (POS) tagging can be done with just a few (e.g., ten) word translation pairs. We use the translation pairs to establish a coarse linear isometric (orthonormal) mapping between monolingual embeddings. This enables the supervised source model expressed in terms of embeddings to be used directly on the target language. We further refine the model in an unsupervised manner by initializing and regularizing it to be close to the direct transfer model. Averaged across six languages, our model yields a 37.5% absolute improvement over the monolingual prototype-driven method (Haghghi and Klein, 2006) when using a comparable amount of supervision. Moreover, to highlight key linguistic characteristics of the generated tags, we use them to predict typological properties of languages, obtaining a 50% error reduction relative to the prototype model.

Part-of-Speech Tagging for Historical English

Yi Yang and Jacob Eisenstein

11:25–11:45

As more historical texts are digitized, there is interest in applying natural language processing tools to these archives. However, the performance of these tools is often unsatisfactory, due to language change and genre differences. Spelling normalization heuristics are the dominant solution for dealing with historical texts, but this approach fails to account for changes in usage and vocabulary. In this empirical paper, we assess the capability of domain adaptation techniques to cope with historical texts, focusing on the classic benchmark task of part-of-speech tagging. We evaluate several domain adaptation methods on the task of tagging Early Modern English and Modern British English texts in the Penn Corpora of Historical English. We demonstrate that the Feature Embedding method for unsupervised domain adaptation outperforms word embeddings and Brown clusters, showing the importance of embedding the entire feature space, rather than just individual words. Feature Embeddings also give better performance than spelling normalization, but the combination of the two methods is better still, yielding a 5% raw improvement in tagging accuracy on Early Modern English texts.

Statistical Modeling of Creole Genesis

Yugo Murawaki

11:45–12:05

Creole languages do not fit into the traditional tree model of evolutionary history because multiple languages are involved in their formation. In this paper, we present several statistical models to explore the nature of creole genesis. After reviewing quantitative studies on creole genesis, we first tackle the question of whether creoles are typologically distinct from non-creoles. By formalizing this question as a binary classification problem, we demonstrate that a linear classifier fails to separate creoles from non-creoles although the two groups have substantially different distributions in the feature space. We then model a creole language as a mixture of source languages plus a special restructured. We find a pervasive influence of the restructured in creole genesis and some statistical universals in it, paving the way for more elaborate statistical models.

Shallow Parsing Pipeline - Hindi-English Code-Mixed Social Media Text

Arnav Sharma, Sakshi Gupta, Raveesh Motlani, Piyush Bansal, Manish Shrivastava, Radhika Mamidi, and Dipti Misra Sharma

12:05–12:15

In this study, the problem of shallow parsing of Hindi-English code-mixed social media text (CSMT) has been addressed. We have annotated the data, developed a language identifier, a normalizer, a part-of-speech tagger and a shallow parser. To the best of our knowledge, we are the first to attempt shallow parsing on CSMT. The pipeline developed has been made available to the research community with the goal of enabling better text analysis of Hindi English CSMT. The pipeline is accessible at [<http://bit.ly/csmt-parser-api>].

Parallel Session 8C: Word Embeddings II

Grande Ballroom C

Chair: Hinrich Schütze

Bilingual Learning of Multi-sense Embeddings with Discrete Autoencoders

Simon Šuster, Ivan Titov, and Gertjan van Noord

10:45–11:05

We present an approach to learning multi-sense word embeddings relying both on monolingual and bilingual information. Our model consists of an encoder, which uses monolingual and bilingual context (i.e. a parallel sentence) to choose a sense for a given word, and a decoder which predicts context words based on the chosen sense. The two components are estimated jointly. We observe that the word representations induced from bilingual data outperform the monolingual counterparts across a range of evaluation tasks, even though crosslingual information is not available at test time.

Polyglot Neural Language Models: A Case Study in Cross-Lingual Phonetic Representation Learning

Julia Tsvetkov, Sunayana Sitaram, Manaal Faruqui, Guillaume Lample, Patrick Littell, David Mortensen, Alan W Black, Lori Levin, and Chris Dyer

11:05–11:25

We introduce polyglot language models, recurrent neural network models trained to predict symbol sequences in many different languages using shared representations of symbols and conditioning on typological information about the language to be predicted. We apply these to the problem of modeling phone sequences—a domain in which universal symbol inventories and cross-linguistically shared feature representations are a natural fit. Intrinsic evaluation on held-out perplexity, qualitative analysis of the learned representations, and extrinsic evaluation in two downstream applications that make use of phonetic features show (i) that polyglot models better generalize to held-out data than comparable monolingual models and (ii) that polyglot phonetic feature representations are of higher quality than those learned monolingually.

Learning Distributed Representations of Sentences from Unlabelled Data

Felix Hill, Kyunghyun Cho, and Anna Korhonen

11:25–11:45

Unsupervised methods for learning distributed representations of words are ubiquitous in today's NLP research, but far less is known about the best ways to learn distributed phrase or sentence representations from unlabelled data. This paper is a systematic comparison of models that learn such representations. We find that the optimal approach depends critically on the intended application. Deeper, more complex models are preferable for representations to be used in supervised systems, but shallow log-linear models work best for building representation spaces that can be decoded with simple spatial distance metrics. We also propose two new unsupervised representation-learning objectives designed to optimise the trade-off between training time, domain portability and performance.

[TACL] Learning to Understand Phrases by Embedding the Dictionary

Felix Hill, Kyunghyun Cho, Anna Korhonen, and Yoshua Bengio

11:45–12:05

Distributional models that learn rich semantic word representations are a success story of recent NLP research. However, developing models that learn useful representations of phrases and sentences has proved far harder. We propose using the definitions found in everyday dictionaries as a means of bridging this gap between lexical and phrasal semantics. Neural language embedding models can be effectively trained to map dictionary definitions (phrases) to (lexical) representations of the words defined by those definitions. We present two applications of these architectures ‘reverse dictionaries’ that return the name of a concept given a definition or description and general-knowledge crossword question answerers. On both tasks, neural language embedding models trained on definitions from a handful of freely-available lexical resources perform as well or better than existing commercial systems that rely on significant task-specific engineering. The results highlight the effectiveness of both neural embedding architectures and definition-based training for developing models that understand phrases and sentences.

Retrofitting Sense-Specific Word Vectors Using Parallel Text

Allyson Ettinger, Philip Resnik, and Marine Carpuat

12:05–12:15

Jauhar et al. (2015) recently proposed to learn sense-specific word representations by “retrofitting” standard distributional word representations to an existing ontology. We observe that this approach does not require an ontology, and can be generalized to any graph defining word senses and relations between them. We create such a graph using translations learned from parallel corpora. On a set of lexical semantic tasks, representations learned using parallel text perform roughly as well as those derived from WordNet, and combining the two

representation types significantly improves performance.

NAACL Business Meeting

Date: Wednesday, June 15, 2016

Time: 1:00–2:00 PM

Venue: Grande Ballroom A

All attendees are encouraged to participate in the business meeting.

Session 9 Overview – Wednesday, June 15, 2016

Track A	Track B	Track C	
<i>Argumentation and Discourse</i> Grande Ballroom A	<i>Misc Semantics</i> Grande Ballroom B	<i>Text Categorization</i> Grande Ballroom C	
End-to-End Argumentation Mining in Student Essays <i>Persing and Ng</i>	Automatic Generation and Scoring of Positive Interpretations from Negated Statements <i>Blanco and Sarabi</i>	Hierarchical Attention Networks for Document Classification <i>Yang, Yang, Dyer, He, Smola, and Hovy</i>	2:15
Cross-Domain Mining of Argumentative Text through Distant Supervision <i>Al-Khatib, Wachsmuth, Hagen, Köhler, and Stein</i>	Learning Natural Language Inference with LSTM <i>Wang and Jiang</i>	Dependency Based Embeddings for Sentence Classification Tasks <i>Komninos and Manandhar</i>	2:35
A Study of the Impact of Persuasive Argumentation in Political Debates <i>Cano-Basave and He</i>	Activity Modeling in Email <i>Qadir, Gamon, Pantel, and Awadallah</i>	Deep LSTM based Feature Mapping for Query Classification <i>Shi, Yao, Tian, and Jiang</i>	2:55
Lexical Coherence Graph Modeling Using Word Embeddings <i>Mesgar and Strube</i>	Clustering Paraphrases by Word Sense <i>Cocos and Callison-Burch</i>	Dependency Sensitive Convolutional Neural Networks for Modeling Sentences and Documents <i>Zhang, Lee, and Radev</i>	3:15
Using Context to Predict the Purpose of Argumentative Writing Revisions <i>Zhang and Litman</i>	Unsupervised Learning of Prototypical Fillers for Implicit Semantic Role Labeling <i>Schenk and Chiarcos</i>	MGNC-CNN: A Simple Approach to Exploiting Multiple Word Embeddings for Sentence Classification <i>Zhang, Roller, and Wallace</i>	3:35

Parallel Session 9

Parallel Session 9A: Argumentation and Discourse

Grande Ballroom A

Chair: Cristian Danescu-Niculescu-Mizil

End-to-End Argumentation Mining in Student Essays

Isaac Persing and Vincent Ng

14:15–14:35

Understanding the argumentative structure of a persuasive essay involves addressing two challenging tasks: identifying the components of the essay's argument and identifying the relations that occur between them. We examine the under-investigated task of end-to-end argument mining in persuasive student essays, where we (1) present the first results on end-to-end argument mining in student essays using a pipeline approach; (2) address error propagation inherent in the pipeline approach by performing joint inference over the outputs of the tasks in an Integer Linear Programming (ILP) framework; and (3) propose a novel objective function that enables F-score to be maximized directly by an ILP solver. We evaluate our joint-inference approach with our novel objective function on a publicly-available corpus of 90 essays, where it yields an 18.5% relative error reduction in F-score over the pipeline system.

Cross-Domain Mining of Argumentative Text through Distant Supervision

Khalid Al-Khatib, Henning Wachsmuth, Matthias Hagen, Jonas Köhler, and Benno Stein 14:35–14:55

Argumentation mining is considered as a key technology for future search engines and automated decision making. In such applications, argumentative text segments have to be mined from large and diverse document collections. However, most existing argumentation mining approaches tackle the classification of argumentativeness only for a few manually annotated documents from narrow domains and registers. This limits their practical applicability. We hence propose a distant supervision approach that acquires argumentative text segments automatically from online debate portals. Experiments across domains and registers show that training on such a corpus improves the effectiveness and robustness of mining argumentative text. We freely provide the underlying corpus for research.

A Study of the Impact of Persuasive Argumentation in Political Debates

Amparo Elizabeth Cano-Basave and Yulan He

14:55–15:15

Persuasive communication is the process of shaping, reinforcing and changing others' responses. In political debates, speakers express their views towards the debated topics through not only the choice of discourse content but also an argumentation process. In this work we study the use of semantic frames for modelling argumentation in speakers' discourse. We investigate the impact of a speaker's argumentation style and their effect in influencing an audience in supporting their candidature. We model the influence index of each candidate based on results of national polls occurring immediately after a debate took place and present a system which ranks speakers in terms of their relative influence based on a combination of content and persuasive argumentation features. Our results show that although content alone is predictive of a speaker's influence rank, persuasive argumentation also affects such indices.

Lexical Coherence Graph Modeling Using Word Embeddings

Mohsen Mesgar and Michael Strube

15:15–15:35

Coherence is established by semantic connections between sentences of a text which can be modeled by lexical relations. In this paper, we introduce the lexical coherence graph (LCG), a new graph-based model to represent lexical relations among sentences. The frequency of subgraphs (coherence patterns) of this graph captures the connectivity style of sentence nodes in this graph. The coherence of a text is encoded by a vector of these frequencies. We evaluate the LCG model on the readability ranking task. The results of the experiments show that the LCG model obtains higher accuracy than state-of-the-art coherence models. Using larger subgraphs yields higher accuracy because they capture more structural information. However, larger subgraphs can be sparse. We adapt Kneser-Ney smoothing to smooth subgraphs' frequencies. Smoothing improves performance.

Using Context to Predict the Purpose of Argumentative Writing Revisions

Fan Zhang and Diane Litman

15:35–15:45

While there is increasing interest in automatically recognizing the argumentative structure of a text, recognizing the argumentative purpose of revisions to such texts has been less explored. Furthermore, existing revision classification approaches typically ignore contextual information. We propose two approaches for utilizing

contextual information when predicting argumentative revision purposes: developing contextual features for use in the classification paradigm of prior work, and transforming the classification problem to a sequence labeling task. Experimental results using two corpora of student essays demonstrate the utility of contextual information for predicting argumentative revision purposes.

Parallel Session 9B: Misc Semantics

Grande Ballroom B

Chair: Steven Bethard

Automatic Generation and Scoring of Positive Interpretations from Negated Statements

Eduardo Blanco and Zahra Sarabi

14:15–14:35

This paper presents a methodology to extract positive interpretations from negated statements. First, we automatically generate plausible interpretations using well-known grammar rules and manipulating semantic roles. Second, we score plausible alternatives according to their likelihood. Manual annotations show that the positive interpretations are intuitive to humans, and experimental results show that the scoring task can be automated.

Learning Natural Language Inference with LSTM

Shuohang Wang and Jing Jiang

14:35–14:55

Natural language inference (NLI) is a fundamentally important task in natural language processing that has many applications. The recently released Stanford Natural Language Inference (SNLI) corpus has made it possible to develop and evaluate learning-centered methods such as deep neural networks for natural language inference (NLI). In this paper, we propose a special long short-term memory (LSTM) architecture for NLI. Our model builds on top of a recently proposed neural attention model for NLI but is based on a significantly different idea. Instead of deriving sentence embeddings for the premise and the hypothesis to be used for classification, our solution uses a match-LSTM to perform word-by-word matching of the hypothesis with the premise. This LSTM is able to place more emphasis on important word-level matching results. In particular, we observe that this LSTM remembers important mismatches that are critical for predicting the contradiction or the neutral relationship label. On the SNLI corpus, our model achieves an accuracy of 86.1%, outperforming the state of the art.

Activity Modeling in Email

Ashequl Qadir, Michael Gamon, Patrick Pantel, and Ahmed Hassan Awadallah

14:55–15:15

We introduce a latent activity model for workplace emails, positing that communication at work is purposeful and organized by activities. We pose the problem as probabilistic inference in graphical models that jointly capture the interplay between latent activities and the email contexts they govern, such as the recipients, subject and body. The model parameters are learned using maximum likelihood estimation with an expectation maximization algorithm. We present three variants of the model that incorporate the recipients, co-occurrence of the recipients, and email body and subject. We demonstrate the model's effectiveness in an email recipient recommendation task and show that it outperforms a state-of-the-art generative model. Additionally, we show that the activity model can be used to identify email senders who engage in similar activities, resulting in further improvements in recipient recommendation.

Clustering Paraphrases by Word Sense

Anne Cocos and Chris Callison-Burch

15:15–15:35

Automatically generated databases of English paraphrases have the drawback that they return a single list of paraphrases for an input word or phrase. This means that all senses of polysemous words are grouped together, unlike WordNet which partitions different senses into separate synsets. We present a new method for clustering paraphrases by word sense, and apply it to the Paraphrase Database (PPDB). We investigate the performance of hierarchical and spectral clustering algorithms, and systematically explore different ways of defining the similarity matrix that they use as input. Our method produces sense clusters that are qualitatively and quantitatively good, and that represent a substantial improvement to the PPDB resource.

Unsupervised Learning of Prototypical Fillers for Implicit Semantic Role Labeling

Niko Schenk and Christian Chiaro

15:35–15:45

Gold annotations for supervised implicit semantic role labeling are extremely sparse and costly. As a lightweight alternative, this paper describes an approach based on unsupervised parsing which can do without iSRL-specific training data: We induce prototypical roles from large amounts of explicit SRL annotations paired with their distributed word representations. An evaluation shows competitive performance with supervised methods on the SemEval 2010 data, and our method can easily be applied to predicates (or languages) for which no training annotations are available.

Parallel Session 9C: Text Categorization

Grande Ballroom C

Chair: Jacob Eisenstein

Hierarchical Attention Networks for Document Classification*Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy* 14:15–14:35

We propose a hierarchical attention network for document classification. Our model has two distinctive characteristics: (i) it has a hierarchical structure that mirrors the hierarchical structure of documents; (ii) it has two levels of attention mechanisms applied at the word- and sentence-level, enabling it to attend differentially to more and less important content when constructing the document representation. Experiments conducted on six large scale text classification tasks demonstrate that the proposed architecture outperform previous methods by a substantial margin. Visualization of the attention layers illustrates that the model selects qualitatively informative words and sentences.

Dependency Based Embeddings for Sentence Classification Tasks*Alexandros Komninos and Suresh Manandhar*

14:35–14:55

We compare different word embeddings from a standard window based skipgram model, a skipgram model trained using dependency context features and a novel skipgram variant that utilizes additional information from dependency graphs. We explore the effectiveness of the different types of word embeddings for word similarity and sentence classification tasks. We consider three common sentence classification tasks: question type classification on the TREC dataset, binary sentiment classification on Stanford's Sentiment Treebank and semantic relation classification on the SemEval 2010 dataset. For each task we use three different classification methods: a Support Vector Machine, a Convolutional Neural Network and a Long Short Term Memory Network. Our experiments show that dependency based embeddings outperform standard window based embeddings in most of the settings, while using dependency context embeddings as additional features improves performance in all tasks regardless of the classification method.

Deep LSTM based Feature Mapping for Query Classification*Yangyang Shi, Kaisheng Yao, Le Tian, and Daxin Jiang*

14:55–15:15

Traditional convolutional neural network (CNN) based query classification uses linear feature mapping in its convolution operation. The recurrent neural network (RNN), differs from a CNN in representing word sequence with their ordering information kept explicitly. We propose using a deep long-short-term-memory (DLSTM) based feature mapping to learn feature representation for CNN. The DLSTM, which is a stack of LSTM units, has different order of feature representations at different depth of LSTM unit. The bottom LSTM unit equipped with input and output gates, extracts the first order feature representation from current word position. To extract higher order nonlinear feature representation, the LSTM unit at higher position gets input from two parts. First part is the lower LSTM unit's memory cell from previous word. Second part is the lower LSTM unit's hidden output from current word. In this way, taking advantage of different types of gates in LSTM unit, the DLSTM captures the nonlinear nonconsecutive interaction within n-grams. Using an architecture that combines a stack of the DLSTM layers with a tradition CNN layer, we have observed new state-of-the-art query classification accuracy on benchmark data sets for query classification.

Dependency Sensitive Convolutional Neural Networks for Modeling Sentences and Documents*Rui Zhang, Honglak Lee, and Dragomir R. Radev*

15:15–15:35

The goal of sentence and document modeling is to accurately represent the meaning of sentences and documents for various Natural Language Processing tasks. In this work, we present Dependency Sensitive Convolutional Neural Networks (DSCNN) as a general-purpose classification system for both sentences and documents. DSCNN hierarchically builds textual representations by processing pretrained word embeddings via Long Short-Term Memory networks and subsequently extracting features with convolution operators. Compared with existing recursive neural models with tree structures, DSCNN does not rely on parsers and expensive phrase labeling, and thus is not restricted to sentence-level tasks. Moreover, unlike other CNN-based models that analyze sentences locally by sliding windows, our system captures both the dependency information within each sentence and relationships across sentences in the same document. Experiment results demonstrate that our approach is achieving state-of-the-art performance on several tasks, including sentiment analysis, question type classification, and subjectivity classification.

MGNC-CNN: A Simple Approach to Exploiting Multiple Word Embeddings for Sentence Classification

Ye Zhang, Stephen Roller, and Byron C. Wallace

15:35–15:45

We introduce a novel, simple convolution neural network (CNN) architecture - multi-group norm constraint CNN (MGNC-CNN)—that capitalizes on multiple sets of word embeddings for sentence classification. MGNC-CNN extracts features from input embedding sets independently and then joins these at the penultimate layer in the network to form a final feature vector. We then adopt a group regularization strategy that differentially penalizes weights associated with the subcomponents generated from the respective embedding sets. This model is much simpler than comparable alternative architectures and requires substantially less training time. Furthermore, it is flexible in that it does not require input word embeddings to be of the same dimensionality. We show that MGNC-CNN consistently outperforms baseline models.

Best paper awards

Time: 4:15–5:45

Chair: Owen Rambow

Location: Pavilion

Improving sentence compression by learning to predict gaze

Sigrid Klerke, Yoav Goldberg, and Anders Søgaard

We show how eye-tracking corpora can be used to improve sentence compression models, presenting a novel multi-task learning algorithm based on multi-layer LSTMs. We obtain performance competitive with state-of-the-art approaches to sentence compression.

Feuding Families and Former Friends: Unsupervised Learning for Dynamic Fictional Relationships

Mohit Iyyer, Anupam Guha, Snigdha Chaturvedi, Jordan Boyd-Graber, and Hal Daumé III

Understanding how a fictional relationship between two characters changes over time (e.g., from best friends to sworn enemies) is a key challenge in digital humanities scholarship. We present a novel unsupervised neural network for this task that incorporates dictionary learning to generate interpretable, accurate relationship trajectories. While previous work on characterizing literary relationships relies on plot summaries annotated with predefined labels, our model jointly learns a set of global relationship descriptors as well as a trajectory over these descriptors for each relationship in a dataset of raw text from novels. We find that our model learns descriptors of events (e.g., marriage or murder) as well as interpersonal states (love, sadness). Our model outperforms topic model baselines on two crowdsourced tasks, and we also find interesting correlations to annotations in an existing dataset.

Learning to Compose Neural Networks for Question Answering

Jacob Andreas, Marcus Rohrbach, Trevor Darrell, and Dan Klein

We describe a question answering model that applies to both natural images and structured knowledge bases. The model uses natural language strings to automatically assemble deep networks from a collection of composable neural modules. Parameters for these modules are learned jointly with network-assembly parameters via reinforcement learning, with only (word, question, answer) triples as supervision. Our approach, which we term a _dynamic neural module network_, achieves state-of-the-art results on benchmark datasets in both visual and structured domains.

6

Workshops

Thursday–Friday

Seabreeze 1/2	SemEval: 10th International Workshop on Semantic Evaluation	p.105
---------------	---	-------

Thursday

Marina 2	Human-Computer Question Answering	p.119
Spinnaker 2	Fifth Workshop on Computational Linguistics for Literature (clfl2016)	p.121
Executive 4	Computational Linguistics and Clinical Psychology — From Linguistic Signal to Clinical Reality (CLPsych)	p.122
Marina 3	7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA 2016)	p.125
Executive 3AB	11th Workshop on Innovative Use of NLP for Building Educational Applications (BEA11)	p.127
Spinnaker 1	2nd Workshop on Semantics-Driven Machine Translation (SedMT2016)	p.130
Marina 1	Coreference Resolution Beyond OntoNotes Excursion	p.131
MCAS Miramar		p.133

Friday

Spinnaker 1	Workshop on Computational Approaches to Deception Detection	p.134
Spinnaker 2	Workshop on Discontinuous Structures in Natural Language Processing	p.135
Marina 1	The 4th Workshop on EVENTS: Definition, Detection, Coreference, and Representation	p.136
Marina 2	The 4th Workshop on Metaphor in NLP (Meta4NLP)	p.137
Marina 3	Workshop on Multilingual and Crosslingual Methods in NLP	p.138
Executive 3AB	AKBC: 5th Workshop on Automated Knowledge Base Construction	p.140
Executive 4	TextGraphs: 10th Workshop on Graph-based Methods for Natural Language Processing	p.142

Workshop 1: SemEval: 10th International Workshop on Semantic Evaluation

Organizers: *Steven Bethard, Daniel Cer, Marine Carpuat, David Jurgens, Preslav Nakov, and Torsten Zesch*

Venue: Seabreeze 1/2

Thursday, June 16, 2016

Welcome

9:00–9:15 **Opening Remarks (SemEval organizers)**

9:15–10:30 **Sentiment Analysis**

9:15–9:30 SemEval-2016 Task 4: Sentiment Analysis in Twitter
Preslav Nakov, Alan Ritter, Sara Rosenthal, Fabrizio Sebastiani, and Veselin Stoyanov

9:30–9:45 SemEval-2016 Task 5: Aspect Based Sentiment Analysis
Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammad AL-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphee De Clercq, Veronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia Loukachevitch, Evgeniy Kotelnikov, Núria Bel, Salud María Jiménez-Zafra, and Gülsen Eryiğit

9:45–10:00 SemEval-2016 Task 6: Detecting Stance in Tweets
Saif Mohammad, Svetlana Kiritchenko, Parinaz Sobhani, Xiaodan Zhu, and Colin Cherry

10:00–10:15 SemEval-2016 Task 7: Determining Sentiment Intensity of English and Arabic Phrases
Svetlana Kiritchenko, Saif Mohammad, and Mohammad Salameh

10:15–10:30 **Sentiment Analysis Discussion (Task Organizers)**

10:30–11:00 **Coffee Break**

11:00–12:30 **Poster Session: Sentiment Analysis**

- CUFE at SemEval-2016 Task 4: A Gated Recurrent Model for Sentiment Classification
Mahmoud Nabil, Amir Atyia, and Mohamed Aly
- QCRI at SemEval-2016 Task 4: Probabilistic Methods for Binary and Ordinal Quantification
Giovanni Da San Martino, Wei Gao, and Fabrizio Sebastiani
- SteM at SemEval-2016 Task 4: Applying Active Learning to Improve Sentiment Classification
Stefan Räbiger, Mishal Kazmi, Yücel Saygin, Peter Schüller, and Myra Spiliopoulou
- I2RNTU at SemEval-2016 Task 4: Classifier Fusion for Polarity Classification in Twitter
Zhengchen Zhang, Chen Zhang, Dongyan Huang, Weisi Lin, and Minghui Dong

- LyS at SemEval-2016 Task 4: Exploiting Neural Activation Values for Twitter Sentiment Classification and Quantification
David Vilares, Yerai Doval, Miguel A. Alonso, and Carlos Gómez-Rodríguez
- TwiSE at SemEval-2016 Task 4: Twitter Sentiment Classification
Georgios Balikas and Massih-Reza Amini
- ISTI-CNR at SemEval-2016 Task 4: Quantification on an Ordinal Scale
Andrea Esuli
- aueb.twitter.sentiment at SemEval-2016 Task 4: A Weighted Ensemble of SVMs for Twitter Sentiment Analysis
Stavros Giorgis, Apostolos Rousas, John Pavlopoulos, Prodromos Malakasiotis, and Ion Androutsopoulos
- thecerealkiller at SemEval-2016 Task 4: Deep Learning based System for Classifying Sentiment of Tweets on Two Point Scale
Vikrani Yadav
- NTNUSentEval at SemEval-2016 Task 4: Combining General Classifiers for Fast Twitter Sentiment Analysis
Brage Ekroll Jahren, Valerij Fredriksen, Björn Gambäck, and Lars Bungum
- UDLAP at SemEval-2016 Task 4: Sentiment Quantification Using a Graph Based Representation
Esteban Castillo, Ofelia Cervantes, Darnes Vilariño, and David Báez
- GTI at SemEval-2016 Task 4: Training a Naive Bayes Classifier using Features of an Unsupervised System
Jonathan Juncal-Martínez, Tamara Álvarez-López, Milagros Fernández-Gavilanes, Enrique Costa-Montenegro, and Francisco Javier González-Castaño
- Aicyber at SemEval-2016 Task 4: i-vector based sentence representation
Steven Du and Xi Zhang
- SwissCheese at SemEval-2016 Task 4: Sentiment Classification Using an Ensemble of Convolutional Neural Networks with Distant Supervision
Jan Deriu, Maurice Gonzenbach, Fatih Uzdilli, Aurelien Lucchi, Valeria De Luca, and Martin Jaggi
- PUT at SemEval-2016 Task 4: The ABC of Twitter Sentiment Analysis
Mateusz Lango, Dariusz Brzezinski, and Jerzy Stefanowski
- mib at SemEval-2016 Task 4a: Exploiting lexicon based features for Sentiment Analysis in Twitter
Vittoria Cozza and Marinella Petrocchi
- MDSENT at SemEval-2016 Task 4: A Supervised System for Message Polarity Classification
Hang Gao and Tim Oates
- CICBUAPnlp at SemEval-2016 Task 4-A: Discovering Twitter Polarity using Enhanced Embeddings
Helena Gomez, Darnes Vilariño, Grigori Sidorov, and David Pinto Avendaño
- Finki at SemEval-2016 Task 4: Deep Learning Architecture for Twitter Sentiment Analysis
Dario Stojanovski, Gjorgji Strezoski, Gjorgji Madjarov, and Ivica Dimitrovski
- Tweester at SemEval-2016 Task 4: Sentiment Analysis in Twitter Using Semantic-Affective Model Adaptation
Elisavet Palogiannidi, Athanasia Kolovou, Fenia Christopoulou, Filippos Kokkinos, Elias Iosif, Nikolaos Malandrakis, Haris Papageorgiou, Shrikanth Narayanan, and Alexandros Potamianos

- UofL at SemEval-2016 Task 4: Multi Domain word2vec for Twitter Sentiment Classification
Omar Abdelwahab and Adel Elmaghraby
- NRU-HSE at SemEval-2016 Task 4: Comparative Analysis of Two Iterative Methods Using Quantification Library
Nikolay Karpov, Alexander Porshnev, and Kirill Rudakov
- INSIGHT-1 at SemEval-2016 Task 4: Convolutional Neural Networks for Sentiment Classification and Quantification
Sebastian Ruder, Parsa Ghaffari, and John G. Breslin
- UNIMELB at SemEval-2016 Tasks 4A and 4B: An Ensemble of Neural Networks and a Word2Vec Based Model for Sentiment Classification
Steven Xu, Huizhi Liang, and Timothy Baldwin
- SentiSys at SemEval-2016 Task 4: Feature-Based System for Sentiment Analysis in Twitter
Hussam Hamdan
- DSIC-ELIRF at SemEval-2016 Task 4: Message Polarity Classification in Twitter using a Support Vector Machine Approach
Victor Martínez Morant, Lluís-F Hurtado, and Ferran Pla
- SENSEI-LIF at SemEval-2016 Task 4: Polarity embedding fusion for robust sentiment analysis
Mickael Rouvier and Benoit Favre
- DiegoLab16 at SemEval-2016 Task 4: Sentiment Analysis in Twitter using Centroids, Clusters, and Sentiment Lexicons
Abeed Sarker and Graciela Gonzalez
- VCU-TSA at Semeval-2016 Task 4: Sentiment Analysis in Twitter
Gerard Briones, Kasun Amarasinghe, and Bridget McInnes
- UniPI at SemEval-2016 Task 4: Convolutional Neural Networks for Sentiment Classification
Giuseppe Attardi and Daniele Sartiano
- IIP at SemEval-2016 Task 4: Prioritizing Classes in Ensemble Classification for Sentiment Analysis of Tweets
Jasper Friedrichs
- PotTS at SemEval-2016 Task 4: Sentiment Analysis of Twitter Using Character-level Convolutional Neural Networks.
Uladzimir Sidarenka
- INESC-ID at SemEval-2016 Task 4-A: Reducing the Problem of Out-of-Embedding Words
Silvio Amir, Ramón Astudillo, Wang Ling, Mario J. Silva, and Isabel Trancoso
- SentimentalITsts at SemEval-2016 Task 4: building a Twitter sentiment analyzer in your backyard
Cosmin Floorean, Oana Bejenaru, Eduard Apostol, Octavian Ciobanu, Adrian Iftene, and Diana Trandabat
- Minions at SemEval-2016 Task 4: or how to build a sentiment analyzer using off-the-shelf resources?
Calin-Cristian Ciubotariu, Marius-Valentin Hrisca, Mihail Gliga, Diana Darabana, Diana Trandabat, and Adrian Iftene
- YZU-NLP Team at SemEval-2016 Task 4: Ordinal Sentiment Classification Using a Recurrent Convolutional Network
Yunchao He, Liang-Chih Yu, Chin-Sheng Yang, K. Robert Lai, and Weiyi Liu
- ECNU at SemEval-2016 Task 4: An Empirical Investigation of Traditional NLP Features and Word Embedding Features for Sentence-level and Topic-level Sentiment Analysis in Twitter
Yunxiao Zhou, Zhihua Zhang, and Man Lan

- OPAL at SemEval-2016 Task 4: the Challenge of Porting a Sentiment Analysis System to the “Real” World
Alexandra Balahur
 - Know-Center at SemEval-2016 Task 5: Using Word Vectors with Typed Dependencies for Opinion Target Expression Extraction
Stefan Falk, Andi Rexha, and Roman Kern
 - NileTMRG at SemEval-2016 Task 5: Deep Convolutional Neural Networks for Aspect Category and Sentiment Extraction
Talaat Khalil and Samhaa R. El-Beltagy
 - XRCE at SemEval-2016 Task 5: Feedbacked Ensemble Modeling on Syntactic-Semantic Knowledge for Aspect Based Sentiment Analysis
Caroline Brun, Julien Perez, and Claude Roux
 - NLANGP at SemEval-2016 Task 5: Improving Aspect Based Sentiment Analysis using Neural Network Features
Zhiqiang Toh and Jian Su
 - bunji at SemEval-2016 Task 5: Neural and Syntactic Models of Entity-Attribute Relationship for Aspect-based Sentiment Analysis
Toshihiko Yanase, Kohsuke Yanai, Misa Sato, Toshinori Miyoshi, and Yoshiki Niwa
 - IHS-RD-Belarus at SemEval-2016 Task 5: Detecting Sentiment Polarity Using the Heatmap of Sentence
Maryna Chernyshevich
 - BUTknot at SemEval-2016 Task 5: Supervised Machine Learning with Term Substitution Approach in Aspect Category Detection
Jakub Machacek
 - IIT-TUDA at SemEval-2016 Task 5: Beyond Sentiment Lexicon: Combining Domain Dependency and Distributional Semantics Features for Aspect Based Sentiment Analysis
Ayush Kumar, Sarah Kohail, Amit Kumar, Asif Ekbal, and Chris Biemann
 - GTI at SemEval-2016 Task 5: SVM and CRF for Aspect Detection and Unsupervised Aspect-Based Sentiment Analysis
Tamara Álvarez-López, Jonathan Juncal-Martínez, Milagros Fernández-Gavilanes, Enrique Costa-Montenegro, and Francisco Javier González-Castaño
 - AUEB-ABSA at SemEval-2016 Task 5: Ensembles of Classifiers and Embeddings for Aspect Based Sentiment Analysis
Dionysios Xenos, Panagiotis Theodorakakos, John Pavlopoulos, Prodromos Malakasiotis, and Ion Androutsopoulos
 - AKTSKI at SemEval-2016 Task 5: Aspect Based Sentiment Analysis for Consumer Reviews
Shubham Pateria and Prafulla Choubeiy
 - MayAnd at SemEval-2016 Task 5: Syntactic and word2vec-based approach to aspect-based polarity detection in Russian
Vladimir Mayorov and Ivan Andrianov
 - INSIGHT-1 at SemEval-2016 Task 5: Deep Learning for Multilingual Aspect-based Sentiment Analysis
Sebastian Ruder, Parsa Ghaffari, and John G. Breslin
 - TGB at SemEval-2016 Task 5: Multi-Lingual Constraint System for Aspect Based Sentiment Analysis
Fatih Samet Çetin, Ezgi Yıldırım, Can Özbel, and Gülsen Eryiğit
 - UWB at SemEval-2016 Task 5: Aspect Based Sentiment Analysis
Tomáš Hercig, Tomáš Brychcín, Lukáš Svoboda, and Michal Konkol
 - SentiSys at SemEval-2016 Task 5: Opinion Target Extraction and Sentiment Polarity Detection
Hussam Hamdan
-

- COMMIT at SemEval-2016 Task 5: Sentiment Analysis with Rhetorical Structure Theory
Kim Schouten and Flavius Frasincar
- ECNU at SemEval-2016 Task 5: Extracting Effective Features from Relevant Fragments in Sentence for Aspect-Based Sentiment Analysis in Reviews
Mengxiao Jiang, Zhihua Zhang, and Man Lan
- UFAL at SemEval-2016 Task 5: Recurrent Neural Networks for Sentence Classification
Aleš Tamchyna and Kateřina Veselovská
- UWATERLOO at SemEval-2016 Task 5: Minimally Supervised Approaches to Aspect-Based Sentiment Analysis
Olga Vechtomova and Anni He
- INF-UFRGS-OPINION-MINING at SemEval-2016 Task 6: Automatic Generation of a Training Corpus for Unsupervised Identification of Stance in Tweets
Marcelo Dias and Karin Becker
- pkudblab at SemEval-2016 Task 6 : A Specific Convolutional Neural Network System for Effective Stance Detection
Wan Wei, Xiao Zhang, Xuqin Liu, Wei Chen, and Tengjiao Wang
- USFD at SemEval-2016 Task 6: Any-Target Stance Detection on Twitter with Autoencoders
Isabelle Augenstein, Andreas Vlachos, and Kalina Bontcheva
- IUCL at SemEval-2016 Task 6: An Ensemble Model for Stance Detection in Twitter
Can Liu, Wen Li, Bradford Demarest, Yue Chen, Sara Couture, Daniel Dakota, Nikita Haduong, Noah Kaufman, Andrew Lamont, Manan Pancholi, Kenneth Steimel, and Sandra Kübler
- Tohoku at SemEval-2016 Task 6: Feature-based Model versus Convolutional Neural Network for Stance Detection
Yuki Igarashi, Hiroya Komatsu, Sosuke Kobayashi, Naoaki Okazaki, and Kentaro Inui
- UWB at SemEval-2016 Task 6: Stance Detection
Peter Krejzl and Josef Steinberger
- DeepStance at SemEval-2016 Task 6: Detecting Stance in Tweets Using Character and Word-Level CNNs
Prashanth Vijayaraghavan, Ivan Sysoev, Soroush Vosoughi, and Deb Roy
- NLDS-UCSC at SemEval-2016 Task 6: A Semi-Supervised Approach to Detecting Stance in Tweets
Amita Misra, Brian Ecker, Theodore Handleman, Nicolas Hahn, and Marilyn A. Walker
- ltl.uni-due at SemEval-2016 Task 6: Stance Detection in Social Media Using Stacked Classifiers
Michael Wojatzki and Torsten Zesch
- CU-GWU Perspective at SemEval-2016 Task 6: Ideological Stance Detection in Informal Text
Heba Elfardy and Mona Diab
- JU_NLP at SemEval-2016 Task 6: Detecting Stance in Tweets using Support Vector Machines
Braja Gopal Patra, Dipankar Das, and Sivaji Bandyopadhyay
- IDINTNU at SemEval-2016 Task 6: Detecting Stance in Tweets Using Shallow Features and GloVe Vectors for Word Representation
Henrik Bøhler, Petter Asla, Erwin Marsi, and Rune Sætre

- ECNU at SemEval 2016 Task 6: Relevant or Not? Supportive or Not? A Two-step Learning System for Automatic Detecting Stance in Tweets
Zhihua Zhang and Man Lan
- MITRE at SemEval-2016 Task 6: Transfer Learning for Stance Detection
Guido Zarrella and Amy Marsh
- TakeLab at SemEval-2016 Task 6: Stance Classification in Tweets Using a Genetic Algorithm Based Ensemble
Martin Tutek, Ivan Sekulic, Paula Gombar, Ivan Paljak, Filip Culinovic, Filip Boltuzic, Mladen Karan, Domagoj Alagić, and Jan Šnajder
- LSIS at SemEval-2016 Task 7: Using Web Search Engines for English and Arabic Unsupervised Sentiment Intensity Prediction
Amal Htait, Sébastien Fournier, and Patrice Bellot
- iLab-Edinburgh at SemEval-2016 Task 7: A Hybrid Approach for Determining Sentiment Intensity of Arabic Twitter Phrases
Eshrag Refaei and Verena Rieser
- UWB at SemEval-2016 Task 7: Novel Method for Automatic Sentiment Intensity Determination
Ladislav Lenc, Pavel Král, and Václav Rajtmář
- NileTMRG at SemEval-2016 Task 7: Deriving Prior Polarities for Arabic Sentiment Terms
Samhaa R. El-Beltagy
- ECNU at SemEval-2016 Task 7: An Enhanced Supervised Learning Method for Lexicon Sentiment Intensity Ranking
Feixiang Wang, Zhihua Zhang, and Man Lan

12:30–2:00 **Lunch**

- 2:00–3:30 **Textual Similarity, Question Answering and Semantic Analysis**
- 2:15–2:30 SemEval-2016 Task 2: Interpretable Semantic Textual Similarity
Eneko Agirre, Aitor González-Agirre, Inigo Lopez-Gazpio, Montse Marítxalar, German Rigau, and Larraitz Uriar
- 2:30–2:45 SemEval-2016 Task 3: Community Question Answering
Preslav Nakov, Lluís Márquez, Alessandro Moschitti, Walid Magdy, Hamdy Mubarak, abed Alhakim Freihat, Jim Glass, and Bilal Randeree
- 2:45–3:00 SemEval-2016 Task 10: Detecting Minimal Semantic Units and their Meanings (DiMSUM)
Nathan Schneider, Dirk Hovy, Anders Johannsen, and Marine Carpuat
- 3:00–3:15 SemEval 2016 Task 11: Complex Word Identification
Gustavo Paetzold and Lucia Specia
- 3:15–3:30 **Textual Similarity and Question Answering Discussion (Task Organizers)**

3:30–4:00 **Coffee Break**

- 4:00–5:30 **Poster Session: Textual Similarity, and Question Answering**
- FBK HLT-MT at SemEval-2016 Task 1: Cross-lingual Semantic Similarity Measurement Using Quality Estimation Features and Compositional Bilingual Word Embeddings
Duygu Ataman, Jose G. C. De Souza, Marco Turchi, and Matteo Negri
 - VRep at SemEval-2016 Task 1 and Task 2: A System for Interpretable Semantic Similarity
Sam Henry and Allison Sands
 - UTA DLNLP at SemEval-2016 Task 1: Semantic Textual Similarity: A Unified Framework for Semantic Processing and Evaluation
Peng Li and Heng Huang

- UWB at SemEval-2016 Task 1: Semantic Textual Similarity using Lexical, Syntactic, and Semantic Information
Tomáš Brychcín and Lukáš Svoboda
- HHU at SemEval-2016 Task 1: Multiple Approaches to Measuring Semantic Textual Similarity
Matthias Liebeck, Philipp Pollack, Pashutan Modaresi, and Stefan Conrad
- Samsung Poland NLP Team at SemEval-2016 Task 1: Necessity for diversity; combining recursive autoencoders, WordNet and ensemble methods to measure semantic similarity.
Barbara Rychalska, Katarzyna Pakulska, Krystyna Chodorowska, Wojciech Walczak, and Piotr Andruszkiewicz
- USFD at SemEval-2016 Task 1: Putting different State-of-the-Arts into a Box
Ahmet Aker, Frederic Blain, Andres Duque, Marina Fomicheva, Jurica Seva, and Kashif Shah
- NaCTeM at SemEval-2016 Task 1: Inferring sentence-level semantic similarity from an ensemble of complementary lexical and sentence-level features
Piotr Przybyła, Nhungh T. H. Nguyen, Matthew Shardlow, Georgios Kontonasios, and Sophia Ananiadou
- ECNU at SemEval-2016 Task 1: Leveraging Word Embedding From Macro and Micro Views to Boost Performance for Semantic Textual Similarity
Junfeng Tian and Man Lan
- SAARSHEFF at SemEval-2016 Task 1: Semantic Textual Similarity with Machine Translation Evaluation Metrics and (eXtreme) Boosted Tree Ensembles
Liling Tan, Carolina Scarton, Lucia Specia, and Josef van Genabith
- WOLVESAAr at SemEval-2016 Task 1: Replicating the Success of Monolingual Word Alignment and Neural Embeddings for Semantic Textual Similarity
Hannah Bechara, Rohit Gupta, Liling Tan, Constantin Orasan, Ruslan Mitkov, and Josef van Genabith
- DTSim at SemEval-2016 Task 1: Semantic Similarity Model Including Multi-Level Alignment and Vector-Based Compositional Semantics
Rajendra Banjade, Nabin Maharjan, Dipesh Gautam, and Vasile Rus
- ISCAS_NLP at SemEval-2016 Task 1: Sentence Similarity Based on Support Vector Regression using Multiple Features
Cheng Fu, Bo An, Xianpei Han, and Le Sun
- UMD-TTIC-UW at SemEval-2016 Task 1: Attention-Based Multi-Perspective Convolutional Neural Networks for Textual Similarity Measurement
Hua He, John Wieting, Kevin Gimpel, Jinfeng Rao, and Jimmy Lin
- DLSCU at SemEval-2016 Task 1: Supervised Models of Sentence Similarity
Md Arafat Sultan, Steven Bethard, and Tamara Sumner
- DCU-SEManiacs at SemEval-2016 Task 1: Synthetic Paragraph Embeddings for Semantic Textual Similarity
Chris Hokamp and Piyush Arora
- iUBC at SemEval-2016 Task 2: RNNs and LSTMs for interpretable STS
Indigo Lopez-Gazpio, Eneko Agirre, and Montse Maritxalar
- Rev at SemEval-2016 Task 2: Aligning Chunks by Lexical, Part of Speech and Semantic Equivalence
Karin Verspoor and Timothy Miller

- Inspire at SemEval-2016 Task 2: Interpretable Semantic Textual Similarity Alignment based on Answer Set Programming
Mishal Kazmi and Peter Schüller
 - FBK-HLT-NLP at SemEval-2016 Task 2: A Multitask, Deep Learning Approach for Interpretable Semantic Textual Similarity
Simone Magnolini, Anna Feltracco, and Bernardo Magnini
 - IISCNLP at SemEval-2016 Task 2: Interpretable STS with ILP based Multiple Chunk Aligner
Lavanya Tekumalla and Sharmistha Jat
 - VENSEEVAL at Semeval-2016 Task 2 iSTS - with a full-fledged rule-based approach
Rodolfo Delmonte
 - UW-B at SemEval-2016 Task 2: Interpretable Semantic Textual Similarity with Distributional Semantics for Chunks
Miloslav Konopik, Ondrej Prazak, David Steinberger, and Tomáš Brychcín
 - DTSim at SemEval-2016 Task 2: Interpreting Similarity of Texts Based on Automated Chunking, Chunk Alignment and Semantic Relation Prediction
Rajendra Banjade, Nabin Maharjan, Nobal Bikram Niraula, and Vasile Rus
 - UH-PRHLT at SemEval-2016 Task 3: Combining Lexical and Semantic-based Features for Community Question Answering
Marc Franco-Salvador, Sudipta Kar, Thamar Solorio, and Paolo Rosso
 - RDI_Team at SemEval-2016 Task 3: RDI Unsupervised Framework for Text Ranking
Ahmed Magooda, Amr Gomaa, Ashraf Mahgoub, Hany Ahmed, Mohsen Rashwan, Hazem Raafat, Eslam Kamal, and Ahmad Al Sallab
 - KeLP at SemEval-2016 Task 3: Learning Semantic Relations between Questions and Answers
Simone Filice, Danilo Croce, Alessandro Moschitti, and Roberto Basili
 - SLS at SemEval-2016 Task 3: Neural-based Approaches for Ranking in Community Question Answering
Mitra Mohtarami, Yonatan Belinkov, Wei-Ning Hsu, Yu Zhang, Tao Lei, Kfir Bar, Scott Cyphers, and Jim Glass
 - SUper Team at SemEval-2016 Task 3: Building a Feature-Rich System for Community Question Answering
Tsvetomila Mihaylova, Pepa Gencheva, Martin Boyanov, Ivana Yovcheva, Todor Mihaylov, Momchil Hardalov, Yasen Kiprov, Daniel Balchev, Ivan Koychev, Preslav Nakov, Ivelina Nikolova, and Galia Angelova
 - PMI-cool at SemEval-2016 Task 3: Experiments with PMI and Goodness Polarity Lexicons for Community Question Answering
Daniel Balchev, Yasen Kiprov, Ivan Koychev, and Preslav Nakov
 - UniMelb at SemEval-2016 Task 3: Identifying Similar Questions by combining a CNN with String Similarity Measures
Timothy Baldwin, Huizhi Liang, Bahar Salehi, Doris Hoogeveen, Yitong Li, and Long Duong
 - ICL00 at SemEval-2016 Task 3: Translation-Based Method for CQA System
Yunfang Wu and Minghua Zhang
 - Overfitting at SemEval-2016 Task 3: Detecting Semantically Similar Questions in Community Question Answering Forums with Word Embeddings
Hujie Wang and Pascal Poupart
 - QU-IR at SemEval 2016 Task 3: Learning to Rank on Arabic Community Question Answering Forums with Word Embedding
Rana Malhas, Marwan Torki, and Tamer Elsayed
-

- ECNU at SemEval-2016 Task 3: Exploring Traditional Method and Deep Learning Method for Question Retrieval and Answer Ranking in Community Question Answering
Guoshun Wu and Man Lan
- SemanticZ at SemEval-2016 Task 3: Ranking Relevant Answers in Community Question Answering Using Semantic Similarity Based on Fine-tuned Word Embeddings
Todor Mihaylov and Preslav Nakov
- MTE-NN at SemEval-2016 Task 3: Can Machine Translation Evaluation Help Community Question Answering?
Francisco Guzmán, Preslav Nakov, and Lluís Màrquez
- ConvKN at SemEval-2016 Task 3: Answer and Question Selection for Question Answering on Arabic and English Fora
Alberto Barrón-Cedeño, Giovanni Da San Martino, Shafiq Joty, Alessandro Moschitti, Fahad Al-Obaidli, Salvatore Romeo, Kateryna Tymoshenko, and Antonio Uva
- ITNLP-AiKF at SemEval-2016 Task 3 a qesiton answering system using community QA repository
Chang e Jia
- UFRGS&LIF at SemEval-2016 Task 10: Rule-Based MWE Identification and Predominant-Supersense Tagging
Silvio Cordeiro, Carlos Ramisch, and Aline Villavicencio
- WHUNlp at SemEval-2016 Task DiMSUM: A Pilot Study in Detecting Minimal Semantic Units and their Meanings using Supervised Models
Xin Tang, Fei Li, and Donghong Ji
- UTU at SemEval-2016 Task 10: Binary Classification for Expression Detection (BCED)
Jari Björne and Tapio Salakoski
- UW-CSE at SemEval-2016 Task 10: Detecting Multiword Expressions and Supersenses using Double-Chained Conditional Random Fields
Mohammad Javad Hosseini, Noah A. Smith, and Su-In Lee
- ICL-HD at SemEval-2016 Task 10: Improving the Detection of Minimal Semantic Units and their Meanings with an Ontology and Word Embeddings
Angelika Kirilin, Felix Krauss, and Yannick Versley
- VectorWeavers at SemEval-2016 Task 10: From Incremental Meaning to Semantic Unit (phrase by phrase)
Andreas Scherbakov, Ekaterina Vylomova, Fei Liu, and Timothy Baldwin
- PLUJAGH at SemEval-2016 Task 11: Simple System for Complex Word Identification
Krzysztof Wróbel
- USAAR at SemEval-2016 Task 11: Complex Word Identification with Sense Entropy and Sentence Perplexity
José Manuel Martínez and Liling Tan
- Sensible at SemEval-2016 Task 11: Neural Nonsense Mangled in Ensemble Mess
Gillin Nat
- SV000gg at SemEval-2016 Task 11: Heavy Gauge Complex Word Identification with System Voting
Gustavo Paetzold and Lucia Specia
- Melbourne at SemEval 2016 Task 11: Classifying Type-level Word Complexity using Random Forests with Corpus and Word List Features
Julian Brooke, Alexandra Uitdenbogerd, and Timothy Baldwin

- CLaC at SemEval-2016 Task 11: Exploring linguistic and psycho-linguistic Features for Complex Word Identification
Elnaz Davoodi and Leila Kosseim
- JU_NLP at SemEval-2016 Task 11: Identifying Complex Words in a Sentence
Niloy Mukherjee, Braja Gopal Patra, Dipankar Das, and Sivaji Bandyopadhyay
- MAZA at SemEval-2016 Task 11: Detecting Lexical Complexity Using a Decision Stump Meta-Classifier
Shervin Malmasi and Marcos Zampieri
- LTG at SemEval-2016 Task 11: Complex Word Identification with Classifier Ensembles
Shervin Malmasi, Mark Dras, and Marcos Zampieri
- MacSaar at SemEval-2016 Task 11: Zipfian and Character Features for ComplexWord Identification
Marcos Zampieri, Liling Tan, and Josef van Genabith
- Garuda & Bhasha at SemEval-2016 Task 11: Complex Word Identification Using Aggregated Learning Models
Prafulla Choube and Shubham Pateria
- TALN at SemEval-2016 Task 11: Modelling Complex Words by Contextual, Lexical and Semantic Features
Francesco Ronzano, Ahmed Abura'ed, Luis Espinosa Anke, and Horacio Saggion
- IIIT at SemEval-2016 Task 11: Complex Word Identification using Nearest Centroid Classification
Ashish Palakurthi and Radhika Mamidi
- AmritaCEN at SemEval-2016 Task 11: Complex Word Identification using Word Embedding
Anand Kumar and Soman K P
- CoastalCPH at SemEval-2016 Task 11: The importance of designing your Neural Networks right
Joachim Bingel, Natalie Schluter, and Héctor Martínez Alonso
- HMC at SemEval-2016 Task 11: Identifying Complex Words Using Depth-limited Decision Trees
Maury Quijada and Julie Medero
- UWB at SemEval-2016 Task 11: Exploring Features for Complex Word Identification
Michał Konkol
- AI-KU at SemEval-2016 Task 11: Word Embeddings and Substring Features for Complex Word Identification
Onur Kuru
- Pomona at SemEval-2016 Task 11: Predicting Word Complexity Based on Corpus Frequency
David Kauchak

Friday, June 17, 2016

9:00–10:30 Perspectives

9:00–9:30 **SemEval-2017 Preview (SemEval organizers)**

9:30–10:30 **Invited Talk**

10:30–11:00 **Coffee Break**

11:00–12:30 **Semantic Analysis, Semantic Parsing and Semantic Taxonomy**

11:00–11:15 SemEval-2016 Task 12: Clinical TempEval

Steven Bethard, Guergana Savova, Wei-Te Chen, Leon Derczynski, James Pustejovsky, and Marc Verhagen

11:15–11:30 SemEval-2016 Task 8: Meaning Representation Parsing
Jonathan May

11:30–11:45 SemEval-2016 Task 9: Chinese Semantic Dependency Parsing
Wanxiang Che, Yanqiu Shao, Ting Liu, and Yu Ding

11:45–12:00 SemEval-2016 Task 13: Taxonomy Extraction Evaluation (TExEval-2)
Georgeta Borda, Els Lefever, and Paul Buitelaar

12:00–12:15 SemEval-2016 Task 14: Semantic Taxonomy Enrichment
David Jurgens and Mohammad Taher Pilehvar

12:30–2:00 **Lunch**

2:00–3:30 **Best Of SemEval**

3:15–3:30 LIMSI-COT at SemEval-2016 Task 12: Temporal relation identification using a pipeline of classifiers
Julien Tourille, Olivier Ferret, Aurélie Névéol, and Xavier Tannier

3:30–4:00 **Coffee Break**

4:00–5:30 **Poster Session: Semantic Analysis, Parsing, and Taxonomy**

- RIGA at SemEval-2016 Task 8: Impact of Smatch Extensions and Character-Level Neural Translation on AMR Parsing Accuracy
Guntis Barzdins and Didzis Gosko
- DynamicPower at SemEval-2016 Task 8: Processing syntactic parse trees with a Dynamic Semantics core
Alastair Butler
- M2L at SemEval-2016 Task 8: AMR Parsing with Neural Networks
Yevgeniy Puzikov, Daisuke Kawahara, and Sadao Kurohashi
- ICL-HD at SemEval-2016 Task 8: Meaning Representation Parsing - Augmenting AMR Parsing with a Preposition Semantic Role Labeling Neural Network
Lauritz Brandt, David Grimm, Mengfei Zhou, and Yannick Versley
- UCL+Sheffield at SemEval-2016 Task 8: Imitation learning for AMR parsing with an alpha-bound
James Goodman, Andreas Vlachos, and Jason Naradowsky
- CAMR at SemEval-2016 Task 8: An Extended Transition-based AMR Parser
Chuan Wang, Sameer Pradhan, Xiaoman Pan, Heng Ji, and Nianwen Xue
- The Meaning Factory at SemEval-2016 Task 8: Producing AMRs with Boxer
Johannes Bjerva, Johan Bos, and Hessel Haagsma
- UoF at SemEval-2016 Task 8: Learning Synchronous Hyperedge Replacement Grammar for AMR Parsing
Xiaochang Peng and Daniel Gildea
- CLIPUMD at SemEval-2016 Task 8: Parser for Abstract Meaning Representation using Learning to Search
Sudha Rao, Yogarshi Vyas, Hal Daumé III, and Philip Resnik

- CU-NLP at SemEval-2016 Task 8: AMR Parsing using LSTM-based Recurrent Neural Networks
William Foland and James H. Martin
- CMU at SemEval-2016 Task 8: Graph-based AMR Parsing with Infinite Ramp Loss
Jeffrey Flanigan, Chris Dyer, Noah A. Smith, and Jaime Carbonell
- IHS-RD-Belarus at SemEval-2016 Task 9: Transition-based Chinese Semantic Dependency Parsing with Online Reordering and Bootstrapping.
Artsiom Artsymenia, Palina Doumar, and Maria Yermakovich
- OCLSP at SemEval-2016 Task 9: Multilayered LSTM as a Neural Semantic Dependency Parser
Lifeng Jin, Manjuan Duan, and William Schuler
- OSU_CHGCG at SemEval-2016 Task 9 : Chinese Semantic Dependency Parsing with Generalized Categorical Grammar
Manjuan Duan, Lifeng Jin, and William Schuler
- LIMSI at SemEval-2016 Task 12: machine-learning and temporal information to identify clinical events and time expressions
Cyril Grouin and Véronique MORICEAU
- Hitachi at SemEval-2016 Task 12: A Hybrid Approach for Temporal Information Extraction from Clinical Notes
Sarath P R, Manikandan R, and Yoshiki Niwa
- CDE-IIITH at SemEval-2016 Task 12: Extraction of Temporal Information from Clinical documents using Machine Learning techniques
Veera Raghavendra Chikka
- VUACLT at SemEval 2016 Task 12: A CRF Pipeline to Clinical TempEval
Tomaso Caselli and Roser Morante
- GUIR at SemEval-2016 task 12: Temporal Information Processing for Clinical Narratives
Arman Cohan, Kevin Meurer, and Nazli Goharian
- UtahBMI at SemEval-2016 Task 12: Extracting Temporal Information from Clinical Text
Abdulrahman Khalifa, Sumithra Velupillai, and Stephane Meystre
- ULISBOA at SemEval-2016 Task 12: Extraction of temporal expressions, clinical events and relations using IBEnt
Marcia Barros, André Lamúrias, Gonçalo Figueiró, Marta Antunes, Joana Teixeira, Alexandre Pinheiro, and Francisco M. Couto
- UTA DLNLP at SemEval-2016 Task 12: Deep Learning Based Natural Language Processing System for Clinical Information Identification from Clinical Notes and Pathology Reports
Peng Li and Heng Huang
- Brundlefly at SemEval-2016 Task 12: Recurrent Neural Networks vs. Joint Inference for Clinical Temporal Information Extraction
Jason Fries
- KULEuven-LIIR at SemEval 2016 Task 12: Detecting Narrative Containment in Clinical Records
Artuur Leeuwenberg and Marie-Francine Moens
- CENTAL at SemEval-2016 Task 12: a linguistically fed CRF model for medical and temporal information extraction
Charlotte Hansart, Damien De Meyere, Patrick Watrin, André Bittar, and Cédrik Fairon
- UTHealth at SemEval-2016 Task 12: an End-to-End System for Temporal Information Extraction from Clinical Notes
Hee-Jin Lee, Hua Xu, Jingqi Wang, Yaoyun Zhang, Sungrim Moon, Jun Xu, and Yonghui Wu

- NUIG-UNLP at SemEval-2016 Task 13: A Simple Word Embedding-based Approach for Taxonomy Extraction
Joel Pocostales
- USAAR at SemEval-2016 Task 13: Hyponym Endocentricity
Liling Tan, Francis Bond, and Josef van Genabith
- JUNLP at SemEval-2016 Task 13: A Language Independent Approach for Hypernym Identification
Promita Maitra and Dipankar Das
- QASSIT at SemEval-2016 Task 13: On the integration of Semantic Vectors in Pretopological Spaces for Lexical Taxonomy Acquisition
Guillaume Cleuziou and Jose G. Moreno
- TAXI at SemEval-2016 Task 13: a Taxonomy Induction Method based on Lexico-Syntactic Patterns, Substrings and Focused Crawling
Alexander Panchenko, Stefano Faralli, Eugen Ruppert, Steffen Remus, Hubert Naets, Cédrick Fairon, Simone Paolo Ponchetto, and Chris Biemann
- Duluth at SemEval 2016 Task 14: Extending Gloss Overlaps to Enrich Semantic Taxonomies
Ted Pedersen
- TALN at SemEval-2016 Task 14: Semantic Taxonomy Enrichment Via Sense-Based Embeddings
Luis Espinosa Anke, Francesco Ronzano, and Horacio Saggion
- MSEjrKU at SemEval-2016 Task 14: Taxonomy Enrichment by Evidence Ranking
Michael Schlichtkrull and Héctor Martínez Alonso
- Deftor at SemEval-2016 Task 14: Taxonomy enrichment using definition vectors
Hristo Tanev and Agata Rotondi
- UMNDuluth at SemEval-2016 Task 14: WordNet’s Missing Lemmas
Jon Rusert and Ted Pedersen
- VCU at Semeval-2016 Task 14: Evaluating definitional-based similarity measure for semantic taxonomy enrichment
Bridget McInnes
- GWU NLP at SemEval-2016 Shared Task 1: Matrix Factorization for Crosslingual STS
Hanan Aldarmaki and Mona Diab
- CNRC: Experiments in Crosslingual Semantic Textual Similarity
Chi-ku Lo, Cyril Goutte, and Michel Simard
- MayoNLP at SemEval-2016 Task 1: Semantic Textual Similarity based on Lexical Semantic Net and Deep Learning Semantic Model
Naveed Afzal, Yanshan Wang, and Hongfang Liu
- UoB-UK: A Flexible and Extendable System for Semantic Text Similarity using Types, Surprise and Phrase Linking
Harish Tayyar Madabushi, Mark Buhagiar, and Mark Lee
- BIT at SemEval-2016 Task 1: Sentence Similarity Based on Alignments and Vector with the Weight of Information Content
Hao Wu, Heyan Huang, and Wenpeng Lu
- RICOH at SemEval-2016 Task 1: IR-based Semantic Textual Similarity Estimation
Hideo Itoh
- IHS-RD-Belarus at SemEval-2016 Task 1: Multistage Approach for Measuring Semantic Similarity
Maryna Beliuha and Maryna Chernyshevich
- JUNITMZ: Identifying Semantic Similarity Using Levenshtein Ratio
Sandip Sarkar, Dipankar Das, Partha Pakray, and Alexander Gelbukh

- Amrita_CEN at SemEval-2016 Task Semantic Textual Similarity : Semantic Relation from Word Embeddings in Higher Dimension
Barathi Ganesh HB, Anand Kumar M, and Soman KP
- NUIG-UNLP at SemEval-2016 Task 1: Soft Alignment and Deep Learning for Semantic Textual Similarity
John Philip McCrae, Kartik Asooja, Nitish Aggarwal, and Paul Buitelaar
- SEMSIM: A Multi-Feature Approach to Semantic Text Similarity
Luigi Di Caro and Guido Boella
- LIPN-IIMAS at SemEval-2016 Task STS: Random Forest Regression Experiments on Align-and-Differentiate and Word Embeddings penalizing strategies
Oscar William Lightgow Serrano, Ivan Vladimir Meza Ruiz, Albert Manuel Orozco Camacho, Jorge Garcia Flores, and Davide Buscaldi
- UNBNLP at SemEval-2016 Task 1: Semantic Textual Similarity: A Unified Framework for Semantic Processing and Evaluation
Milton King, Waseem Gharbieh, SoHyun Park, and Paul Cook
- ASOBEK at SemEval-2016 Task 1: Sentence Representation with Character N-gram Embeddings for Semantic Textual Similarity
Asli Eyecioğlu and Bill Keller
- SimiHawk at SemEval-2016 Task 1: A Deep Ensemble System for Semantic Textual Similarity
Peter Potash, William Boag, Alexey Romanov, Vasili Ramanishka, and Anna Rumshisky
- SERGIOJIMENEZ at SemEval-2016 Task-1: Effectively Combining Paraphrase Database, String Matching, WordNet, and Word Embedding for Semantic Textual Similarity
Sergio Jimenez
- RTM at SemEval-2016 Task 1: Predicting Semantic Similarity with Referential Translation Machines and Related Statistics
Ergun Bicici
- DalGTM at SemEval-2016 Task 1: Importance-Aware Compositional Approach to Short Text Similarity
Jie Mei, Aminul Islam, and Evangelos Milios

Workshop 2: Human-Computer Question Answering

Organizers: *Mohit Iyyer, He He, Jordan Boyd-Graber, and Hal Daumé III*

Venue: Marina 2

Thursday, June 16, 2016

9:00–9:15 Welcome

Invited Talks I

9:15–9:50 Invited Talk by Ray Mooney

9:50–10:25 Invited Talk by Jason Weston

10:25–10:40 Coffee Break

Talks by Authors of Accepted Papers

- 10:40–11:00 Watson Discovery Advisor: Question-answering in an industrial setting
Charley Beller, Graham Katz, Allen Ginsberg, Chris Phipps, Sean Bethard, Paul Chase, Elinna Shek, and Kristen Summers
- 11:00–11:20 Crowdsourcing for (almost) Real-time Question Answering
Denis Savenkov, Scott Weitzner, and Eugene Agichtein
- 11:20–11:40 Attention-Based Convolutional Neural Network for Machine Comprehension
Wenpeng Yin, Sebastian Ebert, and Hinrich Schütze

11:40–12:00 Invited Talk by Eunsol Choi

12:00–1:15 Lunch

Invited Talks II

1:15–1:50 Invited Talk by Peter Clark

1:50–2:25 Invited Talk by Zhengdong Lu

2:25–3:00 Invited Talk by Richard Socher

Poster Session

3:00–3:15 Coffee Break

3:15–4:00 Open-domain Factoid Question Answering via Knowledge Graph Search
Ahmad Aghaeabrahimian

3:15–4:00 Neural Enquirer: Learning to Query Tables in Natural Language
Pengcheng Yin, Zhengdong Lu, Hang Li, and Kao Ben

3:15–4:00 Neural Generative Question Answering
Jun Yin, Xin Jiang, Zhengdong Lu, Lifeng Shang, Hang Li, and Xiaoming Li

3:15–4:00 “A Distorted Skull Lies in the Bottom Center...” Identifying Paintings from Text Descriptions

Anupam Guha, Mohit Iyyer, and Jordan Boyd-Graber

3:15–4:00 Using Confusion Graphs to Understand Classifier Error
Davis Yoshida and Jordan Boyd-Graber

3:15–4:00 Paraphrase for Open Question Answering: New Dataset and Methods

Ying Xu, Pascual Martínez-Gómez, Yusuke Miyao, and Randy Goebel

Exhibition Match

4:00–5:15 Best Shared Task System vs. Human Quiz Bowl Team

Workshop 3: Fifth Workshop on Computational Linguistics for Literature (clfl2016)

Organizers: *Anna Feldman, Anna Kazantseva, and Stan Szpakowicz*

Venue: Spinnaker 2

Thursday, June 16, 2016

9:00–10:00 **Session 1**

9:00–9:05 **Welcome**

9:05–10:00 **Genesis reads Macbeth: The role of stories in human intelligence: invited talk, Patrick Henry Winston, MIT**

10:00–10:30 Supervised Machine Learning for Hybrid Meter
Alex Estes and Christopher Hench

10:30–11:00 **Coffee break**

11:00–12:30 **Session 2**

11:00–11:30 Automatic Text Generation by Learning from Literary Structures
Angel Daza, Hiram Calvo, and Jesús Figueroa-Nazuno

11:30–12:00 Intersecting Word Vectors to Take Figurative Language to New Heights
Andrea Gagliano, Emily Paul, Kyle Booten, and Marti A. Hearst

12:00–12:30 Gender-Distinguishing Features in Film Dialogue
Alexandra Schofield and Leo Mehr

12:30–2:00 **Lunch break**

2:00–3:30 **Session 3**

2:00–2:05 **Interactive Text Mining Suite, a teaser to accompany a demo**

2:05–3:00 **“The Not-Moth”: Poetic Expression in Array Spaces of Computational Constellation. Invited talk, Loss Pequeño Glazier, SUNY Buffalo**

3:00–3:30 Reconstructing Ancient Literary Texts from Noisy Manuscripts
Moshe Koppel, Moty Michaely, and Alex Tal

3:30–4:00 **Coffee break**

4:00–5:15 **Session 4**

4:00–4:30 Syntax Matters for Rhetorical Structure: The Case of Chiasmus
Marie Dubremetz and Joakim Nivre

4:30–5:00 Bilingual Chronological Classification of Hafez’s Poems
Arya Rahgozar and Diana Inkpen

5:00–5:15 **Wrap-up**

Workshop 4: Computational Linguistics and Clinical Psychology — From Linguistic Signal to Clinical Reality (CLPsych)

Organizers: *Kristy Hollingshead and Lyle Ungar*

Venue: Executive 4

Thursday, June 16, 2016

9:00–9:20 **Opening Remarks (Kristy Hollingshead and Lyle Ungar)**

9:20–10:30 **Oral Presentations, Session 1**

- Detecting late-life depression in Alzheimer's disease through analysis of speech and language
Kathleen C. Fraser, Frank Rudzicz, and Graeme Hirst
- Towards Early Dementia Detection: Fusing Linguistic and Non-Linguistic Clinical Data
Joseph Bullard, Cecilia Ovesdotter Alm, Xumin Liu, Qi Yu, and Rubén Proaño

10:30–11:00 **Break**

11:00–11:45 **Poster Presentations**

- Self-Reflective Sentiment Analysis
Benjamin Shickel, Martin Heesacker, Sherry Benton, Ashkan Ebadi, Paul Nickerson, and Parisa Rashidi
- Is Sentiment in Movies the Same as Sentiment in Psychotherapy? Comparisons Using a New Psychotherapy Sentiment Database
Michael Tanana, Aaron Dembe, Christina S. Soma, Zac Imel, David Atkins, and Vivek Sri Kumar
- Building a Motivational Interviewing Dataset
Verónica Pérez-Rosas, Rada Mihalcea, Kenneth Resnicow, Satinder Singh, and Lawrence An
- Crazy Mad Nutters: The Language of Mental Health
Jena D. Hwang and Kristy Hollingshead
- The language of mental health problems in social media
George Gkotsis, Anika Oellrich, Tim Hubbard, Richard Dobson, Maria Liakata, Sumithra Velupillai, and Rina Dutta
- Exploring Autism Spectrum Disorders Using HLT
Julia Parish-Morris, Mark Liberman, Neville Ryant, Christopher Cieri, Leila Bateman, Emily Ferguson, and Robert Schultz

11:45–1:00 **Lunch**

1:00–2:45 **Oral Presentations, Session 2**

- Generating Clinically Relevant Texts: A Case Study on Life-Changing Events
Mayuresh Oak, Anil Behera, Titus Thomas, Cecilia Ovesdotter Alm, Emily Prud'hommeaux, Christopher Homan, and Raymond Ptucha

- Don't Let Notes Be Misunderstood: A Negation Detection Method for Assessing Risk of Suicide in Mental Health Records
George Gkotsis, Sumithra Velupillai, Anika Oellrich, Harry Dean, Maria Liakata, and Rina Dutta
- Exploratory Analysis of Social Media Prior to a Suicide Attempt
Glen Coppersmith, Kim Ngo, Ryan Leary, and Anthony Wood

2:45–3:00 **Break**

3:00–3:15 **Shared Task Introduction**

- CLPsych 2016 Shared Task: Triaging content in online peer-support forums
David N. Milne, Glen Pink, Ben Hachey, and Rafael A. Calvo

3:15–3:40 **Shared Task Poster Presentations, Session 1**

- Data61-CSIRO systems at the CLPsych 2016 Shared Task
Sunghwan Mac Kim, Yufei Wang, Stephen Wan, and Cecile Paris
- Predicting Post Severity in Mental Health Forums
Shervin Malmasi, Marcos Zampieri, and Mark Dras
- Classifying ReachOut posts with a radial basis function SVM
Chris Brew
- Triaging Mental Health Forum Posts
Arman Cohan, Sydney Young, and Nazli Goharian
- Mental Distress Detection and Triage in Forum Posts: The LT3 CLPsych 2016 Shared Task System
Bart Desmet, Gilles Jacobs, and Véronique Hoste
- Text Analysis and Automatic Triage of Posts in a Mental Health Forum
Ehsaneddin Asgari, Soroush Nasiriany, and Mohammad R.K. Mofrad
- The UMD CLPsych 2016 Shared Task System: Text Representation for Predicting Triage of Forum Posts about Mental Health
Meir Friedenberg, Hadi Amiri, Hal Daumé III, and Philip Resnik

3:40–4:00 **Break**

4:00–4:25 **Shared Task Poster Presentations, Session 2**

- Using Linear Classifiers for the Automatic Triage of Posts in the 2016 CLPsych Shared Task
Juri Opitz
- The GW/UMD CLPsych 2016 Shared Task System
Ayah Zirikly, Varun Kumar, and Philip Resnik
- Semi-supervised CLPsych 2016 Shared Task System Submission
Nicolas Rey-Villamizar, Prasha Shrestha, Thamar Solorio, Farig Sadegue, Steven Bethard, and Ted Pedersen
- Combining Multiple Classifiers Using Global Ranking for ReachOut.com Post Triage
Chen-Kai Wang, Hong-Jie Dai, Chih-Wei Chen, Jitendra Jonnagaddala, and Nai-Wen Chang
- Classification of mental health forum posts
Glen Pink, Will Radford, and Ben Hachey
- Automatic Triage of Mental Health Online Forum Posts: CLPsych 2016 System Description
Hayda Almeida, Marc Queudot, and Marie-Jean Meurs
- Automatic Triage of Mental Health Forum Posts
Benjamin Shickel and Parisa Rashidi
- Text-based experiments for Predicting mental health emergencies in online web forum posts
Hector-Hugo Franco-Peña and Liliana Mamani Sanchez

4:25–4:45 **Closing Remarks**

Workshop 5: 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA 2016)

Organizers: *Alexandra Balahur, Erik van der Goot, Piek Vossen, and Andrés Montoyo*

Venue: Marina 3

Thursday, June 16, 2016

8:40–9:00 Opening Remarks and Intro Talk

8:45–9:00 Sentiment Analysis - What are we talking about?
Alexandra Balahur

9:00–10:30 Session 1: Sentiment Analysis Perspectives in Social Media

9:00–9:40 Sentiment, Subjectivity, and Social Analysis Go ToWork: An Industry View - Invited Talk
Seth Grimes

9:40–10:05 Rumor Identification and Belief Investigation on Twitter
Sardar Hamidian and Mona Diab

10:05–10:30 Modelling Valence and Arousal in Facebook posts
Daniel Preojuic-Pietro, H. Andrew Schwartz, Gregory Park, Johannes Eichstaedt, Margaret Kern, Lyle Ungar, and Elisabeth Shulman

10:30–11:00 Coffee Break

11:00–12:30 Session 2: Sentiment Detection for Social Media Applications

11:00–11:40 Purity Homophily in Social Networks - Invited Talk
Morteza Dehghani

11:40–12:05 Hit Songs' Sentiments Harness Public Mood & Predict Stock Market
Rachel Harsley, Bhavesh Gupta, Barbara Di Eugenio, and Huayi Li

12:05–12:30 Fashioning Data - A Social Media Perspective on Fast Fashion Brands
Rupak Chakraborty, Senjuti Kundu, and Prakul Agarwal

12:30–2:00 Lunch Break

2:00–3:30 Session 3: New Approaches in Sentiment Analysis

2:00–2:40 Deep Learning for Sentiment Analysis - Invited Talk
Richard Socher

2:40–3:05 Sentiment Lexicon Creation using Continuous Latent Space and Neural Networks
Pedro Dias Cardoso and Anindya Roy

3:05–3:30 The Effect of Negators, Modals, and Degree Adverbs on Sentiment Composition
Svetlana Kiritchenko and Saif Mohammad

3:30–4:00 Coffee Break

4:00–5:40 Session 4: Overview and Applications of Sentiment Analysis

4:00–4:25 How can NLP Tasks Mutually Benefit Sentiment Analysis? A Holistic Approach to Sentiment Analysis
Lingjia Deng and Janyce Wiebe

4:25–4:50 An Unsupervised System for Visual Exploration of Twitter Conversations
Derrick Higgins, Michael Heilman, Adrianna Jelesnianska, and Keith Ingersoll

4:50–5:10 Threat detection in online discussions
Aksel Wester, Lilja Øvrelid, Erik Veldal, and Hugo Lewi Hammer

5:10–5:30 Classification of comment helpfulness to improve knowledge sharing among medical practitioners.
Pierre André Ménard and Caroline Barrière

5:20–5:30 **Break**

6:00–7:30 **Session 5: Posters**

- Political Issue Extraction Model: A Novel Hierarchical Topic Model That Uses Tweets By Political And Non-Political Authors
Aditya Joshi, Pushpak Bhattacharyya, and Mark Carman
- Early text classification: a Naïve solution
Hugo Jair Escalante, Manuel Montes y Gomez, Luis Villasenor, and Marcelo Luis Errecalde
- Semi-supervised and unsupervised categorization of posts in Web discussion forums using part-of-speech information and minimal features
Krish Perumal and Graeme Hirst
- Linguistic Understanding of Complaints and Praises in User Reviews
Guangyu Zhou and Kavita Ganesh
- Reputation System: Evaluating Reputation among All Good Sellers
Vandana Jha, Savitha R, P Deepa Shenoy, and Venugopal K R
- Improve Sentiment Analysis of Citations with Author Modelling
Zheng Ma, Jinseok Nam, and Karsten Weihe
- Implicit Aspect Detection in Restaurant Reviews using Cooccurrence of Words
Rrubaa Panchendarajan, Nazick Ahamed, Brunthavan Murugaiah, Prakash Sivakumar, Surangika Ranathunga, and Akila Pemasiri
- Domain Adaptation of Polarity Lexicon combining Term Frequency and Bootstrapping
Salud María Jiménez-Zafra, Maite Martin, M. Dolores Molina González, and L. Alfonso Urena Lopez
- Do Enterprises Have Emotions?
Sven Buechel, Udo Hahn, Jan Goldenstein, Sebastian G. M. Händschke, and Peter Walgenbach
- A semantic-affective compositional approach for the affective labelling of adjective-noun and noun-noun pairs
Elisavet Palogiannidi, Elias Iosif, Polychronis Koutsakis, and Alexandros Potamianos
- Fracking Sarcasm using Neural Network
Aniruddha Ghosh and Tony Veale

7:30–8:00 **Closing discussion - “Where do we go from here?”**

- An Hymn of an even Deeper Sentiment Analysis
Manfred Klenner
- Sentiment Analysis in Twitter: A SemEval Perspective
Preslav Nakov
- The Challenge of Sentiment Quantification
Fabrizio Sebastiani
- A Practical Guide to Sentiment Annotation: Challenges and Solutions
Saif Mohammad
- Emotions and NLP: Future Directions
Carlo Strapparava

Workshop 6: 11th Workshop on Innovative Use of NLP for Building Educational Applications (BEA11)

Organizers: *Joel Tetreault, Jill Burstein, Claudia Leacock, and Helen Yannakoudakis*

Venue: Executive 3AB

Thursday June 16, 2016

8:45–9:00 Load Oral Presentations

9:00–9:15 Opening Remarks

9:15–9:40 The Effect of Multiple Grammatical Errors on Processing Non-Native Writing
Courtney Napoles, Aoife Cahill, and Nitin Madnani

9:40–10:05 Text Readability Assessment for Second Language Learners
Menglin Xia, Ekaterina Kochmar, and Ted Briscoe

10:05–10:30 Automatic Generation of Context-Based Fill-in-the-Blank Exercises Using Co-occurrence Likelihoods and Google n-grams
Jennifer Hill and Rahul Simha

10:30–11:00 Break

11:00–11:25 Automated classification of collaborative problem solving interactions in simulated science tasks
Michael Flor, Su-Youn Yoon, Jiangang Hao, Lei Liu, and Alina von Davier

11:25–11:50 Computer-assisted stylistic revision with incomplete and noisy feedback.
A pilot study
Christian M. Meyer and Johann Frerik Koch

11:50–12:15 A Report on the Automatic Evaluation of Scientific Writing Shared Task
Vidas Daudaravicius, Rafael E. Banchs, Elena Volodina, and Courtney Napoles

12:25–2:00 Lunch

2:00–2:45 Poster and Demo Session A

- Topicality-Based Indices for Essay Scoring
Beata Beigman Klebanov, Michael Flor, and Binod Gyawali
- Predicting the Spelling Difficulty of Words for Language Learners
Lisa Beinborn, Torsten Zesch, and Iryna Gurevych
- Characterizing Text Difficulty with Word Frequencies
Xiaobin Chen and Detmar Meurers
- Unsupervised Modeling of Topical Relevance in L2 Learner Text
Ronan Cummins, Helen Yannakoudakis, and Ted Briscoe
- UW-Stanford System Description for AESW 2016 Shared Task on Grammatical Error Detection
Dan Flickinger, Michael Goodman, and Woodley Packard
- Shallow Semantic Reasoning from an Incomplete Gold Standard for Learner Language
Levi King and Markus Dickinson

- The NTNU-YZU System in the AESW Shared Task: Automated Evaluation of Scientific Writing Using a Convolutional Neural Network
Lung-Hao Lee, Bo-Lin Lin, Liang-Chih Yu, and Yuen-Hsien Tseng
- Automated scoring across different modalities
Anastassia Loukina and Aoife Cahill
- Model Combination for Correcting Preposition Selection Errors
Nitin Madnani, Michael Heilman, and Aoife Cahill
- Pictogrammar: an AAC device based on a semantic grammar
Fernando Martínez-Santiago, Miguel Ángel García Cumberras, Arturo Montejo Ráez, and Manuel Carlos Díaz Galiano
- Detecting Context Dependence in Exercise Item Candidates Selected from Corpora
Ildikó Pilán
- Feature-Rich Error Detection in Scientific Writing Using Logistic Regression
Madeline Remse, Mohsen Mesgar, and Michael Strube
- Bundled Gap Filling: A New Paradigm for Unambiguous Cloze Exercises
Michael Wojatzki, Oren Melamud, and Torsten Zesch

2:45–3:30 Poster and Demo Session B

- Evaluation Dataset (DT-Grade) and Word Weighting Approach towards Constructed Short Answers Assessment in Tutorial Dialogue Context
Rajendra Banjade, Nabin Maharjan, Nobal Bikram Niraula, Dipesh Gautam, Borhan Samei, and Vasile Rus
- Linguistically Aware Information Retrieval: Providing Input Enrichment for Second Language Learners
Maria Chinkina and Detmar Meurers
- Enhancing STEM Motivation through Personal and Communal Values: NLP for Assessment of Utility Value in Student Writing
Beata Beigman Klebanov, Jill Burstein, Judith Harackiewicz, Stacy Priniski, and Matthew Mulholland
- Cost-Effectiveness in Building a Low-Resource Morphological Analyzer for Learner Language
Scott Ledbetter and Markus Dickinson
- Automatically Scoring Tests of Proficiency in Music Instruction
Nitin Madnani, Aoife Cahill, and Brian Riordan
- Combined Tree Kernel-based classifiers for Assessing Quality of Scientific Text
Liliana Mamani Sanchez and Hector-Hugo Franco-Peña
- Augmenting Course Material with Open Access Textbooks
Smitha Milli and Marti A. Hearst
- Exploring the Intersection of Short Answer Assessment, Authorship Attribution, and Plagiarism Detection
Björn Rudzewitz
- Sentence-Level Grammatical Error Identification as Sequence-to-Sequence Correction
Allen Schmaltz, Yoon Kim, Alexander M. Rush, and Stuart M. Shieber
- Combining Off-the-shelf Grammar and Spelling Tools for the Automatic Evaluation of Scientific Writing (AESW) Shared Task 2016
René Witte and Bahar Sateli
- Candidate re-ranking for SMT-based grammatical error correction
Zheng Yuan, Ted Briscoe, and Mariano Felice
- Spoken Text Difficulty Estimation Using Linguistic Features
Su-Youn Yoon, Yeonsuk Cho, and Diane Napolitano

- Automatically Extracting Topical Components for a Response-to-Text Writing Assessment
Zahra Rahimi and Diane Litman

3:30–4:00 **Break**

4:00–4:20 Sentence Similarity Measures for Fine-Grained Estimation of Topical Relevance in Learner Essays

Marek Rei and Ronan Cummins

4:20–4:45 Insights from Russian second language readability classification: complexity-dependent training requirements, and feature evaluation of multiple categories

Robert Reynolds

4:45–5:10 Investigating Active Learning for Short-Answer Scoring

Andrea Horbach and Alexis Palmer

5:10–5:25 **Closing Remarks**

Workshop 7: 2nd Workshop on Semantics-Driven Machine Translation (SedMT2016)

Organizers: *Deyi Xiong, Kevin Duh, Eneko Agirre, Nora Aranberri, and Houfeng Wang*

Venue: Spinnaker 1

Thursday, June 16, 2016

8:45–9:00 **Opening Remarks**

9:00–10:30 **Session 1 (Chair: Yvette Graham)**

9:00–10:00 **Invited Talk by Johan Bos**

10:00–10:15 Deterministic natural language generation from meaning representations
for machine translation

Alastair Butler

10:15–10:30 Extending Phrase-Based Translation with Dependencies by Using Graphs
Liangyou Li, Andy Way, and Qun Liu

10:30–11:00 **Coffee Break**

11:00–12:00 **Session 2 (Chair: Nora Aranberri)**

11:00–12:00 **Invited Talk by Kyunghyun Cho**

12:00–2:00 **Lunch**

2:00–3:30 **Session 3 (Chair: Andrew Finch)**

2:00–3:00 **Invited Talk by Jan Hajic**

3:00–3:15 The Naming Sharing Structure and its Cognitive Meaning in Chinese and
English

Shili Ge and Rou Song

3:15–3:30 Towards Semantic-based Hybrid Machine Translation between Bulgarian
and English

Kiril Simov, Petya Osenova, and Alexander Popov

3:30–4:00 **Coffee Break**

4:00–5:45 **Session 4 (Chair: Nora Aranberri)**

4:00–5:00 **Invited Talk by Martha Palmer**

5:00–5:45 **Panel**

5:45–6:00 **Closing**

Workshop 8: Coreference Resolution Beyond OntoNotes

Organizers: *Maciej Ograniczuk and Vincent Ng*

Venue: Marina 1

Thursday, June 16, 2016

9:00–10:30 **Session 1**

9:00–9:10 **Introduction**

9:10–10:10 **Invited talk: The (Non)Utility of Semantics for Coreference Resolution (CORBON Remix) (Michael Strube)**

10:10–10:30 Sense Anaphoric Pronouns: Am I One?

Marta Recasens, Zhichao Hu, and Olivia Rhinehart

10:30–11:00 **Coffee break**

11:00–12:30 **Session 2**

11:00–11:30 Experiments on bridging across languages and genres

Yulia Grishina

11:30–12:00 Bridging Relations in Polish: Adaptation of Existing Typologies

Maciej Ograniczuk and Magdalena Zawiławska

12:00–12:30 Beyond Identity Coreference: Contrasting Indicators of Textual Coherence in English and German

Kerstin Kunz, Ekaterina Lapshinova-Koltunski, and José Manuel Martínez

12:30–2:00 **Lunch break**

2:00–3:30 **Session 3**

2:00–3:00 **Invited talk: A Bayesian Model of Pronoun Production and Interpretation (Andrew Kehler)**

3:00–3:30 Exploring the steps of Verb Phrase Ellipsis

Zhengzhong Liu, Edgar González Pellicer, and Dan Gillick

3:30–4:00 **Coffee break**

4:00–5:30 **Session 4**

4:00–4:20 Anaphoricity in Connectives: A Case Study on German

Manfred Stede and Yulia Grishina

4:20–4:30 **One minute madness for posters**

4:30–5:30 **Poster session**

- Abstract Coreference in a Multilingual Perspective: a View on Czech and German

Anna Nedoluzhko and Ekaterina Lapshinova-Koltunski

- Antecedent Prediction Without a Pipeline

Sam Wiseman, Alexander M. Rush, and Stuart M. Shieber

- Bridging Corpus for Russian in comparison with Czech

Anna Roitberg and Anna Nedoluzhko

- Coreference Resolution for the Basque Language with BART
Ander Soraluze, Olatz Arregi, Xabier Arregi, Arantza Diaz de Ilarraz, Mijail Kabadgov, and Massimo Poesio
- Error analysis for anaphora resolution in Russian: new challenging issues for anaphora resolution task in a morphologically rich language
Svetlana Toldova, Ilya Azerkovich, Alina Ladygina, Anna Roitberg, and Maria Vasilyeva
- How to Handle Split Antecedents in Tamil?
Vijay Sundar Ram and Sobha Lalitha Devi
- When Annotation Schemes Change Rules Help: A Configurable Approach to Coreference Resolution beyond OntoNotes
Amir Zeldes and Shuo Zhang

Optional Excursion

Thursday, June 16, 2016, 6:00

Marine Corps Air Station Miramar

The optional offsite excursion is a trip to the Marine Corps Air Station Miramar. This trip includes luxury motor coach transportation, guided officers tour of flight deck and airplanes, and Officers' Club dinner with open beer/wine bar.

The excursion must be booked separately at registration. Space is limited.

Workshop 9: Workshop on Computational Approaches to Deception Detection

Organizers: *Tommaso Fornaciari, Eileen Fitzpatrick, and Joan Bachenko*

Venue: Spinnaker 1

Friday, June 17, 2016

9:00–10:00 **Keynote: Multimodal Deception Detection using Real-life Trial Data (Rada Mihalcea)**

10:00–11:45 **Session 1**

- Account Deletion Prediction on RuNet: A Case Study of Suspicious Twitter Accounts Active During the Russian-Ukrainian Crisis
Svitlana Volkova and Eric Bell
- Fake News or Truth? Using Satirical Cues to Detect Potentially Misleading News
Victoria Rubin, Niall Conroy, Yimin Chen, and Sarah Cornwell
- Using the verifiability of details as a test of deception: A conceptual framework for the automation of the verifiability approach
Bennett Kleinberg, Galit Nahari, and Bruno Verschueren

11:45–12:00 **Break**

12:00–1:10 **Session 2**

- Estimating the amenability of new domains for deception detection
Eileen Fitzpatrick and Joan Bachenko
- The Use of Second Life for Deception Detection Research
Stephen Kunath and Kevin McCabe

1:10–3:00 **Break**

3:00–4:10 **Session 3**

- Identifying Individual Differences in Gender, Ethnicity, and Personality from Dialogue for Deception Detection
Sarah Ita Levitan, Yocheved Levitan, Guozhen An, Michelle Levine, Rivka Levitan, Andrew Rosenberg, and Julia Hirschberg
- Individual Differences in Strategic Deception
Scott Appling and Erica Briscoe

4:10–5:00 **Demo**

5:00–5:30 **Discussion**

Workshop 10: Workshop on Discontinuous Structures in Natural Language Processing

Organizers: *Wolfgang Maier, Sandra Kübler, and Constantin Orasan*

Venue: Spinnaker 2

Friday, June 17, 2016

9:30–10:00 An LFG Account of Discontinuous Nominal Expressions
Liselotte Snijders

10:00–10:30 Non-projectivity and valency
Zdeňka Urešová, Eva Fučíková, and Jan Hajič

10:30–11:00 **Coffee break**

11:00–12:15 **Invited Talk: Finite automata for free word order languages (David Chiang)**

12:15–12:45 Machine Translation of Non-Contiguous Multiword Units
Anabela Barreiro and Fernando Batista

12:45–2:30 **Lunch break**

2:30–3:00 Discontinuous VP in Bulgarian
Elisaveta Balabanova

3:00–3:30 Discontinuous Genitives in Hindi/Urdu
Sebastian Sulger

4:00–4:30 Discontinuous parsing with continuous trees
Wolfgang Maier and Timm Lichte

4:30–5:00 Discontinuity (Re)²-visited: A Minimalist Approach to Pseudoprojective Constituent Parsing
Yannick Versley

5:00–5:45 **Panel discussion**

Workshop 11: The 4th Workshop on EVENTS: Definition, Detection, Coreference, and Representation

Organizers: *Eduard Hovy, Teruko Mitamura, Martha Palmer, and Tim O'Gorman*

Venue: Marina 1

Friday, June 17, 2016

9:00–10:30 **Events and Ontologies: A debate between Colin Baker, Sergei Nirenburg and Marjorie McShane**

11:00–12:30 **Trusting human or automated analysis: A debate between Dan Roth and Eduard Hovy**

1:00–2:30 **Poster Session**

- "Making the News": Identifying Noteworthy Events in News Articles
Shyam Upadhyay, Christos Christodoulopoulos, and Dan Roth
- Annotation of causal and aspectual structure of events in RED: a preliminary report
William Croft, Pavlina Peskova, and Michael Regan
- Multimodal Use of an Upper-Level Event Ontology
Claire Bonial, David Tahmoush, Susan Windisch Brown, and Martha Palmer
- A Comparison of Event Representations in DEFT
Ann Bies, Zhiyi Song, Jeremy Getman, Joe Ellis, Justin Mott, Stephanie Strassel, Martha Palmer, Teruko Mitamura, Marjorie Freedman, Heng Ji, and Tim O'Gorman
- Event Nugget and Event Coreference Annotation
Zhiyi Song, Ann Bies, Stephanie Strassel, Joe Ellis, Teruko Mitamura, Hoa Trang Dang, Yukari Yamakawa, and Sue Holm
- Constructing a Dictionary Describing Feature Changes of Arguments in Event Sentences
Tetsuaki Nakamura and Daisuke Kawahara
- CaTeRS: Causal and Temporal Relation Scheme for Semantic Annotation of Event Structures
Nasrin Mostafazadeh, Alyson Grealish, Nathanael Chambers, James Allen, and Lucy Vanderwende

2:30–3:30 **Comparison of event annotations on shared task (discussion led by Martha Palmer)**

4:00–5:00 **Event Sequencing, Annotation and Evaluation (discussion led by Teruko Mitamura)**

Workshop 12: The 4th Workshop on Metaphor in NLP (Meta4NLP)

Organizers: *Beata Beigman Klebanov, Ekaterina Shutova, and Patricia Lichtenstein*

Venue: Marina 2

Friday, June 17, 2016

9:00–9:05 **Opening remarks**

9:05–10:00 **Invited talk: Lera Boroditsky** “Metaphors we think and feel with”

10:00–10:30 Finding metaphorical triggers through source (not target) domain
lexicalization patterns
Jenny Lederer

10:30–11:00 **Coffee break**

11:00–11:30 **NAACL paper:** *Beata Beigman Klebanov, Chee Wee Leong, Elkin Dario Gutierrez, Ekaterina Shutova and Michael Flor* “Semantic classifications for detection of verb metaphors”

11:30–12:00 Detecting novel metaphor using selectional preference information
Hessel Haagsma and Johannes Bjerva

12:00–12:30 **NAACL paper:** *Maximilian Köper and Sabine Schulte im Walde* “Distinguishing Literal and Non-Literal Usage of German Particle Verbs”

12:30–2:00 **Lunch**

2:00–2:30 Supervised Metaphor Detection using Conditional Random Fields
Sunny Rai, Shampa Chakraverty, and Devendra K. Talyal

2:30–3:00 Token-Level Metaphor Detection using Neural Networks
Erik-Lân Do Dinh and Iryna Gurevych

3:00–3:30 Round Up The Usual Suspects: Knowledge-Based Metaphor Generation
Tony Veale

3:30–4:00 **Coffee break**

4:00–5:00 **Invited talk: Seana Coulson** “Creative language: metaphor and conceptual blending”

Workshop 13: Workshop on Multilingual and Crosslingual Methods in NLP

Organizers: *Dipanjan Das, Chris Dyer, Manaal Faruqui, and Yulia Tsvetkov*

Venue: Marina 3

Friday, June 17, 2016

9:15–9:30 **Opening Remarks (Yulia Tsvetkov)**

9:30–10:10 **Invited talk: Evaluation by Compression (Kevin Knight)**

10:10–10:50 **Invited talk: Multi-way, Multilingual Neural Machine Translation (Kyunghyun Cho)**

10:50–11:10 **Coffee Break**

11:10–11:50 **Invited talk: The Case for a Coarse-grained Multilingual Representation of Case and Adposition Semantics (Nathan Schneider)**

11:50–12:30 **Invited talk: To be decided (Chris Dyer)**

12:30–1:30 **Lunch and setting posters**

1:30–1:50 Comparing Fifty Natural Languages and Twelve Genetic Languages Using Word Embedding Language Divergence (WELD) as a Quantitative Measure of Language Distance
Ehsaneddin Asgari and Mohammad R.K. Mofrad

2:00–3:30 **Posters and coffee**

- Learning Cross-lingual Representations with Matrix Factorization
Hanan Aldarmaki and Mona Diab
- Should Have, Would Have, Could Have. Investigating Verb Group Representations for Parsing with Universal Dependencies.
Miryam de Lhoneux and Joakim Nivre
- Cross-lingual Dependency Transfer : What Matters? Assessing the Impact of Pre- and Post-processing
Ophélie Lacroix, Guillaume Wisniewski, and François Yvon
- Enhancing Automatic Wordnet Construction Using Word Embeddings
Feras Al Tarouti and Jugal Kalita
- Cross-lingual alignment transfer: a chicken-and-egg story?
Lauriane Aufrant, Guillaume Wisniewski, and François Yvon
- Leveraging Data-Driven Methods in Word-Level Language Identification for a Multilingual Alpine Heritage Corpus
Ada Wan
- Learning Translations for Tagged Words: Extending the Translation Lexicon of an ITG for Low Resource Languages
Markus Saers and Dekai Wu
- Comparing Fifty Natural Languages and Twelve Genetic Languages Using Word Embedding Language Divergence (WELD) as a Quantitative Measure of Language Distance
Ehsaneddin Asgari and Mohammad R.K. Mofrad

- 3:30–4:10 **Invited talk: Cross-lingual and Unsupervised Learning of Semantic Representations (Ivan Titov)**
- 4:10–4:50 **Invited talk: Unsupervised Modeling of Code-Switching and Orthographic Variation, and its Application to the Study of Digital Humanities (Dan Garrette)**
- 4:50–5:30 **Invited talk: Cross-lingual Learning of Universalized Morphosemantics (David Yarowsky)**
- 5:30–5:45 **Best paper and poster awards**

Workshop 14: AKBC: 5th Workshop on Automated Knowledge Base Construction

Organizers: *Jay Pujara, Tim Rocktäschel, Danqi Chen, and Sameer Singh*

Venue: Executive 3AB

Friday, June 17, 2016

9:00–9:10 Opening Remarks

9:10–9:40 Joint Compositional Learning from Text and Knowledge Bases (Kristina Toutanova)

9:40–10:10 The Allen AI Science Challenge: Results, Lessons, and Open Questions (Oren Etzioni)

10:10–11:00 Morning Poster Session and Coffee Break

- Using Graphs of Classifiers to Impose Constraints on Semi-supervised Relation Extraction
Lidong Bing, William Cohen, Bhuvan Dhingra, and Richard Wang
- Discovering Entity Knowledge Bases on the Web
Andrew Chisholm, Will Radford, and Ben Hachey
- IKE - An Interactive Tool for Knowledge Extraction
Bhavana Dalvi, Sumithra Bhakthavatsalam, Chris Clark, Peter Clark, Oren Etzioni, Anthony Fader, and Dirk Groeneveld
- Incorporating Selectional Preferences in Multi-hop Relation Extraction
Rajarshi Das, Arvind Neelakantan, David Belanger, and Andrew McCallum
- Knowledge Base Population for Organization Mentions in Email
Ning Gao, Mark Dredze, and Douglas Oard
- Enriching Wikidata with Frame Semantics
Hatem Mousselly Sergieh and Iryna Gurevych
- Demonyms and Compound Relational Nouns in Nominal Open IE
Harinder Pal and Mausam
- But What Do We Actually Know?
Simon Razniewski, Fabian Suchanek, and Werner Nutt
- Learning Knowledge Base Inference with Neural Theorem Provers
Tim Rocktäschel and Sebastian Riedel

11:00–11:30 Andrew McCallum’s Mysterious Production of Facts (Talk TBA) (Andrew McCallum)

11:30–12:00 Look Ma, No Neurons: Using Explicit Inference Rules to Complete a KB (William Cohen)

12:00–1:20 Lunch Break and Morning Posters

1:20–1:50 Contributed Talks 1

**1:20–1:35 The Physics of Text: Ontological Realism in Information Extraction
*Stuart Russell, Ole Torp Lassen, Justin Uang, and Wei Wang***

1:35–1:50 Know2Look: Commonsense Knowledge for Visual Search
Sreyasi Nag Chowdhury, Niket Tandon, and Gerhard Weikum

1:50–2:15 Meaningful Discourses (Talk TBA) (Christopher Manning)

2:15–2:40 Common Sense and Language (Benjamin Van Durme)

2:40–3:10 Contributed Talks 2

2:40–2:55 Row-less Universal Schema

Patrick Verga and Andrew McCallum

2:55–3:10 An Attentive Neural Architecture for Fine-grained Entity Type Classification

Sonse Shimaoka, Pontus Stenetorp, Kentaro Inui, and Sebastian Riedel

3:10–4:00 Afternoon Poster Session and Coffee Break

- Know2Look: Commonsense Knowledge for Visual Search
Sreyasi Nag Chowdhury, Niket Tandon, and Gerhard Weikum
- Regularizing Relation Representations by First-order Implications
Thomas Demeester, Tim Rocktäschel, and Sebastian Riedel
- Applying Universal Schemas for Domain Specific Ontology Expansion
Paul Groth, Sujit Pal, Darin McBeath, Brad Allen, and Ron Daniel
- Design of Word Association Games using Dialog Systems for Acquisition of Word Association Knowledge
Yuichiro Machida, Daisuke Kawahara, Sadao Kurohashi, and Manabu Sassano
- Call for Discussion: Building a New Standard Dataset for Relation Extraction Tasks
Teresa Martin, Fiete Botschen, Ajay Nagesh, and Andrew McCallum
- The Physics of Text: Ontological Realism in Information Extraction
Stuart Russell, Ole Torp Lassen, Justin Uang, and Wei Wang
- A Comparison of Weak Supervision methods for Knowledge Base Construction
Ameet Soni, Dileep Viswanathan, Niranjan Pachaiyappan, and Sriraam Natarajan
- An Attentive Neural Architecture for Fine-grained Entity Type Classification
Sonse Shimaoka, Pontus Stenetorp, Kentaro Inui, and Sebastian Riedel
- A Factorization Machine Framework for Testing Bigram Embeddings in Knowledgebase Completion
Johannes Welbl, Guillaume Bouchard, and Sebastian Riedel
- Row-less Universal Schema
Patrick Verga and Andrew McCallum

4:00–4:25 Querying Unnormalized and Incomplete Knowledge Bases (Percy Liang)

4:25–4:50 Memory Networks for Language Understanding: Successes and Challenges (Antoine Bordes)

4:50–5:30 Afternoon Speaker Panel (Christopher Manning, Benjamin Van Durme, Percy Liang, Antoine Bordes)

5:30–5:45 Closing Remarks

5:45–6:15 Evening Poster Session

Workshop 15: TextGraphs: 10th Workshop on Graph-based Methods for Natural Language Processing

Organizers: *V.G. Vinod Vydiswaran, Martin Riedl, and Tanmoy Chakraborty*

Venue: Executive 4

Friday, June 17, 2016

9:00–9:15 **Introduction**

9:15–10:15 **Invited Talk: Inducing Semantic Representations (Ivan Titov)**

10:15–10:30 Embedding Senses for Efficient Graph-based Word Sense Disambiguation
Luis Nieto Piña and Richard Johansson

10:30–11:00 **Coffee Break**

11:00–11:25 Context Tailoring for Text Normalization
Seniz Demir

11:25–11:40 Cross-Lingual Question Answering Using Common Semantic Space
Amir Pouran Ben Veyseh

11:40–12:05 Network Motifs May Improve Quality Assessment of Text Documents
Thomas Arnold and Karsten Weihe

12:05–12:20 Better Together: Combining Language and Social Interactions into a Shared Representation
Yi-Yu Lai, Chang Li, Dan Goldwasser, and Jennifer Neville

12:20–12:40 **Invited Demo: General purpose semantic platform as an information retrieval system (Graphiq)**

12:40–2:00 **Lunch**

2:00–2:15 Visualization of Dynamic Reference Graphs
Ivan Rodin, Ekaterina Chernyak, Mikhail Dubov, and Boris Mirkin

2:15–3:15 **Invited Talk: To be finalized (Eduard Hovy)**

3:15–3:30 **Conclusion**

Anti-harassment policy

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

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

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

The NAACL board members are listed at:

<http://naacl.org/officers/officers-2016.html>

The full policy and its implementation is defined at:

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

Index

- Abdelali, Ahmed, 47, 73
Abdelwahab, Omar, 107
Abramova, Ekaterina, 39
Abura'ed, Ahmed, 114
Adel, Heike, 44, 62
Afantenos, Stergos, 28
Afzal, Naveed, 117
Agarwal, Prakul, 125
Agarwal, Vipul, 48
Aggarwal, Khushboo, 48
Aggarwal, Nitish, 118
Aghaebrahimian, Ahmad, 119
Agichtein, Eugene, 119
Agirre, Eneko, 110, 111, 130
Agrawal, Aishwarya, 77
Ahamed, Nazick, 126
Ahmed, Hany, 112
Ahrendt, Simon, 45
Aker, Ahmet, 111
Al Sailab, Ahmad, 112
Al Tarouti, Feras, 138
Al-Ayyoub, Mahmoud, 105
Al-Khatib, Khalid, 96
Al-Obaidli, Fahad, 113
AL-Smadi, Mohammad, 105
Alagić, Domagoj, 110
Aldarmaki, Hanan, 117, 138
Allen, Brad, 141
Allen, James, 62, 136
Almeida, Hayda, 123
Alonso, Miguel A., 106
Alpert-Abrams, Hannah, 42
Álvarez-López, Tamara, 106, 108
Aly, Mohamed, 105
Amarasinghe, Kasun, 107
Ambati, Bharat Ram, 42, 72
Amini, Aida, 75
Amini, Massih-Reza, 106
Amir, Eyal, 41
Amir, Silvio, 107
Amiri, Hadi, 123
An, Bo, 111
An, Guozhen, 134
An, Lawrence, 122
Ananiadou, Sophia, 111
Anastasakos, Tasos, 48
Anastasopoulos, Antonios, 69
Andreas, Jacob, 101
Andrianov, Ivan, 108
Androutopoulos, Ion, 105, 106, 108
Andruszkiewicz, Piotr, 111
Angelova, Galia, 112
Antunes, Marta, 116
Apidianaki, Marianna, 47, 105
Apostol, Eduard, 107
Appling, Scott, 134
Aranberri, Nora, 130
Arnold, Thomas, 142
Arora, Piyush, 111
Arora, Raman, 71
Arregi, Olatz, 132
Arregi, Xabier, 132
Arthur, Philip, 53
Artstein, Ron, 47
Artsymenia, Artsiom, 116
Asgari, Ehsaneddin, 123, 138
Asher, Nicholas, 28
Asla, Petter, 109

- Asooja, Kartik, 118
Astudillo, Ramón, 107
Ataman, Duygu, 110
Atkins, David, 122
Attardi, Giuseppe, 107
Atyia, Amir, 105
Aufrant, Lauriane, 72, 138
Augenstein, Isabelle, 109
Auli, Michael, 26, 33
Awadallah, Ahmed Hassan, 98
Azerkovich, Ilya, 132
- B. Hashemi, Homa, 47
Bachenko, Joan, 134
Báez, David, 106
Balabanova, Elisaveta, 135
Balahur, Alexandra, 108, 125
Balaji, Peddamuthu, 79
Balchev, Daniel, 112
Baldwin, Timothy, 43, 107, 112, 113
Balikas, Georgios, 106
Ballesteros, Miguel, 33, 34
Banchs, Rafael E., 127
Bandyopadhyay, Sivaji, 109, 114
Banjade, Rajendra, 111, 112, 128
Bansal, Mohit, 59, 71, 90
Bansal, Piyush, 90
Bar, Kfir, 112
Baroni, Marco, 40
Barreiro, Anabela, 135
Barrière, Caroline, 126
Barrón-Cedeño, Alberto, 113
Barros, Marcia, 116
Barzdins, Guntis, 115
Barzilay, Regina, 22, 73, 88, 90
Basiли, Roberto, 112
Bateman, Leila, 122
Batista, Fernando, 135
Batra, Dhruv, 62, 77
Bechara, Hannah, 111
Beck, Daniel, 65
Becker, Karin, 109
Behera, Anil, 122
Beigman Klebanov, Beata, 127, 128
Beinborn, Lisa, 127
Bejenaru, Oana, 107
Bel, Núria, 105
Belanger, David, 66, 140
Belinkov, Yonatan, 112
Beliuha, Maryna, 117
Bell, Eric, 134
Beller, Charley, 119
Bellot, Patrice, 110
Ben, Kao, 119
Bender, Emily M., 9, 10
Bengio, Yoshua, 65, 92
- Benton, Sherry, 122
Berant, Jonathan, 53
Bertero, Dario, 28
Bethard, Sean, 119
Bethard, Steven, 98, 105, 111, 115, 123
Bhakthavatsalam, Sumithra, 140
Bhatt, Himanshu Sharad, 79
Bhattacharya, Prasanta, 46
Bhattacharyya, Pushpak, 9, 18, 126
Bicici, Ergun, 118
Biemann, Chris, 53, 108, 117
Bies, Ann, 136
Biklula, Ravi, 48
Bing, Lidong, 140
Bingel, Joachim, 114
Binoun, Yoav, 80
Birch, Alexandra, 24
Bird, Steven, 69
Bisazza, Arianna, 38
Bisk, Yonatan, 33, 59
Bittar, André, 116
Bjerva, Johannes, 115, 137
Björne, Jari, 113
Black, Alan W, 92
Blain, Frederic, 111
Blanco, Eduardo, 98
Boag, William, 118
Bocionek, Christine, 61
Boella, Guido, 118
Bogdanova, Dasha, 88
Bøhler, Henrik, 109
Boies, Daniel, 48
Boltuzic, Filip, 110
Bonadiman, Daniele, 88
Bond, Francis, 117
Bonial, Claire, 136
Bontcheva, Kalina, 109
Booten, Kyle, 74, 121
Bordea, Georgeta, 115
Bos, Johan, 115
Botschen, Fiete, 141
Bouamor, Houda, 73
Bouchard, Guillaume, 141
Boyanov, Martin, 112
Boyd-Graber, Jordan, 67, 69, 101, 119
Bragg, Jonathan, 66
Brandt, Lauritz, 115
Breslin, John G., 107, 108
Brew, Chris, 123
Briones, Gerard, 107
Briscoe, Erica, 134
Briscoe, Ted, 40, 127, 128
Brockett, Chris, 28
Brooke, Julian, 113
Brun, Caroline, 108
Brunk, Cliff, 90

- Brychcín, Tomáš, 108, 111, 112
Brzezinski, Dariusz, 106
Buechel, Sven, 126
Buhagiar, Mark, 117
Buitelaar, Paul, 115, 118
Bulat, Luana, 46
Bullard, Joseph, 122
Bungum, Lars, 106
Burstein, Jill, 127, 128
Buscaldi, Davide, 118
Butler, Alastair, 115, 130
Byrne, Bill, 65
- Çetin, Fatih Samet, 108
C. De Souza, Jose G., 110
Cahill, Aoife, 127, 128
Callison-Burch, Chris, 79, 98
Calvo, Hiram, 121
Calvo, Rafael A., 123
Cano-Basave, Amparo Elizabeth, 96
Carbonell, Jaime, 59, 116
Cardie, Claire, 37
Carman, Mark, 126
Carpuat, Marine, 69, 76, 92, 105, 110
Caselli, Richard, 76
Caselli, Tommaso, 116
Cassidy, Taylor, 79
Castelli, Vittorio, 88
Castillo, Esteban, 106
Castro Ferreira, Thiago, 41
Celikyilmaz, Asli, 48
Cer, Daniel, 105
Cervantes, Ofelia, 106
Chai, Joyce Y., 31
Chakraborty, Rupak, 125
Chakraborty, Tanmoy, 142
Chakraverty, Shampa, 137
Chambers, Nathanael, 62, 136
Chan, Ho Yin Ricky, 80
Chandar, Sarath, 9, 12, 31
Chandramohan, Senthilkumar, 48
Chang, Baobao, 41
Chang, Ching-Yun, 41
Chang, Nai-Wen, 123
Chang, Shih-fu, 79
Chase, Paul, 119
Chaturvedi, Snigdha, 101
Che, Wanxiang, 115
Chen, Chih-Wei, 123
Chen, Danqi, 140
Chen, Huadong, 77
Chen, Huanhuan, 71
Chen, Hu, 40
Chen, Jiajun, 77
Chen, Lu, 61
Chen, Wei-Te, 115
- Chen, Wei, 109
Chen, Xiaobin, 127
Chen, Xinlei, 56
Chen, Yimin, 134
Chen, Yue, 109
Cheng, Doreen, 61
Cheng, Shambo, 77
Chernyak, Ekaterina, 142
Chernyshevich, Maryna, 108, 117
Cherry, Colin, 25, 65, 75, 105
Chiang, David, 24, 69
Chiarcos, Christian, 98
Chikka, Veera Raghavendra, 116
Chinkina, Maria, 128
Chisholm, Andrew, 140
Cho, Eunah, 79
Cho, Kyunghyun, 37, 65, 92
Cho, Yeonsuk, 128
Chodorowska, Krystyna, 111
Choi, Jinho D., 34, 56
Chopra, Sumit, 26
Choubey, Prafulla, 108, 114
Chow, Marisa, 48
Christodoulopoulos, Christos, 136
Christopoulou, Fenia, 106
Chrupała, Grzegorz, 40
Cieri, Christopher, 122
Ciobanu, Octavian, 107
Ciubotariu, Calin-Cristian, 107
Clark, Chris, 140
Clark, Peter, 140
Clark, Stephen, 33, 46
Cleuziou, Guillaume, 117
Cocos, Anne, 98
Cohan, Arman, 116, 123
Cohen, William, 9, 16, 140
Cohn, Trevor, 65, 69, 77
Collins, Michael, 53
Conrad, Stefan, 111
Conroy, Niall, 134
Constant, Matthieu, 73
Contractor, Danish, 26
Cook, Paul, 118
Coppersmith, Glen, 123
Cordeiro, Silvio, 113
Cornwell, Sarah, 134
Costa-Montenegro, Enrique, 106, 108
Cotterell, Ryan, 55
Couto, Francisco M., 116
Couture, Sara, 109
Cozza, Vittoria, 106
Croce, Danilo, 112
Crocker, Matthew, 80
Croft, William, 136
Cromierès, Fabien, 24
Crook, Paul, 48

- Cubaud, Pierre, 47
Cui, Yiming, 69
Culinovic, Filip, 110
Cummins, Ronan, 127, 129
Currey, Anna, 82
Cyphers, Scott, 112
- Da San Martino, Giovanni, 105, 113
Dai, Hong-Jie, 123
Dai, Xin-Yu, 77
Dakota, Daniel, 109
Dalvi, Bhavana, 140
Dandapat, Sandipan, 79
Danescu-Niculescu-Mizil, Cristian, 28, 45, 96
Dang, Hoa Trang, 136
Daniel, Ron, 141
Darabana, Diana, 107
Darrell, Trevor, 101
Darwish, Kareem, 47
Das, Dipanjan, 53, 67, 138
Das, Dipankar, 109, 114, 117
Das, Rajarshi, 140
Daudaravicius, Vidas, 127
Daumé III, Hal, 69, 101, 115, 119, 123
Davier, Alina von, 127
Davoodi, Elnaz, 114
Daza, Angel, 121
De Clercq, Orphee, 105
De Luca, Valeria, 106
de Marneffe, Marie-Catherine, 39
De Meyere, Damien, 116
Dean, Harry, 123
Dehdari, Jon, 47, 75, 82
Dehghani, Morteza, 125
Delmonte, Rodolfo, 112
Delogu, Francesca, 80
Demarest, Bradford, 109
Dembe, Aaron, 122
Demberg, Vera, 45, 78
Demeester, Thomas, 141
Demir, Seniz, 142
Deng, Lingjia, 125
Deoskar, Tejaswini, 42
Derczynski, Leon, 115
Deriu, Jan, 106
Dernoncourt, Franck, 44
Desmet, Bart, 123
Devlin, Jacob, 77
Dey, Anik, 80
Dhingra, Bhuwan, 140
Di Caro, Luigi, 118
Di Eugenio, Barbara, 125
Diab, Mona, 109, 117, 125, 138
Dias Cardoso, Pedro, 125
Dias, Marcelo, 109
Díaz de Ilarza, Arantza, 132
- Díaz Galiano, Manuel Carlos, 128
Dickinson, Markus, 127, 128
Dimitrovski, Ivica, 106
Ding, Yu, 115
Do Dinh, Erik-Lân, 137
Dobson, Richard, 122
Dolan, Bill, 28
Dong, Minghui, 105
Dounar, Palina, 116
Doval, Yerai, 106
Dras, Mark, 114, 123
Dredze, Mark, 71, 72, 78, 140
Drozd, Aleksandr, 48
Du, Steven, 106
Duan, Manjuan, 116
Dubov, Mikhail, 142
Dubremetz, Marie, 121
Dueck, Amylou, 76
Duh, Kevin, 130
Dumas Ang, Siena, 88
Duong, Long, 69, 112
Duque, Andres, 111
Durrani, Nadir, 47, 73
Durrett, Greg, 77
Dutta, Rina, 122, 123
Dyer, Chris, 33, 34, 38, 55, 59, 65, 92, 99, 116, 138
- Ebadi, Ashkan, 122
Ebert, Sebastian, 61, 119
Ecker, Brian, 109
Eichstaedt, Johannes, 125
Eisenstein, Jacob, 38, 90, 99
Eisner, Jason, 55
Ekbal, Asif, 108
El-Beltagy, Samhaa R., 108, 110
Elfandy, Heba, 109
Ellis, Joe, 136
Elmaghraby, Adel, 107
Elsayed, Tamer, 112
Erk, Katrin, 74
Errecalde, Marcelo Luis, 126
Eryiğit, Gülşen, 105, 108
Escalante, Hugo Jair, 126
Espinosa Anke, Luis, 114, 117
Estes, Alex, 121
Esuli, Andrea, 106
Ettinger, Allyson, 92
Etzioni, Oren, 88, 140
Eyecioglu, Asli, 118
- Fader, Anthony, 140
Fairon, Cédrick, 116, 117
Falk, Stefan, 108
Faralli, Stefano, 117
Farhadi, Ali, 32

- Faruqui, Manaal, 55, 71, 92, 138
Favre, Benoit, 107
Fei, Geli, 44
Feizollahi, Zhaleh, 48
Feldman, Anna, 121
Felice, Mariano, 128
Feltracco, Anna, 112
Feng, Yukun, 48
Ferguson, Emily, 122
Fernández, Raquel, 39, 40
Fernández-Gavilanes, Milagros, 106, 108
Ferraro, Francis, 77
Ferreira, William, 75
Ferret, Olivier, 115
Figueiró, Gonçalo, 116
Figueroa-Nazuno, Jesús, 121
Filice, Simone, 73, 112
Finch, Andrew, 41, 73
Firat, Orhan, 65
Fitzpatrick, Eileen, 134
Flanigan, Jeffrey, 59, 116
Flickinger, Dan, 9, 10, 127
Flor, Michael, 127
Floorean, Cosmin, 107
Florian, Radu, 88
Foland, William, 116
Fomicheva, Marina, 111
Fornaciari, Tommaso, 134
Foster, Jennifer, 88
Fournier, Sébastien, 110
Francis-Landau, Matthew, 77
Franco-Peña, Hector-Hugo, 123, 128
Franco-Salvador, Marc, 112
Fraser, Kathleen C., 122
Frasincar, Flavius, 109
Frazier, Peter, 37
Fredriksen, Valerij, 106
Freedman, Marjorie, 136
Freihat, abed Alhakim, 110
Friedenberg, Meir, 123
Friedrichs, Jasper, 107
Fries, Jason, 116
Fu, Cheng, 111
Fung, Pascale, 28, 80
Fučíková, Eva, 135

Gaddy, David, 90
Gagliano, Andrea, 121
Gainer, Alesia, 47
Galanis, Dimitris, 105
Galley, Michel, 28, 77
Gambäck, Björn, 106
Gamon, Michael, 98
Ganesan, Kavita, 126
Ganesh HB, Barathi, 118
Gao, Hang, 106

Gao, Jianfeng, 28
Gao, Ning, 140
Gao, Qiaozhi, 31
Gao, Wei, 105
García Cumbreras, Miguel Ángel, 128
Garcia Flores, Jorge, 118
Garrette, Dan, 42
Gautam, Dipesh, 111, 128
Gašić, Milica, 28, 29
Ge, Shili, 130
Gelbukh, Alexander, 117
Gella, Spandana, 31
Gemechu, Debela Tesfaye, 49
Genabith, Josef van, 47, 69, 75, 111, 114, 117
Gencheva, Pepa, 112
Georgila, Kallirroi, 47
Getman, Jeremy, 136
Ghaffari, Parsa, 107, 108
Gharbieh, Waseem, 118
Ghosh, Aniruddha, 126
Gildea, Daniel, 115
Gillick, Dan, 90, 131
Gimpel, Kevin, 111
Ginsberg, Allen, 119
Giorgis, Stavros, 106
Girju, Roxana, 41
Girlea, Codruța, 41
Girshick, Ross, 77
Gispert, Adrià de, 65
Gkotsis, George, 122, 123
Gladkova, Anna, 48
Glass, Jim, 110, 112
Gliga, Mihail, 107
Goebel, Randy, 120
Goharian, Nazli, 116, 123
Goldberg, Sean, 76
Goldberg, Yoav, 101
Goldenstein, Jan, 126
Goldsmith, John, 47
Goldwasser, Dan, 142
Gomaa, Amr, 112
Gombar, Paula, 110
Gomez, Helena, 106
Gómez-Rodríguez, Carlos, 106
González Pellicer, Edgar, 131
Gonzalez, Graciela, 76, 107
Gonzalez-Agirre, Aitor, 110
González-Castaño, Francisco Javier, 106, 108
Gonzenbach, Maurice, 106
Goodman, James, 115
Goodman, Michael, 127
Goot, Erik van der, 125
Gormley, Matthew R., 71
Gosko, Didzis, 115
Goutte, Cyril, 117
Graham, Yvette, 24

- Grealish, Alyson, 136
Greenberg, Clayton, 80
Grimes, Seth, 125
Grimm, David, 115
Grishina, Yulia, 131
Grishman, Ralph, 37
Groeneveld, Dirk, 140
Groth, Paul, 141
Grouin, Cyril, 116
Guan, Rachel, 79
Guestrin, Carlos, 80
Guha, Anupam, 101, 119
Guo, Hongyu, 67
Guo, Wu, 41
Gupta, Bhavesh, 125
Gupta, Pankaj, 44
Gupta, Rohit, 111
Gupta, Sakshi, 90
Gurevych, Iryna, 127, 137, 140
Guzmán, Francisco, 47, 73, 113
Gyawali, Binod, 127
- Ha, Thanh-Le, 79
Haagsma, Hessel, 115, 137
Haas, Carolin, 59
Hachey, Ben, 123, 140
Haddow, Barry, 24
Haduong, Nikita, 109
Haffari, Gholamreza, 38, 59, 65, 77
Hagen, Matthias, 96
Hahn, Nicolas, 109
Hahn, Udo, 126
Hajishirzi, Hannaneh, 75, 88
Hajič, Jan, 135
Hakkani-Tur, Dilek, 55
Hamdan, Hussam, 107, 108
Hamidian, Sardar, 125
Hammer, Hugo Lewi, 126
Han, Jiawei, 79
Han, Xianpei, 111
Handleman, Theodore, 109
Händschke, Sebastian G. M., 126
Hansart, Charlotte, 116
Hao, Jiangang, 127
Hao, Shuai, 44
Harackiewicz, Judith, 128
Hardalov, Momchil, 112
Harsley, Rachel, 125
He, Anni, 109
He, He, 69, 119
He, Hua, 67, 111
He, Lei, 44
He, Xiaodong, 62, 77, 99
He, Yulan, 96
He, Yunchao, 44, 107
Hearst, Marti A., 74, 121, 128
- Heesacker, Martin, 122
Heilman, Michael, 126, 128
Hench, Christopher, 121
Hendler, James, 79
Henry, Sam, 110
Hercig, Tomáš, 108
Higgins, Derrick, 126
Hill, Felix, 92
Hill, Jennifer, 127
Hintz, Gerold, 49
Hirschberg, Julia, 134
Hirst, Graeme, 122, 126
Hoang, Cong Duy Vu, 65, 77
Hokamp, Chris, 111
Holenstein, Roman, 48
Hollingshead, Kristy, 122
Holm, Sue, 136
Homan, Christopher, 122
Hoogeveen, Doris, 112
Horbach, Andrea, 129
Hosseini, Mohammad Javad, 113
Hoste, Véronique, 123
Hoste, Veronique, 105
Hovy, Dirk, 74, 81, 110
Hovy, Eduard, 56, 70, 88, 99, 136
Hrisca, Marius-Valentin, 107
Hsu, Kuang-Yi, 79
Hsu, Wei-Ning, 112
Htait, Amal, 110
Hu, Jun, 44
Hu, Yu, 41
Hu, Zhichao, 131
Huang, Chien-Tsung, 79
Huang, Dongyan, 105
Huang, Heng, 110, 116
Huang, Heyan, 117
Huang, Shujian, 77
Huang, Ting-Hao (Kenneth), 77
Hubbard, Tim, 122
Hurtado, Lluís-F, 107
Hwa, Rebecca, 39, 47
Hwang, Jena D., 122
Hwang, Mei-Yuh, 40
- Iftene, Adrian, 107
Igarashi, Yuki, 109
Iglesias, Gonzalo, 65
Imel, Zac, 122
Ingersoll, Keith, 126
Inkpen, Diana, 121
Inui, Kentaro, 63, 109, 141
Iosif, Elias, 106, 126
Islam, Aminul, 118
Itoh, Hideo, 117
Itti, Laurent, 49
Iyyer, Mohit, 101, 119

- Jaakkola, Tommi, 88, 90
Jacobs, Gilles, 123
Jaech, Aaron, 55
Jagannatha, Abhyuday N, 42
Jaggi, Martin, 106
Jahren, Brage Ekroll, 106
Jain, Prachi, 26
Jain, Rohit, 81
Jain, Sarthak, 82
Jat, Sharmistha, 79, 112
Jelesnianska, Adrianna, 126
Jeong, Minwoo, 48
Jha, Vandana, 126
Ji, Donghong, 113
Ji, Heng, 34, 37, 75, 79, 115, 136
Ji, Yangfeng, 38
Jia, Chang e, 113
Jiang, Daxin, 99
Jiang, Hui, 40, 41
Jiang, Jing, 98
Jiang, Mengxiao, 109
Jiang, Xin, 119
Jimenez, Sergio, 118
Jiménez-Zafra, Salud Marfa, 105, 126
Jin, Lifeng, 116
Johannsen, Anders, 110
Johansson, Richard, 142
Johnson, Mark, 42
Johnson, Travis A., 76
Jonnagaddala, Jitendra, 123
Jørgensen, Anna, 74
Joshi, Aditya, 126
Joshi, Hrishikesh, 88
Joty, Shafiq, 56, 113
Juncal-Martínez, Jonathan, 106, 108
Jurafsky, Dan, 56
Jurgens, David, 105, 115
- K P, Soman, 114
K R, Venugopal, 126
Kabadjov, Mijail, 132
Kalita, Jugal, 138
Kamal, Eslam, 112
Kambadur, Prabhanjan, 72
Kaplan, Lance, 79
Kar, Sudipta, 112
Karakanta, Alina, 82
Karan, Mladen, 110
Karnick, Harish, 49
Karpov, Nikolay, 107
Katz, Graham, 119
Kauchak, David, 114
Kaufman, Noah, 109
Kawahara, Daisuke, 115, 136, 141
Kawakami, Kazuya, 34
Kazantseva, Anna, 121
- Kazmi, Mishal, 105, 112
Keller, Bill, 118
Keller, Frank, 31
Kern, Margaret, 125
Kern, Roman, 108
Khalifa, Abdulrahman, 116
Khalil, Talaat, 108
Khan, Omar, 48
Khaprä, Mitesh M., 9, 12, 18, 31
Kiela, Douwe, 31, 46
Kilgour, Kevin, 79
Kim, Hyun, 43
Kim, Sunghwan Mac, 123
Kim, Yoon, 128
Kim, Young-Bum, 34, 48
Kim, Yu-Seop, 34
King, Levi, 127
King, Milton, 118
Kiprov, Yasen, 112
Kirilin, Angelika, 113
Kiritchenko, Svetlana, 61, 73, 105, 125
Klebanov, Beata Beigman, 137
Klein, Dan, 77, 101
Kleinberg, Bennett, 134
Klenner, Manfred, 126
Klerke, Sigrid, 101
Knight, Kevin, 2, 24, 34, 77
Kobayashi, Sosuke, 63, 109
Koch, Johann Frerik, 127
Kochmar, Ekaterina, 127
Kohail, Sarah, 108
Köhler, Jonas, 96
Kohli, Pushmeet, 62, 77
Kokkinos, Filippos, 106
Kolovou, Athanasia, 106
Komatsu, Hiroya, 109
Komninos, Alexandros, 99
Koncel-Kedziorowski, Rik, 55, 75, 88
Kondrak, Grzegorz, 75
Konkol, Michal, 108, 114
Konopik, Miloslav, 112
Kontonatsios, Georgios, 111
Köper, Maximilian, 39
Koppel, Moshe, 121
Korhonen, Anna, 92
Korolov, Rostyslav, 79
Kosseim, Leila, 114
Kotelnikov, Evgeniy, 105
Koutsakis, Polychronis, 126
Koychev, Ivan, 112
KP, Soman, 118
Krahmer, Emiel, 41
Král, Pavel, 110
Krauss, Felix, 113
Krawczyk, Elizabeth, 48
Krejzl, Peter, 109

- Krishnamurthy, Jayant, 53
Krstovski, Kriste, 42, 74
Krüger, Bastian, 79
Kübler, Sandra, 109, 135
Kumar, Amit, 108
Kumar, Anand, 114
Kumar, Ayush, 108
Kumar, Ravi, 28
Kumar, Varun, 123
Kunath, Stephen, 134
Kunchukuttan, Anoop, 9, 18
Kuncoro, Adhiguna, 33
Kundu, Senjuti, 125
Kunz, Kerstin, 131
Kurata, Gakuto, 44
Kurisinkel, Litton J, 81
Kurohashi, Sadao, 24, 115, 141
Kurtz, Michael J., 42
Kuru, Onur, 114
Kushman, Nate, 75
Kwiatkowski, Tom, 53
Kyaw Thu, Ye, 73
- Lacroix, Ophélie, 72, 138
Ladygina, Alina, 132
Lai, K. Robert, 44, 107
Lai, Yi-Yu, 142
Lalitha Devi, Sobha, 132
Lamont, Andrew, 81, 109
Lample, Guillaume, 34, 92
Lamúrias, André, 116
Lan, Man, 107, 109–111, 113
Lango, Mateusz, 106
Lapata, Mirella, 31, 38, 53
Lapshinova-Koltunski, Ekaterina, 69, 131
Lassen, Ole Torp, 140, 141
Lau, Jey Han, 43
Lazaridou, Angeliki, 40
Le Roux, Joseph, 73
Leacock, Claudia, 127
Leary, Ryan, 123
Ledbetter, Scott, 128
Lederer, Jenny, 137
Lee, Hee-Jin, 116
Lee, Honglak, 99
Lee, Jackson, 47
Lee, Ji Young, 44
Lee, Jong-Hyeok, 43
Lee, Kenton, 33
Lee, Lung-Hao, 44, 128
Lee, Mark, 117
Lee, Su-In, 113
Leeuwenberg, Artuur, 116
Lefever, Els, 115
Lei, Tao, 73, 88, 112
Lenc, Ladislav, 110
- Lepage, Yves, 81
Leuski, Anton, 47
Levin, Lori, 92
Levine, Michelle, 134
Levitán, Rivka, 134
Levitán, Sarah Ita, 134
Levitán, Yocheved, 134
Lewis, Mike, 33, 53
Lhoneux, Miryam de, 138
Li, Chang, 142
Li, Chung-Min, 79
Li, Fei, 113
Li, Hang, 9, 14, 70, 119
Li, Hao, 75
Li, Hongzhi, 79
Li, Huayi, 125
Li, Jianfeng, 69
Li, Jiwei, 28, 56
Li, Junyi Jessy, 76
Li, Liangyou, 130
Li, Peng, 110, 116
Li, Sujian, 41
Li, Wen, 109
Li, Xiaoming, 119
Li, Yitong, 112
Li, Zhengyu, 71
Liakata, Maria, 122, 123
Liang, Chao-Chun, 79
Liang, Huizhi, 107, 112
Liang, Percy, 53
Liberman, Mark, 122
Lichte, Timm, 135
Lichtenstein, Patricia, 137
Liebeck, Matthias, 111
Liew, Jasy Suet Yan, 81
Lightgow Serrano, Oscar William, 118
Lin, Bo-Lin, 128
Lin, Christopher H., 66
Lin, Jimmy, 67, 111
Lin, Ruixi, 80
Lin, Weisi, 105
Ling, Wang, 26, 107
Ling, Xiao, 66
Ling, Zhen-Hua, 41
Litman, Diane, 26, 46, 47, 96, 129
Littell, Patrick, 92
Liu, Angli, 66
Liu, Bing, 44
Liu, Can, 109
Liu, Changsheng, 39
Liu, Changsong, 31
Liu, Fei, 26, 113
Liu, Hongfang, 117
Liu, Jiawei, 71
Liu, Lei, 127
Liu, Lemao, 41

- Liu, Linqing, 49
Liu, Quan, 41
Liu, Qun, 24, 70, 130
Liu, Ting, 115
Liu, Weiyi, 107
Liu, Xiaohu, 48
Liu, Xumin, 122
Liu, Xuqin, 109
Liu, Zhengzhong, 70, 131
Liu, Zitao, 26
Lo, Chi-kiu, 117
Locke, Dona, 76
Lopez-Gazpio, Inigo, 110, 111
Loukachevitch, Natalia, 105
Loukina, Anastassia, 128
Lu, Di, 79
Lu, Jianwei, 49
Lu, Liang, 38
Lu, Wei, 56
Lu, Wenpeng, 117
Lu, Yao, 49
Lu, Zhengdong, 9, 14, 119
Lucchi, Aurelien, 106
Luo, Wencan, 26, 34
Luo, Ye, 49
- M, Anand Kumar, 118
Ma, Hao, 9, 13
Ma, Xuezhe, 70
Ma, Zheng, 126
Machacek, Jakub, 108
Machida, Yuichiro, 141
Madhyastha, Pranava Swaroop, 45
Madjarov, Gjorgji, 106
Madnani, Nitin, 127, 128
Magdy, Walid, 110
Magnini, Bernardo, 112
Magnolini, Simone, 112
Magooda, Ahmed, 112
Maharjan, Nabin, 111, 112, 128
Mahgoub, Ashraf, 112
Maier, Wolfgang, 135
Maillard, Jean, 31
Maitra, Promita, 117
Malakasiotis, Prodromos, 106, 108
Malandrakis, Nikolaos, 106
Malhas, Rana, 112
Malmasi, Shervin, 114, 123
Mamani Sanchez, Liliana, 123, 128
Mamidi, Radhika, 90, 114
Manandhar, Suresh, 99, 105
Marcu, Daniel, 34, 59
Marin, Alex, 48
Maritxalar, Montse, 110, 111
Márquez, Lluís, 56, 88, 110, 113
Marsh, Amy, 110
- Marsi, Erwin, 109
Martin, James H., 116
Martin, Maite, 126
Martin, Teresa, 141
Martineau, Justin, 61
Martínez Alonso, Héctor, 114, 117
Martinez Morant, Victor, 107
Martínez, José Manuel, 113, 131
Martínez-Gómez, Pascual, 120
Martínez-Santiago, Fernando, 128
Matsumoto, Yuji, 74
Matsuoka, Satoshi, 48
Mausam, 26, 140
May, Jonathan, 77, 115
Mayorov, Vladimir, 108
McBeath, Darin, 141
McCabe, Kevin, 134
McCallum, Andrew, 66, 140, 141
McClosky, David, 71
McCrae, John Philip, 118
McInnes, Bridget, 107, 117
Medero, Julie, 114
Mediani, Mohammed, 79
Mehdad, Yashar, 26
Mehr, Leo, 121
Mehrotra, Rishabh, 46
Mei, Hongyuan, 59
Mei, Jie, 118
Melamud, Oren, 71, 128
Ménard, Pierre André, 126
Mesgar, Mohsen, 96, 128
Meurer, Kevin, 116
Meurers, Detmar, 127, 128
Meurs, Marie-Jean, 123
Meyer, Christian M., 127
Meystre, Stephane, 116
Meza Ruiz, Ivan Vladimir, 118
Miao, Shen-Yu, 79
Michaely, Moty, 121
Mihalcea, Rada, 122
Mihaylov, Todor, 112, 113
Mihaylova, Tsvetomila, 112
Milios, Evangelos, 118
Miller, Timothy, 111
Milli, Smitha, 128
Milne, David N., 123
Mirkin, Boris, 142
Mishra, Pruthwik, 81
Misra, Amita, 109
Misra, Ishan, 77
Mitamura, Teruko, 136
Mitchell, Margaret, 77
Mitchell, Meg, 31
Mitchell, Tom M., 37, 62
Mitkov, Ruslan, 111
Miura, Akiva, 24

- Miyao, Yusuke, 120
Miyoshi, Toshinori, 108
Mizumoto, Tomoya, 74
Modaresi, Pashuton, 111
Moens, Marie-Francine, 116
Mofrad, Mohammad R.K., 123, 138
Mogadala, Aditya, 56
Mohammad, Saif M., 61, 73
Mohammad, Saif, 105, 125, 126
Mohanty, Sharada Prasanna, 80
Mohtarami, Mitra, 112
Molina González, M. Dolores, 126
Montejo Ráez, Arturo, 128
Montes y Gomez, Manuel, 126
Montoyo, Andrés, 125
Monz, Christof, 38
Moon, Sungrim, 116
Mooney, Ray, 62
Moosavi, Nafise Sadat, 70
Morante, Roser, 116
Moreno, Jose G., 117
Moreno-Noguer, Francesc, 45
Morey, Mathieu, 28
MORICEAU, Véronique, 116
Mortensen, David, 92
Moschitti, Alessandro, 34, 73, 88, 110, 112, 113
Mostafazadeh, Nasrin, 62, 77, 136
Motlani, Ravesh, 49, 90
Mott, Justin, 136
Mousselly Sergieh, Hatem, 140
Mozuraitis, Mindaugas, 80
Mrkšić, Nikola, 28, 29
Mubarak, Hamdy, 47, 110
Muis, Aldrian Obaja, 56
Mukherjee, Niloy, 114
Mulholland, Matthew, 128
Müller, Markus, 79
Muralidaran, Vigneshwaran, 81
Murawaki, Yugo, 90
Murugaiah, Brunthavan, 126
Musa, Ryan, 33
- Nabil, Mahmoud, 105
Naets, Hubert, 117
Nag Chowdhury, Sreyasi, 141
Nagata, Masaaki, 75
Nagesh, Ajay, 141
Nahari, Galit, 134
Nakamura, Satoshi, 24, 53
Nakamura, Tetsuaki, 136
Nakazawa, Toshiaki, 24
Nakov, Preslav, 56, 105, 110, 112, 113, 126
Nam, Jinseok, 126
Napoles, Courtney, 79, 127
Napolitano, Diane, 128
- Naradowsky, Jason, 115
Narayanan, Shrikanth, 106
Nasiriany, Soroush, 123
Nat, Gillin, 113
Natarajan, Sriraam, 141
Nedoluzhko, Anna, 131
Neelakantan, Arvind, 140
Negri, Matteo, 110
Nenkova, Ani, 76
Neubig, Graham, 24, 53, 55
Névéol, Aurélie, 115
Neville, Jennifer, 142
Ng, Vincent, 70, 96, 131
Ngo, Kim, 123
Nguyen, Dat Quoc, 42
Nguyen, Huy, 46
Nguyen, Nhung T. H., 111
Nguyen, Thai Son, 79
Nguyen, Thien Huu, 37
Nickerson, Paul, 122
Niculae, Vlad, 45
Niehues, Jan, 79
Nieto Piña, Luis, 142
Nikolova, Ivelina, 112
Niraula, Nobal Bikram, 112, 128
Nivre, Joakim, 121, 138
Niwa, Yoshiki, 108, 116
Noord, Gertjan van, 92
Nutt, Werner, 140
- Ó Séaghdfa, Diarmuid, 29
Oak, Mayuresh, 122
Oard, Douglas, 140
Oates, Tim, 106
Oellrich, Anika, 122, 123
O'Gorman, Tim, 136
Ogrodniczuk, Maciej, 131
Okazaki, Naoaki, 63, 109
Opitz, Juri, 123
Orasan, Constantin, 111, 135
Ordóñez, Vicente, 32
Orozco Camacho, Albert Manuel, 118
Osborne, Miles, 72
Osenova, Petya, 130
Ostendorf, Mari, 28, 55
Ovesdotter Alm, Cecilia, 122
Övrelid, Lilja, 126
Özbey, Can, 108
- P R, Sarath, 116
Pachaiyappan, Niranjana, 141
Packard, Woodley, 9, 10, 127
Paetzold, Gustavo, 41, 110, 113
Pakray, Partha, 117
Pakulska, Katarzyna, 111
Pal, Harinder, 140

- Pal, Sujit, 141
Palakurthi, Ashish, 114
Paljak, Ivan, 110
Palmer, Alexis, 129
Palmer, Martha, 136
Palogiannidi, Elisavet, 106, 126
Pan, Xiaoman, 34, 115
Pan, Yi-Cheng, 40
Panchendarajan, Rrubaa, 126
Panchenko, Alexander, 117
Pancholi, Manan, 109
Panic, Danko, 48
Pantel, Patrick, 98
Papageorgiou, Haris, 105, 106
Parikh, Devi, 62, 77
Paris, Cecile, 123
Parish-Morris, Julia, 122
Park, Gregory, 125
Park, Se-Young, 66
Park, Seong-Bae, 66
Park, SoHyun, 118
Parker, D. Stott, 78
Pateria, Shubham, 108, 114
Patra, Braja Gopal, 109, 114
Patwardhan, Siddharth, 71
Paul, Emily, 121
Paul, Michael, 24
Pavlopoulos, John, 106, 108
Pedersen, Ted, 117, 123
Pemasiri, Akila, 126
Peng, Xiaochang, 115
Peng, Yangtuo, 40
Perez, Julien, 108
Pérez-Rosas, Verónica, 122
Perret, Jérémie, 28
Persing, Isaac, 96
Perumal, Krish, 126
Peskova, Pavlina, 136
Petrocchi, Marinella, 106
Phipps, Chris, 119
Pilán, Ildikó, 128
Pilehvar, Mohammad Taher, 115
Pillias, Clément, 47
Pinheiro, Alexandre, 116
Pink, Glen, 123
Pinter, Yuval, 56
Pinto Avendaño, David, 106
Pla, Ferran, 107
Plas, Lonneke van der, 55
Pocostales, Joel, 117
Poesio, Massimo, 132
Pollack, Philipp, 111
Pontiki, Maria, 105
Ponzetto, Simone Paolo, 117
Popov, Alexander, 130
Porshnev, Alexander, 107
Post, Matt, 79
Potamianos, Alexandros, 106, 126
Potash, Peter, 118
Poupard, Pascal, 112
Pouran Ben Veyseh, Amir, 142
Pradhan, Sameer, 115
Prazak, Ondrej, 112
Preoțiuc-Pietro, Daniel, 125
Priniiski, Stacy, 128
Proaño, Rubén, 122
Prud'hommeaux, Emily, 122
Przybyla, Piotr, 111
Ptucha, Raymond, 122
Puduppully, Ratish, 43
Pujara, Jay, 140
Pusse, Florian, 78
Pustejovsky, James, 115
Puzikov, Yevgeniy, 115
Qadir, Ashequl, 77, 98
Qian, Yao, 44
Qin, Bing, 105
Qu, Lizhen, 42
Quattoni, Ariadna, 45
Queudot, Marc, 123
Quijada, Maury, 114
R, Manikandan, 116
R, Savitha, 126
Raafat, Hazem, 112
Räbiger, Stefan, 105
Radev, Dragomir R., 26, 99
Radford, Will, 123, 140
Radostev, Vasiliy, 48
Rahgozar, Arya, 121
Rahimi, Zahra, 129
Rai, Sunny, 137
Rajendran, Janarthan, 31
Rajtmajer, Václav, 110
Ramanishka, Vasili, 118
Rambow, Owen, 101
Ramesh, Nikhil, 48
Ramisa, Arnau, 45
Ramisch, Carlos, 113
Ranathunga, Surangika, 126
Randeree, Bilal, 110
Rao, Jinfeng, 111
Rao, Sudha, 115
Rapoport, Ari, 43
Rashidi, Parisa, 122, 123
Rashwan, Mohsen, 112
Rastogi, Pushpendre, 55
Ravi, Sujith, 28
Ravindran, Balaraman, 31
Razniewski, Simon, 140
Recasens, Marta, 131

- Reddy, Siva, 53, 72
Refaee, Eshrag, 110
Regan, Michael, 136
Rei, Marek, 129
Reichart, Roi, 43, 56
Reitter, Ehud, 86
Remse, Madeline, 128
Remus, Steffen, 117
Ren, Xiang, 79
Resnicow, Kenneth, 122
Resnik, Philip, 92, 115, 123
Rettiger, Achim, 56
Rexha, Andi, 108
Rey-Villamizar, Nicolas, 123
Reynolds, Robert, 129
Rhinehart, Olivia, 131
Ribeiro, Marco, 80
Richardson, John, 24
Riedel, Sebastian, 140, 141
Riedl, Martin, 53, 142
Rieser, Verena, 110
Riezler, Stefan, 59
Rigau, German, 110
Riloff, Ellen, 61, 77
Riordan, Brian, 128
Ritter, Alan, 105
Robichaud, Jean-Phillipe, 48
Rochette, Alexandre, 48
Rocktäschel, Tim, 140, 141
Rodin, Ivan, 142
Rodríguez, Miguel, 76
Rohrbach, Marcus, 101
Roitberg, Anna, 131, 132
Rojas-Barahona, Lina M., 28, 29
Roller, Stephen, 74, 100
Romanov, Alexey, 118
Romeo, Salvatore, 113
Ronzano, Francesco, 114, 117
Rosenberg, Andrew, 134
Rosenthal, Sara, 105
Rosso, Paolo, 112
Roth, Benjamin, 62, 66
Roth, Dan, 46, 62, 78, 136
Rothe, Sascha, 61
Rotondi, Agata, 117
Rousas, Apostolos, 106
Rouvier, Mickael, 107
Roux, Claude, 108
Roy, Anindya, 125
Roy, Deb, 109
Roy, Shourya, 79
Roy, Subhro, 75, 78
Rubin, Victoria, 134
Rubino, Raphael, 69
Rudakov, Kirill, 107
Ruder, Sebastian, 107, 108
Rudzewitz, Björn, 128
Rudzicz, Frank, 122
Rumshisky, Anna, 118
Ruppenhofer, Josef, 61
Ruppert, Eugen, 117
Rus, Vasile, 111, 112, 128
Rusert, Jon, 117
Rush, Alexander M., 26, 33, 70, 128, 131
Russell, Stuart, 140, 141
Ryant, Neville, 122
Rychalska, Barbara, 111
Sabharwal, Ashish, 88
Sadeque, Farig, 123
Saers, Markus, 138
Sætre, Rune, 109
Sagae, Kenji, 33
Saggion, Horacio, 114, 117
Sajjad, Hassan, 73
Sakti, Sakriani, 53
Salakoski, Tapio, 113
Salameh, Mohammad, 75, 105
Salazar, Cristobal, 72
Salehi, Bahar, 112
Samei, Borhan, 128
Sands, Allison, 110
Sarabi, Zahra, 98
Sarikaya, Ruhi, 34, 48
Sarkar, Sandip, 117
Sarker, Abeed, 107
Sartiano, Daniele, 107
Sassano, Manabu, 141
Sateli, Bahar, 128
Sato, Misa, 108
Savenkov, Denis, 119
Savova, Guergana, 115
Sayeed, Asad, 78
Saygin, Yücel, 105
Scarton, Carolina, 111
Schenk, Niko, 98
Scherbakov, Andreas, 113
Schlichtkrull, Michael, 117
Schluter, Natalie, 114
Schmaltz, Allen, 128
Schneider, Nathan, 110
Schofield, Alexandra, 121
Schouten, Kim, 109
Schulder, Marc, 61
Schuler, William, 116
Schüller, Peter, 105, 112
Schulte im Walde, Sabine, 39
Schultz, Robert, 122
Schütze, Hinrich, 44, 55, 61, 62, 92, 119
Schwartz, H. Andrew, 125
Schwartz, Roy, 43
Sebastiani, Fabrizio, 105, 126

- Sekulic, Ivan, 110
Semwal, Deepali, 79
Sennrich, Rico, 24
Seva, Jurica, 111
Sha, Lei, 41
Shah, Kashif, 45, 111
Shang, Lifeng, 119
Shao, Yanqiu, 115
Shapiro, Ari, 47
Shardlow, Matthew, 111
Sharma, Arnav, 90
Sharma, Dipti Misra, 80, 81, 90
Sharma, Vasu, 49
Shek, Elinna, 119
Shen, Tianxiao, 73
Shenoy, P Deepa, 126
Sheth, Amit, 61
Shi, Yangyang, 40, 99
Shickel, Benjamin, 122, 123
Schieber, Stuart M., 70, 128, 131
Shimaoka, Sonse, 141
Shrestha, Prasha, 123
Shrivastava, Manish, 43, 90
Shulman, Elisabeth, 125
Shutova, Ekaterina, 31, 137
Si, Mei, 79
Sidarenka, Uladzimir, 107
Siddique, Farhad Bin, 80
Sidorov, Grigori, 106
Silva, Mario J., 107
Simard, Michel, 117
Simha, Rahul, 127
Simo-Serra, Edgar, 45
Simov, Kiril, 130
Singh, Sameer, 80, 140
Singh, Satinder, 122
Singla, Parag, 26
Sirts, Kairit, 42
Sitaram, Sunayana, 92
Sivakumar, Prakash, 126
Smith, David, 42, 74
Smith, Noah A., 33, 59, 113, 116
Smola, Alex, 99
Snijders, Liselotte, 135
Sobhani, Parinaz, 67, 105
Socher, Richard, 125
Soderland, Stephen, 66
Søgaard, Anders, 74, 101
Solorio, Thamar, 112, 123
Soma, Christina S., 122
Song, Hyun-Je, 66
Song, Rou, 130
Song, Zhiyi, 136
Soni, Ameet, 141
Soong, Frank K., 44
Soraluze, Ander, 132
Specia, Lucia, 41, 45, 110, 111, 113
Sperber, Matthias, 79
Spiliopoulou, Myra, 105
Srikumar, Vivek, 122
Srivastava, Manish, 80
Stede, Manfred, 131
Steedman, Mark, 42, 53, 72
Stefanowski, Jerzy, 106
Steimel, Kenneth, 109
Stein, Benno, 96
Steinberger, David, 112
Steinberger, Josef, 109
Stenetorp, Pontus, 141
Stent, Amanda, 26
Stojanovski, Dario, 106
Stoyanov, Veselin, 105
Strapparava, Carlo, 126
Strassel, Stephanie, 136
Strezoski, Gjorgji, 106
Stromberg, Logan, 48
Strube, Michael, 70, 96, 128
Strubell, Emma, 66
Stüker, Sebastian, 79
Su, Jian, 108
Su, Keh-Yih, 79
Su, Pei-Hao, 28, 29
Subramanian, Sandeep, 34
Subramanya, Amarnag, 90
Suchanek, Fabian, 140
Sui, Zhifang, 41
Sulger, Sebastian, 135
Sultan, Md Arafat, 67, 72, 88, 111
Sumita, Eiichiro, 41, 73
Summers, Kristen, 119
Sumner, Tamara, 67, 72, 111
Sun, Le, 111
Sundar Ram, Vijay, 132
Suzuki, Jun, 75
Svoboda, Lukáš, 108, 111
Sysoev, Ivan, 109
Szpakowicz, Stan, 121
Szpektor, Idan, 56
Täckström, Oscar, 53
Tahmoush, David, 136
Tal, Alex, 121
Tamchyna, Aleš, 109
Tan, Liling, 47, 75, 111, 113, 114, 117
Tanana, Michael, 122
Tandon, Niket, 141
Tanev, Hristo, 117
Tang, Xin, 113
Tannier, Xavier, 105, 115
Tao, Fangbo, 79
Tayal, Devendra K., 137
Tayyar Madabushi, Harish, 117

- Teferra, Solomon, 49
Teixeira, Joana, 116
Tekumalla, Lavanya, 112
Temnikova, Irina, 73
Teng, Zhiyang, 41
Tetreault, Joel, 40, 72, 127
Theodorakakos, Panagiotis, 108
Thomas, Titus, 122
Thomson, Blaise, 29
Tian, Junfeng, 111
Tian, Le, 99
Tian, Ran, 63
Titov, Ivan, 92
Toda, Tomoki, 53
Toh, Zhiqiang, 108
Toldova, Svetlana, 132
Tomeh, Nadi, 73
Torki, Marwan, 112
Tourille, Julien, 115
Tran, Ke, 38
Tran, Quan Hung, 59
Trancoso, Isabel, 107
Trandabat, Diana, 107
Traum, David, 47
Tsai, Chen-Tse, 46, 62
Tseng, Yuen-Hsien, 128
Tsvetkov, Yulia, 55, 92, 138
Tu, Zhaopeng, 70
Turchi, Marco, 110
Turtle, Howard R., 81
Tutek, Martin, 110
Tymoshenko, Kateryna, 88, 113
- Uang, Justin, 140, 141
Uitdenbogerd, Alexandra, 113
Ungar, Lyle, 122, 125
Upadhyay, Shyam, 136
Urena Lopez, L. Alfonso, 126
Urešová, Zdeňka, 135
Uria, Larraitz, 110
Utiyama, Masao, 41
Uva, Antonio, 113
Uzdilli, Fatih, 106
- Šnajder, Jan, 110
Šuster, Simon, 92
Vanderwende, Lucy, 62, 77, 136
Vandyke, David, 28, 29
Varma, Vasudeva, 81
Vasilyeva, Maria, 132
Vaswani, Ashish, 33, 34, 77
Veale, Tony, 126, 137
Vechtomova, Olga, 109
Velldal, Erik, 126
Velupillai, Sunithra, 116, 122, 123
Verga, Patrick, 66, 141
- Verhagen, Marc, 115
Verschuere, Bruno, 134
Versley, Yannick, 113, 115, 135
Verspoor, Karin, 111
Veselovská, Kateřina, 109
Vieira, Tim, 55
Vijayaraghavan, Prashanth, 109
Vilares, David, 106
Vilaríño, Darnes, 106
Villasenor, Luis, 126
Villavicencio, Aline, 113
Vinyals, Oriol, 90
Viswanathan, Dileep, 141
Vlachos, Andreas, 75, 109, 115
Vogel, Stephan, 73
Volkova, Svitlana, 134
Volodina, Elena, 127
Vosoughi, Soroush, 109
Voss, Clare, 79
Vossen, Piek, 125
Vu, Ngoc Thang, 44
Vu, Thuy, 78
Vyas, Yogarshi, 76, 115
Vydiswaran, V.G. Vinod, 142
Vylomova, Ekaterina, 65, 113
- Wachsmuth, Henning, 96
Waibel, Alex, 79
Waite, Aurelien, 65
Walczak, Wojciech, 111
Walgenbach, Peter, 126
Walker, Marilyn A., 77, 109
Wallace, Byron C., 66, 100
Wallace, William, 79
Walter, Matthew R., 59
Wan, Ada, 138
Wan, Stephen, 123
Wan, Yan, 80
Wang, Chen-Kai, 123
Wang, Chuan, 115
Wang, Daisy Zhe, 76
Wang, Dongang, 79
Wang, Feixiang, 110
Wang, Houfeng, 130
Wang, Hujie, 112
Wang, Jingqi, 116
Wang, Jin, 44
Wang, Longyue, 70
Wang, Lu, 26, 59
Wang, Peilu, 44
Wang, Richard, 140
Wang, Shijin, 69
Wang, Shuohang, 98
Wang, Tengjiao, 109
Wang, Tianlu, 34
Wang, Wei, 140, 141

- Wang, William Yang, 9, 16, 26
Wang, Yanshan, 117
Wang, Yufei, 123
Wani, Nehal J, 80
Waseem, Zeerak, 81
Washington, Jonathan, 81
Watrin, Patrick, 116
Way, Andy, 70, 130
Wei, Wan, 109
Weihe, Karsten, 126, 142
Weikum, Gerhard, 141
Weissenbacher, Davy, 76
Weitzner, Scott, 119
Welbl, Johannes, 141
Weld, Daniel S., 66
Wen, Tsung-Hsien, 28, 29
Wester, Aksel, 126
Wiebe, Janyce, 125
Wiegand, Michael, 61
Wieting, John, 111
Wijaya, Derry Tanti, 62
Windisch Brown, Susan, 136
Wiseman, Sam, 70, 131
Wisniewski, Guillaume, 72, 138
Witte, René, 128
Wojatzki, Michael, 109, 128
Wojtulewicz, Laura, 76
Wood, Anthony, 123
Wróbel, Krzysztof, 113
Wu, Dekai, 138
Wu, Guoshun, 113
Wu, Hao, 117
Wu, Yonghui, 116
Wu, Yunfang, 112
Wubben, Sander, 41
- Xenos, Dionysios, 108
Xia, Menglin, 127
Xiang, Bing, 44
Xiong, Caiming, 31
Xiong, Deyi, 130
Xiong, Wenting, 46
Xu, Hua, 116
Xu, Jian, 71
Xu, Jun, 116
Xu, Steven, 107
Xu, Wenduan, 33
Xu, Ying, 120
Xu, Yipei, 48
Xu, Yong, 47
Xue, Nianwen, 115
- Yadav, Vikrant, 106
Yamakawa, Yukari, 136
Yan, Jinghui, 81
Yanai, Kohsuke, 108
- Yanase, Toshihiko, 108
Yang, Bishan, 37
Yang, Chin-Sheng, 107
Yang, Diyi, 99
Yang, Eunsuk, 34
Yang, Fan, 34
Yang, Shaohua, 31
Yang, Wei, 81
Yang, Yang, 80
Yang, Yi, 90
Yang, Zichao, 99
Yannakoudakis, Helen, 127
Yao, Kaisheng, 40, 65, 99
Yatskar, Mark, 32
Yermakovich, Maria, 116
Yih, Wen-tau, 9, 13
Yıldırım, Ezgi, 108
Yilmaz, Emine, 46
Yin, Jun, 119
Yin, Pengcheng, 119
Yin, Wenpeng, 119
Yoon, Hee-Geun, 66
Yoon, Su-Youn, 127, 128
Yoshida, Davis, 119
Young, Steve, 28, 29
Young, Sydney, 123
Yovcheva, Ivana, 112
Yu, Dong, 40, 48
Yu, Hong, 42
Yu, Liang-Chih, 44, 107, 128
Yu, Mo, 71, 78
Yu, Qi, 122
Yuan, Zheng, 40, 128
Yuret, Deniz, 59
Yvon, François, 47, 72, 138
- Zampieri, Marcos, 114, 123
Zarrella, Guido, 110
Zawiławska, Magdalena, 131
Zeldes, Amir, 46, 132
Zesch, Torsten, 105, 109, 127, 128
Zettlemoyer, Luke, 33
Zhang, Boliang, 34
Zhang, Chen, 105
Zhang, Fan, 47, 96
Zhang, Justine, 28
Zhang, Liangang, 71
Zhang, Minghua, 112
Zhang, Renxian, 49
Zhang, Rui, 99
Zhang, Shuo, 132
Zhang, Tongtao, 79
Zhang, Xiao, 109
Zhang, Xiaojun, 70
Zhang, Xi, 106
Zhang, Xingxing, 38

- Zhang, Xuejie, 44
Zhang, Yaoyun, 116
Zhang, Ye, 100
Zhang, Yuan, 90
Zhang, Yue, 41, 43
Zhang, Yu, 112
Zhang, Zhengchen, 105
Zhang, Zhihua, 107, 109, 110
Zhao, Hai, 44
Zhao, Yanyan, 105
Zhou, Bowen, 44
Zhou, Guangyu, 126
Zhou, Mengfei, 115
Zhou, Yunxiao, 107
Zhu, Song-Chun, 31
Zhu, Xiaodan, 67, 105
Ziering, Patrick, 55
Zirkly, Ayah, 123
Zitnick, C. Lawrence, 77
Zock, Michael, 49
Zoph, Barret, 24, 77
Zukerman, Ingrid, 59
-



YOUR SEARCH ENDS HERE



全球最大的中文搜索引擎

Baidu was founded in 2000 by Internet pioneer Robin Li, with the mission of providing the best way and most equitable way for people to find what they're looking for.



HELP CONNECT CUSTOMERS TO THE WORLD AT EBAY SDR

Proud Sponsor of NAACL, ACL, & EMNLP 2016



**Machine Translation
Natural Language Processing
Computer Vision
Machine Learning**



WE ARE HIRING!

With hundreds of millions of buyers and sellers, and hundreds of millions of live listings at any giving time, eBay Structured Data Research provides an amazing playground for researchers and practitioners to work with data and conduct experiments at massive scale.

We are research scientists, and computer engineers with a passion to change the world of Commerce. Join us as we invent the future of Commerce!

Visit: research.ebay.com

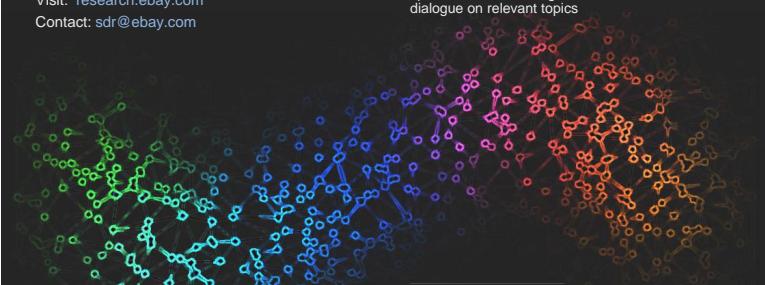
Contact: sdr@ebay.com

OTHER OPPORTUNITIES

Research Internship – Research internships in summer and/or through the year, in the US and Germany

Sponsorship Program – Partner with research institutions and universities for collaborative research that is mutually beneficial and help deepen our understanding of the research discipline and its applicability to business

eBay Speaker Series – Invite distinguished speakers and thought leaders and facilitate dialogue on relevant topics



Since the 1950s when Arthur Lee Samuel developed a checkers player capable of learning from experience, IBM has been a pioneer in using artificial intelligence to play and win against top humans in games of intellect and strategy. Twenty years later IBM defined how people think about speech recognition to this day. Another twenty years and IBM Research made history when chess-playing machine Deep Blue beat Garry Kasparov and became the first chess-playing machine to defeat a reigning world champion. IBM researchers continue to take on new AI challenges in the areas of machine learning, neuroscience, genomics and robotics.

1. 1949: [Checkers Player](#)
2. 1954: [Mark I and Mark II Automatic Language Translators](#)
3. 1972: [Probabilistic Speech Recognition](#)
4. 1984: [First Real Time Demonstration of a Large Vocabulary Speech Recognition Dictation System](#)
5. 1985: [Limited Reasoning](#)
6. 1988: [Statistical Translation Between Languages](#)
7. 1992: [TD-Gammon](#) (Computer Backgammon)
8. 1995: [Reasoning about Knowledge](#)
9. 1997: [Deep Blue Chess](#)
10. 2002: [BLEU: Bilingual Evaluation Understudy](#)
11. 2009: [UIMA: Unstructured Information Management](#)
12. 2009: [More Statistical Machine Translation](#)
13. 2011: [Watson Jeopardy Challenge](#)

WE'RE HIRING!

HELP MAKE VOICEBOX THE INTERFACE
TO THE INTERNET OF THINGS

Natural, Intelligent, Personal

Contact: Phil Cohen
philipc@voicebox.com

www.voicebox.com

in Follow Us @VoiceBox_Tech



From the everyday to the exceptional,
we work to help you achieve more.

The Microsoft Research team pushes technological boundaries to empower people and organizations around the globe.

Learn more at research.microsoft.com

BUILD SOMETHING
THAT TRULY MATTERS



Built by linguists and language lovers, Grammarly's writing app finds and corrects hundreds of complex writing errors — so you don't have to. Millions of users worldwide trust our products.

We're leveraging today's best technology and our deep expertise to find new ways to make health care better and more affordable. We're Optum360, a part of the growing family of UnitedHealth Group. Together, we will leverage our compassion, talent, resources and experience using natural language processing to create the next revolution in health care. Let's go. It's time to do **your life's best work.SM**

We are hiring for the following positions in our San Diego office:

- Research Linguist
- NLP Engineer
- NLP Software Engineer

You will be a key contributor in the design and development of state-of-the-art NLP for medical applications. You will work closely with other experts on current and forward-looking company technologies.

Interested in learning more?

Meet us at the Optum360 booth or visit www.optum360.com.



The Noah's Ark Lab is a research lab of **Huawei Technologies**, located in Hong Kong and Shenzhen. The mission of the lab is to make significant contributions to both the company and society by innovating in data mining, artificial intelligence, and related fields. Founded in 2012, the lab has now grown to be a research organization with many significant achievements. We welcome talented researchers and engineers to join us to realize their dreams.



Join us in building a data-driven world with services that deliver and software that empowers.
We're hiring!

CIVISANALYTICS.COM/CAREERS



research.nuance.com

Inventing a more human conversation with technology

To learn more about Nuance's R&D team of talented scientists, linguists and engineers pioneering human-machine intelligence and communication, visit [research.nuance.com](#)

 NUANCE

NEWSELA

Unlock the written word
for everyone



Daily news on the topics students care about, combined with the assessments, analyses, and insights today's teachers need to guide their students toward reading mastery.

[newsela.com](#)

YAHOO!

We are hiring
research scientists and engineers!

[careers.yahoo.com](#)



Sunnyvale, CA • New York, NY • London, England • Haifa, Israel

 Our areas of research: [research.yahoo.com](#)



NAACL gratefully acknowledges the following sponsors for their support:
Platinum



Gold



Silver



Bronze



Sponsors

Supporter

