2017

55th Annual Meeting ‖
Association for Computational Linguistics ‖
Vancouver, Canada ‖

ACL CONFERENCE HANDBOOK

July 30 - August 4

Cover design by Sharan Freundschuh Atak
Handbook assembled by Christian Federmann (in memoriam Hans-Ulrich Krieger)
Printing by Omnipress of Madison, Wisconsin

Contents

1 Conference Information

Message from the General Chair

Welcome to ACL 2017 on Vancouver! We hope you enjoy the conference and your stay in Canada. This is the 55th annual meeting of the Association for Computational Linguistics. We are anticipating one of the largest ACL conferences ever. We had a record number of papers submitted to the conference, and a record number of industry partners joining us as sponsors of the conference. We are on track to be one of the best attended ACL conferences to date. I hope that this year's conference is intellectually stimulating and that you take home many new ideas and techniques that will help extend your own research.

Each year, the ACL conference is organized by a dedicated team of volunteers. Please thank this year's organizers for their service to the community when you see them at the conference. Without these people, this conference would not happen: Regina Barzilay and Min-Yen Kan (Program Co-Chairs), Priscilla Rasmussen and Anoop Sarkar (Local Organizing Committee), Wei Xu and Jonathan Berant (Workshop Chairs), Maja Popović and Jordan Boyd-Graber (Tutorial Chairs) Wei Lu, Sameer Singh and Margaret Mitchell (Publication Chairs), Heng Ji and Mohit Bansal (Demonstration Chairs), Spandana Gella, Allyson Ettinger, and Matthieu Labeau (Student Research Workshop Organizers), Cecilia Ovesdotter Alm, Mark Dredze, and Marine Carpuat (Faculty Advisors to the Student Research Workshop), Charley Chan (Publicity Chair), Christian Federmann (Conference Handbook Chair), Maryam Siahbani (Student Volunteer Coordinator), and Nitin Madnani (Webmaster and Appmaster).

The organizers have been working for more than a year to put together the conference. Far more than a year in advance, the ACL 2017 Coordinating Committee helped to select the venue and to pick the General Chair and the Program Co-Chairs. This consisted of members from NAACL and ACL executive boards. Representing NAACL we had Hal Daumé III, Michael White, Joel Tetreault, and Emily M. Bender. Representing ACL we had Pushpak Bhattacharyya, Dragomir Radev, Graeme Hirst, Yejin Choi, and Priscilla Rasmussen. I would like to extend a personal thanks to Graeme and Priscilla who often serve as the ACL's institutional memory, and who have helped fill in many details along the way.

I would like to extend a special thanks to our Program Co-Chairs, Regina Barzilay and Min-Yen Kan. They documented their work creating the program by running a blog. They used their blog as a platform for engaging the ACL community in many of the decision making processes including soliciting suggestions for the conference's area chairs and invited speakers. They hosted discussions with Marti Hearst and Joakim Nivre about the value of publishing preprints of submitted paper on arXiv and how they relate to double blind reviewing. They even invited several prominent members of our community to provide last-minute writing advice. If you weren't following the blog in the lead-up to the conference, I highly recommend taking a look through it now. You can find it linked from the ACL 2017 web page.

This year's program looks like it will be excellent! We owe a huge thank you to Regina Barzilay and Min-Yen Kan. They selected this year's papers from 1,318 submissions with the help of 44 area chairs and more than 1,200 reviewers. Thanks to Regina, Min, the area chairs, the reviewers and the authors. Beyond the papers, we have talks by luminaries in the field of NLP, including ACL President Joakim Nivre, invited speakers Mirella Lapata and Noah Smith, and the recipient of this year's Lifetime Achievement Award. We also have an excellent set of workshops and tutorials. On the tutorial day, there will also be a special workshop on Women and Underrepresented Minorities in Natural Language Processing. Thank you to our workshop organizers and tutorial presenters.

This year we are trying to make ACL more family friendly. We are offering on-site subsidized childcare for the first time ever, plus a space for nursing or pumping mothers. Our social event is at the aquarium, which promises to be a family friendly venue. We hope you bring your loved ones with you.

I would like to thank our many sponsors for their generous contributions. Our platinum sponsors are Alibaba, Amazon, Apple, Baidu, Bloomberg, Facebook, Google, Samsung and Tencent. Our gold sponsors are eBay, Elsevier, IBM Research, KPMG, Maluuba, Microsoft, Naver Line, NEC, Recruit Institute of Technology, and SAP. Our silver sponsors are Adobe, Bosch, CVTE, Duolingo, Huawei, Nuance, Oracle, and Sogou. Our bronze sponsors are Grammarly, Toutiao, and Yandex. Our supporters include Newsela and four professional master's degree programs from Brandeis, Columbia, NYU and the University of Washington. We would like to acknowledge the generous support of the National Science Foundation which has awarded a $15,000 grant to the ACL Student Research Workshop. Finally, NVIDIA donated several Titan X GPU cards for us to raffle off during the conference.

Lastly, I would like to thank everyone else who helped to make this conference a success. Thank you to our area chairs, our army of reviewers, our workshop organizers, our tutorial presenters, our invited speakers, and our authors. Best regards to all of you.

Enjoy the conference! I hope that you learn something new that inspires you to push your own research to new heights. Welcome to ACL 2017!

Best Regards,
Chris Callison-Burch
ACL 2017 General Chair

Message from the Program Committee Co-Chairs

Welcome to the 55th Annual Meeting of the Association for Computational Linguistics! This year, ACL received 751 long paper submissions and 567 short paper submissions.[1] Of the long papers, 195 were accepted for presentation at ACL — 117 as oral presentations and 78 as poster presentations (25% acceptance rate). 107 short papers were accepted — 34 as oral and 73 as poster presentations (acceptance rate of 18%). In addition, ACL will also feature 21 presentations of papers accepted in the *Transactions of the Association for Computational Linguistics* (TACL). Including the student research workshop and software demonstrations, the ACL program swells to a massive total of 367 paper presentations on the scientific program, representing the largest ACL program to date.

ACL 2017 will have two distinguished invited speakers: Noah A. Smith (Associate Professor of Computer Science and Engineering at the University of Washington) and Mirella Lapata (Professor in the School of Informatics at the University of Edinburgh). Both are well-renowned for their contributions to the field of computational linguistics and are excellent orators. We are honored that they have accepted our invitation to address the membership at this exciting juncture in our field's history, addressing key issues in representation learning and multimodal machine translation.

To manage the tremendous growth of our field, we introduced some changes to the conference. With the rotation of the annual meeting to the Americas, we anticipated a heavy load of submissions and early on we decided to have both the long and short paper deadlines merged to reduce reviewing load and to force authors to take a stand on their submissions' format. The joint deadline allowed us to only load our reviewers once, and also enabled us to have an extended period for more lengthy dialogue among authors, reviewers and area chairs.

In addition, oral presentations were shortened to fourteen (twelve) minutes for long (short) papers, plus time for questions. While this places a greater demand on speakers to be concise, we believe it is worth the effort, allowing far more work to be presented orally. We also took advantage of the many halls available and expanded the number of parallel talks to five during most of the conference sessions.

In keeping with changes introduced in the ACL community from last year, we continued the practice of recognizing outstanding papers at ACL. The 22 outstanding papers (15 long, 7 short, 1.6% of submissions) represent a broad spectrum of exciting contributions and have been specially placed on the final day of the main conference where the program is focused into two parallel sessions of these outstanding contributions. From these, a best paper and a best short paper those will be announced in the awards session on Wednesday afternoon.

We introduced the chairs' blog[2], where we strove to make the selection process of the internal workings of the scientific committee more transparent. We have publicly documented our calls for area chairs, reviewers and accepted papers selection process. Via the blog, we communicated several innovations in the conference organization workflow, of which we would call attention to two key ones here.

In the review process, we pioneered the use of the Toronto Paper Matching System, a topic model based approach to the assignment of reviewers to papers. We hope this decision will spur other program chairs to adopt the system, as increased coverage will better the reviewer/submission matching process, ultimately leading to a higher quality program.

For posterity, we also introduced the usage of hyperlinks in the bibliography reference sections of papers, and have worked with the ACL Anthology to ensure that digital object identifiers (DOIs) appear in the footer of each paper. These steps will help broaden the long-term impact of the work that our community has on the scientific world at large.

[1] These numbers exclude papers that were not reviewed due to formatting, anonymity, or double submission violations or that were withdrawn prior to review, which was unfortunately a substantial number.

[2] `https://chairs-blog.acl2017.org/`

There are many individuals we wish to thank for their contributions to ACL 2017, some multiple times:

- The 61 area chairs who volunteered for our extra duty. They recruited reviewers, led discussions on each paper, replied to authors' direct comments to them and carefully assessed each submission. Their input was instrumental in guiding the final decisions on papers and selecting the outstanding papers.
- Our full program committee of 1,200 hard-working individuals who reviewed the conference's 1,318 submissions (including secondary reviewers).
- TACL editors-in-chief Mark Johnson, Lillian Lee, and Kristina Toutanova, for coordinating with us on TACL presentations at ACL.
- Noah Smith and Katrin Erk, program co-chairs of ACL 2016 and Ani Nenkova and Owen Rambow, program co-chairs of NAACL 2016, who we consulted several times on short order for help and advice.
- Wei Lu and Sameer Singh, our well-organized publication chairs, with direction and oversight from publication chair mentor Meg Mitchell. Also, Christian Federmann who helped with the local handbook.
- The responsive team at Softconf led by Rich Gerber, who worked quickly to resolve problems and who strove to integrate the use of the Toronto Paper Matching System (TPMS) for our use.
- Priscilla Rasmussen and Anoop Sarkar and the local organization team, especially webmaster Nitin Madnani.
- Chris Callison-Burch, our general chair, who kept us coordinated with the rest of the ACL 2017 team and helped us free our time to concentrate on the key duty of organizing the scientific program.
- Key-Sun Choi, Jing Jiang, Graham Neubig, Emily Pitler, and Bonnie Webber who carefully reviewed papers under consideration for best paper recognition.
- Our senior correspondents for the blog, who contributed guest posts and advice for writing and reviewing: Waleed Ammar, Yoav Artzi, Tim Baldwin, Marco Baroni, Claire Cardie, Xavier Carreras, Hal Daumé, Kevin Duh, Chris Dyer, Marti Hearst, Mirella Lapata, Emily M. Bender, Aurélien Max, Kathy McKeown, Ray Mooney, Ani Nenkova, Joakim Nivre, Philip Resnik, and Joel Tetreault. Without them, the participation of the community through the productive comments, and without you the readership, our blog for disseminating information about the decision processes would not have been possible and a success.

We hope that you enjoy ACL 2017 in Vancouver!

ACL 2017 program co-chairs
Regina Barzilay, Massachusetts Institute of Technology
Min-Yen Kan, National University of Singapore

Organizing Committee

General Chair
Chris Callison-Burch, University of Pennsylvania

Program Committee Co-chairs
Regina Barzilay, Massachusetts Institute of Technology
Min-Yen Kan, National University of Singapore

Local Organizing Committee
Priscilla Rasmussen, ACL
Anoop Sarkar, Simon Fraser University

Workshop Co-chairs
Wei Xu, Ohio State University
Jonathan Berant, Tel Aviv University

Tutorial Co-chairs
Maja Popović, Humboldt-Universität zu Berlin
Jordan Boyd-Graber, University of Colorado, Boulder

Publication Co-chairs
Wei Lu, Singapore University of Technology and Design
Sameer Singh, University of California, Irvine
Margaret Mitchell, Google (Advisory)

Demonstration Chairs
Heng Ji, Rensselaer Polytechnic Institute
Mohit Bansal, University of North Carolina, Chapel Hill

Student Research Workshop Organizers
Spandana Gella, University of Edinburgh
Allyson Ettinger, University of Maryland, College Park
Matthieu Labeau, Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur (LIMSI)

Faculty Advisors to the Student Research Workshop
Cecilia Ovesdotter Alm, Rochester Institute of Technology
Mark Dredze, Johns Hopkins University
Marine Carpuat, University of Maryland, College Park

Publicity Chair
Charley Chan, Bloomberg

Conference Handbook Chair
Christian Federmann, Microsoft

Student Volunteer Coordinator
Maryam Siahbani, University of the Fraser Valley

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

Webmaster and Appmaster
Nitin Madnani, Educational Testing Service

Program Committee

Program Committee Co-chairs
Regina Barzilay, Massachusetts Institute of Technology
Min-Yen Kan, National University of Singapore

Area Chairs

Dialogue and Interactive Systems
- Ron Artstein
- Raquel Fernandez
- Oliver Lemon

Discourse and Pragmatics
- Yangfeng Ji
- Sujian Li
- Bonnie Webber

Summarization and Generation
- Wenjie Li
- Alexander M Rush
- Verena Rieser

Information Extraction and NLP Applications
- Mausam
- Eugene Agichtein
- Chia-Hui Chang
- Jing Jiang
- Sarvnaz Karimi
- Zornitsa Kozareva
- Kang Liu
- Tie-Yan Liu
- Alessandro Moschitti
- Smaranda Muresan

Multilingual
- Omri Abend
- Mona Diab

Sentiment Analysis and Opinion Mining
- Alexandra Balahur
- Lun-Wei Ku
- Saif M Mohammad

Vision, Robotics and Grounding
- Mohit Bansal
- Nate Kushman

Machine Learning
- Grzegorz Chrupała
- Amir Globerson
- Tommi Jaakkola
- Sujith Ravi
- William Yang Wang

Phonology, Morphology and Word Segmentation
- Jason Eisner
- Hinrich Schütze

Semantics
- Manaal Faruqui
- Hannaneh Hajishirzi

Anna Korhonen
Preslav Nakov
Mehroosh Sadrzadeh
Aline Villavicencio

Multidisciplinary
Karën Fort
Michael Piotrowski

Speech
Chiori Hori
Chia-ying Lee

Discourse and Pragmatics
Yangfeng Ji
Sujian Li
Bonnie Webber

Cognitive Modeling and Psycholinguistics
Roger Levy
Anders Søgaard

Machine Translation
Yang Liu
Minh-Thang Luong
Haitao Mi
Graham Neubig
Deyi Xiong

Social Media
Zhiyuan Liu
Shimei Pan
Svitlana Volkova

Biomedical
Aurélie Névéol
Karin Verspoor

Tagging, Chunking, Syntax and Parsing
Emily Pitler
Barbara Plank
Yue Zhang
Hai Zhao

Resources and Evaluation
Sophie Rosset
Wajdi Zaghouani

Meal Info

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

- A full buffet breakfast will be provided each day in the Bayshore Grand Foyer
- Mid-Morning breaks include coffee and tea in the Bayshore Grand Foyer
- Mid-Afternoon breaks include coffee, tea, soda, water, and snacks in the Bayshore Grand Foyer
- A full dinner buffer is provided during the poster sessions on Monday and Tuesday evenings in the Bayshore Grand Ballroom/Foyer

2

Tutorials: Sunday, July 30

Overview

7:30 – 18:00	**Registration**	*Bayshore Grand Foyer*
7:30 – 9:00	**Breakfast**	*Bayshore Grand Foyer*
9:00 – 12:30	**Morning Tutorials**	
	Natural Language Processing for Precision Medicine *Hoifung Poon, Chris Quirk, Kristina Toutanova, and Wen-tau Yih*	*Mackenzie*
	Multimodal Machine Learning *Louis-Philippe Morency and Tadas Baltrusaitis*	*Salon 1*
	Deep Learning for Semantic Composition *Xiaodan Zhu and Edward Grefenstette*	*Salons A/B*
10:30 – 11:00	**Coffee break**	
12:30 – 14:00	**Lunch break**	
14:00 – 17:30	**Afternoon Tutorials**	
	Deep Learning for Dialogue Systems *Yun-Nung Chen, Asli Celikyilmaz, and Dilek Hakkani-Tur*	*Salons A/B*
	Beyond Words: Deep Learning for Multiword Expressions and Collocations *Valia Kordoni*	*Salon 1*
	Making Better Use of the Crowd *Jennifer Wortman Vaughan*	*Mackenzie*
15:30 – 16:00	**Coffee break**	
18:00 – 21:00	**Welcome Reception**	*Bayshore Grand Ballroom*

Message from the Tutorial Co-Chairs

This section contains the abstracts of the ACL 2017 tutorials. This year we had a joint call-for-tutorials, coordinated with the EACL and EMNLP co-chairs (6 co-chairs in total). We received 26 submissions for the joint ACL/EACL/EMNLP call, and it was a difficult task to make a final selection. The six co-chairs applied the following criteria for evaluation: relevance to ACL community, quality of proposal, quality of instructor, estimate of attendance, relevance of area. The tutorials were then assigned to venues trying to respect proposers' preferences and to balance topics across venues. Nine tutorials had ACL as the preferred conference, from which one was rejected, two were redirected to EMNLP and the rest (six of them) was accepted. All six are organised as half-day tutorials.

We are very grateful to Alex Klementiev and Lucia Specia (EACL tutorial chairs), Nathan Schneider and Alexandra Birch (EMNLP tutorial chairs), Priscilla Rasmussen and Anoop Sarkar (local chairs), Wei Lu, Sameer Singh and Margaret Mitchell (publication chairs), Min-Yen Kan and Regina Barzilay (program co-chairs) and of course Chris Callison-Burch (general chair) for various kinds of help, advice and assistance offered during the process of putting the tutorial programme and materials together. Most importantly, we would like to thank the tutorial presenters for the time and effort in preparing and presenting the tutorials.

We hope you will enjoy the tutorials!

ACL 2017 Tutorial Chairs
Maja Popović, Humboldt-Universität zu Berlin
Jordan Boyd-Graber, University of Colorado, Boulder

Tutorial 1

Natural Language Processing for Precision Medicine

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

Sunday, July 30, 2017, 9:00–12:30

Mackenzie

We will introduce precision medicine and showcase the vast opportunities for NLP in this burgeoning field with great societal impact. We will review pressing NLP problems, state-of-the art methods, and important applications, as well as datasets, medical resources, and practical issues. The tutorial will provide an accessible overview of biomedicine, and does not presume knowledge in biology or healthcare. The ultimate goal is to reduce the entry barrier for NLP researchers to contribute to this exciting domain.

Hoifung Poon is a Researcher at Microsoft Research Redmond. His research interests lie in advancing machine learning and natural language processing (NLP) to help automate discovery and decision support in precision medicine. He received his Ph.D. in computer science & engineering at the University of Washington. His past work has been recognized with Best Paper Awards from premier NLP and machine learning venues such as NAACL-09 (unsupervised morphological segmentation), EMNLP-09 (unsupervised semantic parsing), and UAI-11 (sum-product networks).

Chris Quirk is a Principal Researcher at Microsoft Research Redmond. Since joining Microsoft Research in 2001, his research has focused on effective computational systems for aiding human communication, understanding, and task completion. His primary focus is in machine translation, building practical and widely-used system implementations and authoring a number of influential papers. He has also worked in paraphrase, information extraction, and most recently biological applications of natural language processing and machine learning. He has served on numerous program committees, acted Area Chair (ACL 2010, EMNLP 2012), and is currently an action editor of the TACL journal.

Kristina Toutanova is a Staff Research Scientist at Google Research Seattle and affiliate faculty member at the University of Washington. In 2005, she obtained her Ph.D. from the Computer Science Department at Stanford University, where she was advised by Christopher Manning. She focuses on modeling the structure of natural language using machine learning, in the areas of semantic parsing, knowledge extraction, information retrieval, and text-to-text generation. She has coauthored more than 50 publications at refereed conferences and journals, including four papers that have won awards at conferences (EMNLP, NAACL, CoNLL, ECML). She served as a program co-chair for CoNLL 2008 and ACL 2014 and is currently serving as a co-editor-in-chief of the TACL journal.

Wen-tau Yih is a Senior Researcher at Microsoft Research Redmond. His research interests include natural language processing, machine learning and information retrieval. Yih received his Ph.D. in computer science at the University of Illinois at Urbana-Champaign. His work on joint inference using integer linear programming (ILP) helped the UIUC team win the CoNLL-05 shared task on semantic role labeling, and the approach has been widely adopted in the NLP community since then. After joining MSR in 2005, he has worked on email spam filtering, keyword extraction and search & ad relevance. His recent work focuses on continuous semantic representations using neural networks and matrix/tensor decomposition methods, with applications in lexical semantics, knowledge base embedding and question answering. Yih received

the best paper award from CoNLL-2011, an outstanding paper award from ACL-2015 and has served as area chairs (HLT-NAACL-12, ACL-14, EMNLP16,17), program co-chairs (CEAS-09, CoNLL-14) and action/associated editors (TACL, JAIR) in recent years.

Tutorial 2

Multimodal Machine Learning

Louis-Philippe Morency and Tadas Baltrusaitis

Sunday, July 30, 2017, 9:00–12:30

Salon 1

Multimodal machine learning is a vibrant multi-disciplinary research field which addresses some of the original goals of artificial intelligence by integrating and modeling multiple communicative modalities, including linguistic, acoustic and visual messages. With the initial research on audio-visual speech recognition and more recently with image and video captioning projects, this research field brings some unique challenges for multimodal researchers given the heterogeneity of the data and the contingency often found between modalities.

This tutorial builds upon a recent course taught at Carnegie Mellon University during the Spring 2016 semester (CMU course 11-777) and two tutorials presented at CVPR 2016 and ICMI 2016. The present tutorial will review fundamental concepts of machine learning and deep neural networks before describing the five main challenges in multimodal machine learning: (1) multimodal representation learning, (2) translation & mapping, (3) modality alignment, (4) multimodal fusion and (5) co-learning. The tutorial will also present state-of-the-art algorithms that were recently proposed to solve multimodal applications such as image captioning, video descriptions and visual question-answer. We will also discuss the current and upcoming challenges.

Louis-Philippe Morency (`https://www.cs.cmu.edu/~morency/`) is Assistant Professor in the Language Technology Institute at the Carnegie Mellon University where he leads the Multimodal Communication and Machine Learning Laboratory (MultiComp Lab). He received his Ph.D. and Master degrees from MIT Computer Science and Artificial Intelligence Laboratory. In 2008, Dr. Morency was selected as one of "AI's 10 to Watch" by IEEE Intelligent Systems. He has received 7 best paper awards in multiple ACMand IEEE-sponsored conferences for his work on context-based gesture recognition, multimodal probabilistic fusion and computational models of human communication dynamics. Dr. Morency was General Chair for the International Conference on Multimodal Interaction (ICMI 2012) and the NIPS 2010 workshop on Modeling Human Communication Dynamics. He was Program Chair for ICMI 2011, 2014 and 2016, as well as the Tenth International Conference on Creating, Connecting and Collaborating through Computing in January 2012.

Tadas Baltrusaitis (`http://www.cl.cam.ac.uk/~tb346/`) is a post-doctoral associate at the Language Technologies Institute, Carnegie Mellon University. Before this, he was a post-doctoral research at the University of Cambridge, where he also received his PhD degree in 2014. His primary research interests lie in the automatic understanding of non-verbal human behaviour, computer vision, and multimodal machine learning. His papers have won a number of awards for his work on non-verbal human behavior analysis, including ICMI 2014 best student paper award, and ETRA 2016 emerging investigator award. He is also a winner of several challenges in computer vision and multi-modal machine learning, including FERA 2015, and AVEC 2011.

Tutorial 3

Deep Learning for Semantic Composition

Xiaodan Zhu and Edward Grefenstette

Sunday, July 30, 2017, 9:00–12:30

Salons A/B

Learning representation to model the meaning of text has been a core problem in NLP. The last several years have seen extensive interests on distributional approaches, in which text spans of different granularities are encoded as vectors of numerical values. If properly learned, such representation has showed to achieve the state-of-the-art performance on a wide range of NLP problems. In this tutorial, we will cover the fundamentals and the state-of-the-art research on neural network-based modeling for semantic composition, which aims to learn distributed representation for different granularities of text, e.g., phrases, sentences, or even documents, from their sub-component meaning representation, e.g., word embedding.

Xiaodan Zhu is an Assistant Professor of the Department of Electrical and Computer Engineering of Queen's University, Canada. Before that, he was a Research Officer of the National Research Council Canada. His research interests are in Natural Language Processing and Machine Learning. His recent work has focused on deep learning, semantic composition, sentiment analysis, and natural language inference.

Edward Grefenstette is a Staff Research Scientist at DeepMind. His research covers the intersection of Machine Learning, Computer Reasoning, and Natural Language Understanding. Recent publications cover the topics of neural computation, representation learning at the sentence level, recognising textual entailment, and machine reading.

Tutorial 4

Deep Learning for Dialogue Systems

Yun-Nung Chen, Asli Celikyilmaz, and Dilek Hakkani-Tur

Sunday, July 30, 2017, 14:00–17:30

Salons A/B

With the rising trend of artificial intelligence, more and more devices have incorporated goal-oriented spoken dialogue systems. Among popular virtual personal assistants, Microsoftís Cortana, Appleís Siri, Amazon Alexa, Google Assistant, and Facebookís M, have incorporated dialogue system modules in various devices, which allow users to speak naturally in order to finish tasks more efficiently.

The traditional conversational systems have rather complex and/or modular pipelines. The advance of deep learning technologies has recently risen the applications of neural models to dialogue modeling. Nevertheless, applying deep learning technologies for building robust and scalable dialogue systems is still a challenging task and an open research area as it requires deeper understanding of the classic pipelines as well as detailed knowledge on the benchmark of the models of the prior work and the recent state-of-the-art work.

The goal of this tutorial is to provide the audience with developing trend of the dialogue systems, and a roadmap to get them started with the related work. In the first section of the tutorial, we motivate the work on conversation-based intelligent agents, in which the core underlying system is task-oriented dialogue systems. The second and third sections describe different approaches using deep learning for each component in the dialogue system and how it is evaluated. The last two sections focus on discussing the recent trends and current challenges on dialogue system technology and summarize the challenges and conclusions. The tutorial material is available at `http://deepdialogue.miulab.tw/`.

Yun-Nung Chen is currently an assistant professor at the Department of Computer Science, National Taiwan University. She earned her Ph.D. degree from Carnegie Mellon University, where her research interests focus on spoken dialogue system, language understanding, natural language processing, and multi-modal speech applications. She received the Google Faculty Research Awards 2016, two Student Best Paper Awards from IEEE SLT 2010 and IEEE ASRU 2013, a Student Best Paper Nominee from Interspeech 2012, and the Distinguished Master Thesis Award from ACLCLP. Before joining National Taiwan University, she worked in the Deep Learning Technology Center at Microsoft Research Redmond. More information about her can be found at `http://vivianchen.idv.tw`.

Asli Celikyilmaz is currently a researcher at the Deep Learning Technology Center at Microsfot Research. Previously, she was a Research Scientist at Microsoft Bing from 2010 until 2016 focusing on deep learning models for scaling natural user interfaces to multiple domains. She has worked as a Postdoc Researcher at the EECS Department of the UC Berkeley from 2008 until 2010. She has worked with researchers at ICSI @ Berkeley during her postdoc research study. She earned her Ph.D. from University of Toronto, Canada in 2008. Asli's research interests are mainly machine learning and its applications to conversational dialogue systems, mainly natural language understanding and dialogue modeling. In the past she has also focused on research areas including machine intelligence, semantic tagging of natural user utterances of human to machine conversations, text analysis, document summarization, question answering, co-reference resolution, to name a few. Currently she is focusing on reasoning, attention, memory networks as well

as multi-task and transfer learning for conversational dialogue systems. She has been serving as area chair, co-organizer of numerous NLP and speech conferences, such as ACL, NAACL, Interspeech, and IEEE Spoken Language Technologies (SLT). She co-organized a NIPS workshop on Machine Learning for Spoken Language Understanding and Interactions in 2015.

Dilek Hakkani-Tur is a research scientist at Google Research. Prior to joining Google, she was a researcher at Microsoft Research (2010-2016), International Computer Science Institute (ICSI, 2006-2010) and AT&T Labs-Research (20012005). She received her BSc degree from Middle East Technical Univ, in 1994, and MSc and PhD degrees from Bilkent Univ., Department of Computer Engineering, in 1996 and 2000, respectively. Her research interests include natural language and speech processing, spoken dialogue systems, and machine learning for language processing. She has over 50 patents that were granted and co-authored more than 200 papers in natural language and speech processing. She is the recipient of three best paper awards for her work on active learning for dialogue systems, from IEEE Signal Processing Society, ISCA and EURASIP. She was an associate editor of IEEE Transactions on Audio, Speech and Language Processing (20052008), member of the IEEE Speech and Language Technical Committee (2009-2014), area editor for speech and language processing for Elsevier's Digital Signal Processing Journal and IEEE Signal Processing Letters (2011-2013), and currently serves on ISCA Advisory Council (20152018). She is a fellow of IEEE and ISCA.

Tutorial 5

Beyond Words: Deep Learning for Multiword Expressions and Collocations

Valia Kordoni

Sunday, July 30, 2017, 14:00–17:30

Salon 1

Deep learning has recently shown much promise for NLP applications. Traditionally, in most NLP approaches, documents or sentences are represented by a sparse bag-of-words representation. There is now a lot of work which goes beyond this by adopting a distributed representation of words, by constructing a so-called "neural embedding" or vector space representation of each word or document. The aim of this tutorial is to go beyond the learning of word vectors and present methods for learning vector representations for Multiword Expressions and bilingual phrase pairs, all of which are useful for various NLP applications. This tutorial aims to provide attendees with a clear notion of the linguistic and distributional characteristics of Multiword Expressions (MWEs), their relevance for the intersection of deep learning and natural language processing, what methods and resources are available to support their use, and what more could be done in the future. Our target audience are researchers and practitioners in machine learning, parsing (syntactic and semantic) and language technology, not necessarily experts in MWEs, who are interested in tasks that involve or could benefit from considering MWEs as a pervasive phenomenon in human language and communication.

Valia Kordoni is a research professor of computational linguistics at Humboldt University Berlin. She is a leader in EU-funded research in Machine Translation, Computational Semantics, and Machine Learning. She has organized conferences and workshops dedicated to research on MWEs, recently including the EACL 2014 10th Workshop on Multiword Expressions (MWE 2014) in Gothenburg, Sweden, the NAACL 2015 11th Workshop on Multiword Expressions in Denver, Colorado, and the ACL 2016 12th Workshop on Multiword Expressions in Berlin, Germany, among others. She has been the Local Chair of ACL 2016 The 54th Annual Meeting of the Association for Computational Linguistics which took place at the Humboldt University Berlin in August 2016. She has taught a tutorial on Robust Automated Natural Language Processing with Multiword Expressions and Collocations in ACL 2013, as well as a tutorial on Robust Semantic Analysis of Multiword Expressions with FrameNet in EMNLP 2015, together with Miriam R. L. Petruck. She is also the author of Multiword Expressions From Linguistic Analysis to Language Technology Applications (to appear, Springer).

Tutorial 6

Making Better Use of the Crowd

Jennifer Wortman Vaughan

Sunday, July 30, 2017, 14:00–17:30

Mackenzie

Over the last decade, crowdsourcing has been used to harness the power of human computation to solve tasks that are notoriously difficult to solve with computers alone, such as determining whether or not an image contains a tree, rating the relevance of a website, or verifying the phone number of a business.

The natural language processing community was early to embrace crowdsourcing as a tool for quickly and inexpensively obtaining annotated data to train NLP systems. Once this data is collected, it can be handed off to algorithms that learn to perform basic NLP tasks such as translation or parsing.

Usually this handoff is where interaction with the crowd ends. The crowd provides the data, but the ultimate goal is to eventually take humans out of the loop. Are there better ways to make use of the crowd?

In this tutorial, I will begin with a showcase of innovative uses of crowdsourcing that go beyond data collection and annotation. I will discuss applications to natural language processing and machine learning, hybrid intelligence or "human in the loop" AI systems that leverage the complementary strengths of humans and machines in order to achieve more than either could achieve alone, and large scale studies of human behavior online.

I will then spend the majority of the tutorial diving into recent research aimed at understanding who crowdworkers are, how they behave, and what this should teach us about best practices for interacting with the crowd.

I'll start by debunking the common myth among researchers that crowdsourcing platforms are riddled with bad actors out to scam requesters. In particular, I'll describe the results of a research study that showed that crowdworkers on the whole are basically honest.

I'll talk about experiments that have explored how to boost the quality and quantity of crowdwork by appealing to both well-designed monetary incentives (such as performance-based payments) and intrinsic sources of motivation (such as piqued curiosity or a sense of doing meaningful work).

I'll then discuss recent research—both qualitative and quantitative—that has opened up the black box of crowdsourcing to uncover that crowdworkers are not independent contractors, but rather a network with a rich communication structure.

Taken as a whole, this research has a lot to teach us about how to most effectively interact with the crowd. Throughout the tutorial I'll discuss best practices for engaging with crowdworkers that are rarely mentioned in the literature but make a huge difference in whether or not your research studies will succeed. (Here's a few hints: Be respectful. Be responsive. Be clear.)

Jennifer Wortman Vaughan is a Senior Researcher at Microsoft Research, New York City, where she studies algorithmic economics, machine learning, and social computing, with a heavy focus on prediction markets and other forms of crowdsourcing. She is interested in developing general methods that allow us to reason formally about the performance of algorithms with

human components in the same way that traditional computer science techniques allow us to formally reason about algorithms that run on machines alone. Jenn came to Microsoft in 2012 from UCLA, where she was an assistant professor in the computer science department. She completed her Ph.D. at the University of Pennsylvania in 2009, and subsequently spent a year as a Computing Innovation Fellow at Harvard. She is the recipient of Penn's 2009 Rubinoff dissertation award for innovative applications of computer technology, a National Science Foundation CAREER award, a Presidential Early Career Award for Scientists and Engineers (PECASE), and a handful of best paper or best student paper awards. In her "spare" time, Jenn is involved in a variety of efforts to provide support for women in computer science; most notably, she co-founded the Annual Workshop for Women in Machine Learning, which has been held each year since 2006.

Welcome Reception

Sunday, July 30, 2017, 18:00 – 21:00

Westin Bayshore Hotel (Conference Venue)
Bayshore Grand Ballroom
`http://www.westinbayshore.com/`

Catch up with your colleagues at the **Welcome Reception**! It will be held immediately following the Tutorials on Sunday, July 30 at 18:00 in the Bayshore Grand Ballroom of the Westin Bayshore Hotel (the conference venue). Refreshments and a light dinner will be provided, and a cash bar will be available.

3

Main Conference: Monday, July 31

Overview

7:30 – 9:00 **Breakfast** *Bayshore Grand Foyer*

9:00 – 10:00 **Plenary Session. Welcome to ACL 2017!** *Bayshore Grand Ballroom*
Presidential Address: Joakim Nivre

10:00 – 10:30 **Coffee break** *Bayshore Grand Foyer*

Session 1

10:30 – 12:00					
	Information Extraction 1 (NN)	Semantics 1	Discourse 1	Machine Translation 1	Generation 1
	Salons B/C	*Salons E/F*	*Salon D*	*Salon 1*	*Salons 2/3*

12:00 – 13:40 **Lunch break**

Session 2

13:40 – 15:15					
	Question Answering 1	Vision 1	Syntax 1	Machine Learning 1 (NN)	Sentiment 1 (NN)
	Salons B/C	*Salons E/F*	*Salon D*	*Salon 1*	*Salons 2/3*

15:15 – 15:45 **Coffee break** *Bayshore Grand Foyer*

Session 3

15:45 – 17:20					
	Information Extraction 2 / Biomedical 1	Semantics 2 (NN)	Speech 1 / Dialogue 1	Multilingual 1	Phonology 1
	Salons B/C	*Salons E/F*	*Salon D*	*Salon 1*	*Salons 2/3*

18:00 – 21:30 **Poster Session P1 with dinner (includes SRW)**
Bayshore Grand Ballroom/Foyer

Presidential Address: Joakim Nivre

ACL 2017 Presidential Address

Monday, July 31, 2017, 9:00–10:00

Bayshore Grand Ballroom

Abstract: Computational linguistics is a booming field and our association is flourishing with it. As our conferences grow larger and the pace of publishing quickens, there is a constant need to reflect on strategies that will allow us to prosper and grow even stronger in the future. In my presidential address, I will focus on three topics that I think require our attention. The first is equity and diversity, where the ACL executive committee has recently launched a number of actions intended to improve the inclusiveness and diversity of our community, but where there is clearly a need to do more. The second topic is publishing and reviewing, where the landscape is changing very quickly and our current system is starting to strain under the sheer volume of submissions. In particular, there has been an active discussion recently about the pros and cons of preprint publishing and the way it interacts with our standard model for double-blind reviewing. On this topic, I will present the results of a large-scale survey organized by the ACL executive committee to learn more about current practices and views in our community, a survey that will be followed up by a panel and discussion at the ACL business meeting later in the week. The third and final topic is good science and what we can do to promote scientific methodology and research ethics, which is becoming increasingly important in a world where the role of science in society cannot be taken for granted.

Biography: Joakim Nivre is Professor of Computational Linguistics at Uppsala University. He holds a Ph.D. in General Linguistics from the University of Gothenburg and a Ph.D. in Computer Science from Växjö University. His research focuses on data-driven methods for natural language processing, in particular for syntactic and semantic analysis. He is one of the main developers of the transition-based approach to syntactic dependency parsing, described in his 2006 book Inductive Dependency Parsing and implemented in the widely used MaltParser system, and one of the founders of the Universal Dependencies project, which aims to develop cross-linguistically consistent treebank annotation for many languages and currently involves over 150 researchers around the world. He has produced over 200 scientific publications and has more than 11,000 citations according to Google Scholar (June, 2017). He is currently the president of the Association for Computational Linguistics.

Session 1 Overview – Monday, July 31, 2017

Track A	**Track B**	**Track C**	**Track D**	**Track E**	
Information Extraction 1 (NN)	*Semantics 1*	*Discourse 1*	*Machine Translation 1*	*Generation 1*	
Salons B/C	Salons E/F	Salon D	Salon 1	Salons 2/3	
Adversarial Multi-task Learning for Text Classification *P. Liu, X. Qiu, and X. Huang*	Learning Structured Natural Language Representations for Semantic Parsing *J. Cheng, S. Reddy, V. Saraswat, and M. Lapata*	Joint Learning for Event Coreference Resolution *J. Lu and V. Ng*	A Convolutional Encoder Model for Neural Machine Translation *J. Gehring, M. Auli, D. Grangier, and Y. Dauphin*	Neural AMR: Sequence-to-Sequence Models for Parsing and Generation *I. Konstas, S. Iyer, M. Yatskar, Y. Choi, and L. Zettlemoyer*	10:30
Neural End-to-End Learning for Computational Argumentation Mining *S. Eger, J. Daxenberger, and I. Gurevych*	Morph-fitting: Fine-Tuning Word Vector Spaces with Simple Language-Specific Rules *I. Vulić, N. Mrkšić, R. Reichart, D. Ó Séaghdha, S. Young, and A. Korhonen*	Generating and Exploiting Large-scale Pseudo Training Data for Zero Pronoun Resolution *T. Liu, Y. Cui, Q. Yin, W.-N. Zhang, S. Wang, and G. Hu*	Deep Neural Machine Translation with Linear Associative Unit *M. Wang, Z. Lu, J. Zhou, and Q. Liu*	Program Induction by Rationale Generation: Learning to Solve and Explain Algebraic Word Problems *W. Ling, D. Yogatama, C. Dyer, and P. Blunsom*	10:49
Neural Symbolic Machines: Learning Semantic Parsers on Freebase with Weak Supervision *C. Liang, J. Berant, Q. Le, K. D. Forbus, and N. Lao*	Skip-Gram - Zipf + Uniform = Vector Additivity *A. Gittens, D. Achlioptas, and M. W. Mahoney*	Discourse Mode Identification in Essays *W. Song, D. Wang, R. Fu, L. Liu, T. Liu, and G. Hu*	[TACL] A Polynomial-Time Dynamic Programming Algorithm for Phrase-Based Decoding with a Fixed Distortion Limit *Y.-W. Chang and M. Collins*	Automatically Generating Rhythmic Verse with Neural Networks *J. Hopkins and D. Kiela*	11:08
Neural Relation Extraction with Multi-lingual Attention *Y. Lin, Z. Liu, and M. Sun*	The State of the Art in Semantic Representation *O. Abend and A. Rappoport*	[TACL] Winning on the Merits: The Joint Effects of Content and Style on Debate Outcomes *L. Wang, N. Beauchamp, S. Shugars, and K. Qin*	[TACL] Context Gates for Neural Machine Translation *Z. Tu, Y. Liu, Z. Lu, X. Liu, and H. Li*	Creating Training Corpora for NLG Micro-Planners *C. Gardent, A. Shimorina, S. Narayan, and L. Perez-Beltrachini*	11:27
Classifying Temporal Relations by Bidirectional LSTM over Dependency Paths *F. Cheng and Y. Miyao*	AMR-to-text Generation with Synchronous Node Replacement Grammar *L. Song, X. Peng, Y. Zhang, Z. Wang, and D. Gildea*	Lexical Features in Coreference Resolution: To be Used With Caution *N. S. Moosavi and M. Strube*	Alternative Objective Functions for Training MT Evaluation Metrics *M. Stanojević and K. Sima'an*	A Principled Framework for Evaluating Summarizers: Comparing Models of Summary Quality against Human Judgments *M. Peyrard and J. Eckle-Kohler*	11:46

Parallel Session 1

Session Session 1A: Information Extraction 1 (NN)

Adversarial Multi-task Learning for Text Classification
Pengfei Liu, Xipeng Qiu, and Xuanjing Huang 10:30–10:48

Neural network models have shown their promising opportunities for multi-task learning, which focus on learning the shared layers to extract the common and task-invariant features. However, in most existing approaches, the extracted shared features are prone to be contaminated by task-specific features or the noise brought by other tasks. In this paper, we propose an adversarial multi-task learning framework, alleviating the shared and private latent feature spaces from interfering with each other. We conduct extensive experiments on 16 different text classification tasks, which demonstrates the benefits of our approach. Besides, we show that the shared knowledge learned by our proposed model can be regarded as off-the-shelf knowledge and easily transferred to new tasks. The datasets of all 16 tasks are publicly available at `http://nlp.fudan.edu.cn/data/`

Neural End-to-End Learning for Computational Argumentation Mining
Steffen Eger, Johannes Daxenberger, and Iryna Gurevych 10:49–11:07

We investigate neural techniques for end-to-end computational argumentation mining (AM). We frame AM both as a token-based dependency parsing and as a token-based sequence tagging problem, including a multi-task learning setup. Contrary to models that operate on the argument component level, we find that framing AM as dependency parsing leads to subpar performance results. In contrast, less complex (local) tagging models based on BiLSTMs perform robustly across classification scenarios, being able to catch long-range dependencies inherent to the AM problem. Moreover, we find that jointly learning 'natural' subtasks, in a multi-task learning setup, improves performance.

Neural Symbolic Machines: Learning Semantic Parsers on Freebase with Weak Supervision
Chen Liang, Jonathan Berant, Quoc Le, Kenneth D. Forbus, and Ni Lao 11:08–11:26

Harnessing the statistical power of neural networks to perform language understanding and symbolic reasoning is difficult, when it requires executing efficient discrete operations against a large knowledge-base. In this work, we introduce a Neural Symbolic Machine, which contains (a) a neural "programmer", i.e., a sequence-to-sequence model that maps language utterances to programs and utilizes a key-variable memory to handle compositionality (b) a symbolic "computer", i.e., a Lisp interpreter that performs program execution, and helps find good programs by pruning the search space. We apply REINFORCE to directly optimize the task reward of this structured prediction problem. To train with weak supervision and improve the stability of REINFORCE, we augment it with an iterative maximum-likelihood training process. NSM outperforms the state-of-the-art on the WebQuestionsSP dataset when trained from question-answer pairs only, without requiring any feature engineering or domain-specific knowledge.

Neural Relation Extraction with Multi-lingual Attention
Yankai Lin, Zhiyuan Liu, and Maosong Sun 11:27–11:45

Relation extraction has been widely used for finding unknown relational facts from plain text. Most existing methods focus on exploiting mono-lingual data for relation extraction, ignoring massive information from the texts in various languages. To address this issue, we introduce a multi-lingual neural relation extraction framework, which employs mono-lingual attention to utilize the information within mono-lingual texts and further proposes cross-lingual attention to consider the information consistency and complementarity among cross-lingual texts. Experimental results on real-world datasets show that, our model can take advantage of multi-lingual texts and consistently achieve significant improvements on relation extraction as compared with baselines.

Classifying Temporal Relations by Bidirectional LSTM over Dependency Paths
Fei Cheng and Yusuke Miyao 11:46–11:58

Temporal relation classification is becoming an active research field. Lots of methods have been proposed, while most of them focus on extracting features from external resources. Less attention has been paid to a significant advance in a closely related task: relation extraction. In this work, we borrow a state-of-the-art method in relation extraction by adopting bidirectional long short-term memory (Bi-LSTM) along dependency paths (DP). We make a "common root" assumption to extend DP representations of cross-sentence links. In the final comparison to two state-of-the-art systems on TimeBank-Dense, our model achieves comparable performance, without using external knowledge, as well as manually annotated attributes of entities (class,

tense, polarity, etc.).

Session Session 1B: Semantics 1

Learning Structured Natural Language Representations for Semantic Parsing
Jianpeng Cheng, Siva Reddy, Vijay Saraswat, and Mirella Lapata 10:30–10:48

We introduce a neural semantic parser which is interpretable and scalable. Our model converts natural language utterances to intermediate, domain-general natural language representations in the form of predicate-argument structures, which are induced with a transition system and subsequently mapped to target domains. The semantic parser is trained end-to-end using annotated logical forms or their denotations. We achieve the state of the art on SPADES and GRAPHQUESTIONS and obtain competitive results on GEOQUERY and WEBQUESTIONS. The induced predicate-argument structures shed light on the types of representations useful for semantic parsing and how these are dif- ferent from linguistically motivated ones.

Morph-fitting: Fine-Tuning Word Vector Spaces with Simple Language-Specific Rules
Ivan Vulić, Nikola Mrkšić, Roi Reichart, Diarmuid Ó Séaghdha, Steve Young, and Anna Korhonen 10:49–11:07

Morphologically rich languages accentuate two properties of distributional vector space models: 1) the difficulty of inducing accurate representations for low-frequency word forms; and 2) insensitivity to distinct lexical relations that have similar distributional signatures. These effects are detrimental for language understanding systems, which may infer that 'inexpensive' is a rephrasing for 'expensive' or may not associate 'acquire' with 'acquires'. In this work, we propose a novel morph-fitting procedure which moves past the use of curated semantic lexicons for improving distributional vector spaces. Instead, our method injects morphological constraints generated using simple language-specific rules, pulling inflectional forms of the same word close together and pushing derivational antonyms far apart. In intrinsic evaluation over four languages, we show that our approach: 1) improves low-frequency word estimates; and 2) boosts the semantic quality of the entire word vector collection. Finally, we show that morph-fitted vectors yield large gains in the downstream task of dialogue state tracking, highlighting the importance of morphology for tackling long-tail phenomena in language understanding tasks.

Skip-Gram - Zipf + Uniform = Vector Additivity
Alex Gittens, Dimitris Achlioptas, and Michael W. Mahoney 11:08–11:26

In recent years word-embedding models have gained great popularity due to their remarkable performance on several tasks, including word analogy questions and caption generation. An unexpected "side-effect" of such models is that their vectors often exhibit compositionality, i.e., *adding* two word-vectors results in a vector that is only a small angle away from the vector of a word representing the semantic composite of the original words, e.g., "man" + "royal" = "king". This work provides a theoretical justification for the presence of additive compositionality in word vectors learned using the Skip-Gram model. In particular, it shows that additive compositionality holds in an even stricter sense (small distance rather than small angle) under certain assumptions on the process generating the corpus. As a corollary, it explains the success of vector calculus in solving word analogies. When these assumptions do not hold, this work describes the correct nonlinear composition operator. Finally, this work establishes a connection between the Skip-Gram model and the Sufficient Dimensionality Reduction (SDR) framework of Globerson and Tishby: the parameters of SDR models can be obtained from those of Skip-Gram models simply by adding information on symbol frequencies. This shows that Skip-Gram embeddings are optimal in the sense of Globerson and Tishby and, further, implies that the heuristics commonly used to approximately fit Skip-Gram models can be used to fit SDR models.

The State of the Art in Semantic Representation
Omri Abend and Ari Rappoport 11:27–11:45

Semantic representation is receiving growing attention in NLP in the past few years, and many proposals for semantic schemes (e.g., AMR, UCCA, GMB, UDS) have been put forth. Yet, little has been done to assess the achievements and the shortcomings of these new contenders, compare them with syntactic schemes, and clarify the general goals of research on semantic representation. We address these gaps by critically surveying the state of the art in the field.

AMR-to-text Generation with Synchronous Node Replacement Grammar
Linfeng Song, Xiaochang Peng, Yue Zhang, Zhiguo Wang, and Daniel Gildea 11:46–11:58

This paper addresses the task of AMR-to-text generation by leveraging synchronous node replacement grammar. During training, graph-to-string rules are learned using a heuristic extraction algorithm. At test time, a graph transducer is applied to collapse input AMRs and generate output sentences. Evaluated on a standard benchmark, our method gives the state-of-the-art result.

Session Session 1C: Discourse 1

Joint Learning for Event Coreference Resolution
Jing Lu and Vincent Ng 10:30–10:48

While joint models have been developed for many NLP tasks, the vast majority of event coreference resolvers, including the top-performing resolvers competing in the recent TAC KBP 2016 Event Nugget Detection and Coreference task, are pipeline-based, where the propagation of errors from the trigger detection component to the event coreference component is a major performance limiting factor. To address this problem, we propose a model for jointly learning event coreference, trigger detection, and event anaphoricity. Our joint model is novel in its choice of tasks and its features for capturing cross-task interactions. To our knowledge, this is the first attempt to train a mention-ranking model and employ event anaphoricity for event coreference. Our model achieves the best results to date on the KBP 2016 English and Chinese datasets.

Generating and Exploiting Large-scale Pseudo Training Data for Zero Pronoun Resolution
Ting Liu, Yiming Cui, Qingyu Yin, Wei-Nan Zhang, Shijin Wang, and Guoping Hu 10:49–11:07

Most existing approaches for zero pronoun resolution are heavily relying on annotated data, which is often released by shared task organizers. Therefore, the lack of annotated data becomes a major obstacle in the progress of zero pronoun resolution task. Also, it is expensive to spend manpower on labeling the data for better performance. To alleviate the problem above, in this paper, we propose a simple but novel approach to automatically generate large-scale pseudo training data for zero pronoun resolution. Furthermore, we successfully transfer the cloze-style reading comprehension neural network model into zero pronoun resolution task and propose a two-step training mechanism to overcome the gap between the pseudo training data and the real one. Experimental results show that the proposed approach significantly outperforms the state-of-the-art systems with an absolute improvements of 3.1% F-score on OntoNotes 5.0 data.

Discourse Mode Identification in Essays
Wei Song, Dong Wang, Ruiji Fu, Lizhen Liu, Ting Liu, and Guoping Hu 11:08–11:26

Discourse modes play an important role in writing composition and evaluation. This paper presents a study on the manual and automatic identification of narration,exposition, description, argument and emotion expressing sentences in narrative essays. We annotate a corpus to study the characteristics of discourse modes and describe a neural sequence labeling model for identification. Evaluation results show that discourse modes can be identified automatically with an average F1-score of 0.7. We further demonstrate that discourse modes can be used as features that improve automatic essay scoring (AES). The impacts of discourse modes for AES are also discussed.

[TACL] Winning on the Merits: The Joint Effects of Content and Style on Debate Outcomes
Lu Wang, Nick Beauchamp, Sarah Shugars, and Kechen Qin 11:27–11:45

Debate and deliberation play essential roles in politics and government, but most models presume that debates are won mainly via superior style or agenda control. Ideally, however, debates would be won on the merits, as a function of which side has the stronger arguments. We propose a predictive model of debate that estimates the effects of linguistic features and the latent persuasive strengths of different topics, as well as the interactions between the two. Using a dataset of 118 Oxford-style debates, our model's combination of content (as latent topics) and style (as linguistic features) allows us to predict audience-adjudicated winners with 74% accuracy, significantly outperforming linguistic features alone (66%). Our model finds that winning sides employ stronger arguments, and allows us to identify the linguistic features associated with strong or weak arguments.

Lexical Features in Coreference Resolution: To be Used With Caution
Nafise Sadat Moosavi and Michael Strube 11:46–11:58

Lexical features are a major source of information in state-of-the-art coreference resolvers. Lexical features implicitly model some of the linguistic phenomena at a fine granularity level. They are especially useful for representing the context of mentions. In this paper we investigate a drawback of using many lexical features in state-of-the-art coreference resolvers. We show that if coreference resolvers mainly rely on lexical features, they can hardly generalize to unseen domains. Furthermore, we show that the current coreference resolution evaluation is clearly flawed by only evaluating on a specific split of a specific dataset in which there is a notable overlap between the training, development and test sets.

Session Session 1D: Machine Translation 1

A Convolutional Encoder Model for Neural Machine Translation
Jonas Gehring, Michael Auli, David Grangier, and Yann Dauphin 10:30–10:48

The prevalent approach to neural machine translation relies on bi-directional LSTMs to encode the source sentence. We present a faster and simpler architecture based on a succession of convolutional layers. This allows to encode the source sentence simultaneously compared to recurrent networks for which computation is constrained by temporal dependencies. On WMT'16 English-Romanian translation we achieve competitive accuracy to the state-of-the-art and on WMT'15 English-German we outperform several recently published results. Our models obtain almost the same accuracy as a very deep LSTM setup on WMT'14 English-French translation. We speed up CPU decoding by more than two times at the same or higher accuracy as a strong bi-directional LSTM.

Deep Neural Machine Translation with Linear Associative Unit
Mingxuan Wang, Zhengdong Lu, Jie Zhou, and Qun Liu 10:49–11:07

Deep Neural Networks (DNNs) have provably enhanced the state-of-the-art Neural Machine Translation (NMT) with its capability in modeling complex functions and capturing complex linguistic structures. However NMT with deep architecture in its encoder or decoder RNNs often suffer from severe gradient diffusion due to the non-linear recurrent activations, which often makes the optimization much more difficult. To address this problem we propose a novel linear associative units (LAU) to reduce the gradient propagation path inside the recurrent unit. Different from conventional approaches (LSTM unit and GRU), LAUs uses linear associative connections between input and output of the recurrent unit, which allows unimpeded information flow through both space and time The model is quite simple, but it is surprisingly effective. Our empirical study on Chinese-English translation shows that our model with proper configuration can improve by 11.7 BLEU upon Groundhog and the best reported on results in the same setting. On WMT14 English-German task and a larger WMT14 English-French task, our model achieves comparable results with the state-of-the-art.

[TACL] A Polynomial-Time Dynamic Programming Algorithm for Phrase-Based Decoding with a Fixed Distortion Limit
Yin-Wen Chang and Michael Collins 11:08–11:26

Decoding of phrase-based translation models in the general case is known to be NP-complete, by a reduction from the traveling salesman problem (Knight, 1999). In practice, phrase-based systems often impose a hard distortion limit that limits the movement of phrases during translation. However, the impact on complexity after imposing such a constraint is not well studied. In this paper, we describe a dynamic programming algorithm for phrase-based decoding with a fixed distortion limit. The runtime of the algorithm is $O(nd!lh^{d+1})$ where n is the sentence length, d is the distortion limit, l is a bound on the number of phrases starting at any position in the sentence, and h is related to the maximum number of target language translations for any source word. The algorithm makes use of a novel representation that gives a new perspective on decoding of phrase-based models.

[TACL] Context Gates for Neural Machine Translation
Zhaopeng Tu, Yang Liu, Zhengdong Lu, Xiaohua Liu, and Hanh Li 11:27–11:45

In neural machine translation (NMT), generation of a target word depends on both source and target contexts. We find that source contexts have a direct impact on the adequacy of a translation while target contexts affect the fluency. Intuitively, generation of a content word should rely more on the source context and generation of a functional word should rely more on the target context. Due to the lack of effective control over the influence from source and target contexts, conventional NMT tends to yield fluent but inadequate translations. To address this problem, we propose context gates which dynamically control the ratios at which source and target contexts contribute to the generation of target words. In this way, we can enhance both the adequacy and fluency of NMT with more careful control of the information flow from contexts. Experiments show that our approach significantly improves upon a standard attention-based NMT system by +2.3 BLEU points.

Alternative Objective Functions for Training MT Evaluation Metrics
Miloš Stanojević and Khalil Sima'an 11:46–11:58

MT evaluation metrics are tested for correlation with human judgments either at the sentence- or the corpus-level. Trained metrics ignore corpus-level judgments and are trained for high sentence-level correlation only. We show that training only for one objective (sentence or corpus level), can not only harm the performance on the other objective, but it can also be suboptimal for the objective being optimized. To this end we present a metric trained for corpus-level and show empirical comparison against a metric trained for sentence-level

exemplifying how their performance may vary per language pair, type and level of judgment. Subsequently we propose a model trained to optimize both objectives simultaneously and show that it is far more stable than–and on average outperforms–both models on both objectives.

Session Session 1E: Generation 1

Neural AMR: Sequence-to-Sequence Models for Parsing and Generation
Ioannis Konstas, Srinivasan Iyer, Mark Yatskar, Yejin Choi, and Luke Zettlemoyer 10:30–10:48

Sequence-to-sequence models have shown strong performance across a broad range of applications. However, their application to parsing and generating text using Abstract Meaning Representation (AMR) has been limited, due to the relatively limited amount of labeled data and the non-sequential nature of the AMR graphs. We present a novel training procedure that can lift this limitation using millions of unlabeled sentences and careful preprocessing of the AMR graphs. For AMR parsing, our model achieves competitive results of 62.1 SMATCH, the current best score reported without significant use of external semantic resources. For AMR generation, our model establishes a new state-of-the-art performance of BLEU 33.8. We present extensive ablative and qualitative analysis including strong evidence that sequence-based AMR models are robust against ordering variations of graph-to-sequence conversions.

Program Induction by Rationale Generation: Learning to Solve and Explain Algebraic Word Problems
Wang Ling, Dani Yogatama, Chris Dyer, and Phil Blunsom 10:49–11:07

Solving algebraic word problems requires executing a series of arithmetic operations—a program—to obtain a final answer. However, since programs can be arbitrarily complicated, inducing them directly from question-answer pairs is a formidable challenge. To make this task more feasible, we solve these problems by generating answer rationales, sequences of natural language and human-readable mathematical expressions that derive the final answer through a series of small steps. Although rationales do not explicitly specify programs, they provide a scaffolding for their structure via intermediate milestones. To evaluate our approach, we have created a new 100,000-sample dataset of questions, answers and rationales. Experimental results show that indirect supervision of program learning via answer rationales is a promising strategy for inducing arithmetic programs.

Automatically Generating Rhythmic Verse with Neural Networks
Jack Hopkins and Douwe Kiela 11:08–11:26

We propose two novel methodologies for the automatic generation of rhythmic poetry in a variety of forms. The first approach uses a neural language model trained on a phonetic encoding to learn an implicit representation of both the form and content of English poetry. This model can effectively learn common poetic devices such as rhyme, rhythm and alliteration. The second approach considers poetry generation as a constraint satisfaction problem where a generative neural language model is tasked with learning a representation of content, and a discriminative weighted finite state machine constrains it on the basis of form. By manipulating the constraints of the latter model, we can generate coherent poetry with arbitrary forms and themes. A large-scale extrinsic evaluation demonstrated that participants consider machine-generated poems to be written by humans 54% of the time. In addition, participants rated a machine-generated poem to be the best amongst all evaluated.

Creating Training Corpora for NLG Micro-Planners
Claire Gardent, Anastasia Shimorina, Shashi Narayan, and Laura Perez-Beltrachini 11:27–11:45

In this paper, we present a novel framework for semi-automatically creating linguistically challenging micro-planning data-to-text corpora from existing Knowledge Bases. Because our method pairs data of varying size and shape with texts ranging from simple clauses to short texts, a dataset created using this framework provides a challenging benchmark for microplanning. Another feature of this framework is that it can be applied to any large scale knowledge base and can therefore be used to train and learn KB verbalisers. We apply our framework to DBpedia data and compare the resulting dataset with Wen et al. 2016's. We show that while Wen et al.'s dataset is more than twice larger than ours, it is less diverse both in terms of input and in terms of text. We thus propose our corpus generation framework as a novel method for creating challenging data sets from which NLG models can be learned which are capable of handling the complex interactions occurring during in micro-planning between lexicalisation, aggregation, surface realisation, referring expression generation and sentence segmentation. To encourage researchers to take up this challenge, we made available a dataset of 21,855 data/text pairs created using this framework in the context of the WebNLG shared task.

A Principled Framework for Evaluating Summarizers: Comparing Models of Summary Quality against Human Judgments
Maxime Peyrard and Judith Eckle-Kohler 11:46–11:58

We present a new framework for evaluating extractive summarizers, which is based on a principled representation as optimization problem. We prove that every extractive summarizer can be decomposed into an objective function and an optimization technique. We perform a comparative analysis and evaluation of several objec-

tive functions embedded in well-known summarizers regarding their correlation with human judgments. Our comparison of these correlations across two datasets yields surprising insights into the role and performance of objective functions in the different summarizers.

Session 2 Overview – Monday, July 31, 2017

	Track A	**Track B**	**Track C**	**Track D**	**Track E**
	Question Answering 1	*Vision 1*	*Syntax 1*	*Machine Learning 1 (NN)*	*Sentiment 1 (NN)*
	Salons B/C	Salons E/F	Salon D	Salon 1	Salons 2/3
13:40	Gated Self-Matching Networks for Reading Comprehension and Question Answering *W. Wang, N. Yang, F. Wei, B. Chang, and M. Zhou*	Translating Neuralese *J. Andreas, A. Dragan, and D. Klein*	A* CCG Parsing with a Supertag and Dependency Factored Model *M. Yoshikawa, H. Noji, and Y. Matsumoto*	Multi-space Variational Encoder-Decoders for Semi-supervised Labeled Sequence Transduction *C. Zhou and G. Neubig*	Handling Cold-Start Problem in Review Spam Detection by Jointly Embedding Texts and Behaviors *X. Wang, K. Liu, and J. Zhao*
13:59	Generating Natural Answers by Incorporating Copying and Retrieving Mechanisms in Sequence-to-Sequence Learning *S. He, C. Liu, K. Liu, and J. Zhao*	Obtaining referential word meanings from visual and distributional information: Experiments on object naming *S. Zarrieß and D. Schlangen*	A Full Non-Monotonic Transition System for Unrestricted Non-Projective Parsing *D. Fernández-González and C. Gómez-Rodríguez*	Scalable Bayesian Learning of Recurrent Neural Networks for Language Modeling *Z. Gan, C. Li, C. Chen, Y. Pu, Q. Su, and L. Carin*	Learning Cognitive Features from Gaze Data for Sentiment and Sarcasm Classification using Convolutional Neural Network *A. Mishra, K. Dey, and P. Bhattacharyya*
14:18	Coarse-to-Fine Question Answering for Long Documents *E. Choi, D. Hewlett, J. Uszkoreit, I. Polosukhin, A. Lacoste, and J. Berant*	FOIL it! Find One mismatch between Image and Language caption *R. Shekhar, S. Pezzelle, Y. Klimovich, A. Herbelot, M. Nabi, E. Sangineto, and R. Bernardi*	Aggregating and Predicting Sequence Labels from Crowd Annotations *A. T. Nguyen, B. Wallace, J. J. Li, A. Nenkova, and M. Lease*	Learning attention for historical text normalization by learning to pronounce *M. Bollmann, J. Bingel, and A. Søgaard*	An Unsupervised Neural Attention Model for Aspect Extraction *R. He, W. S. Lee, H. T. Ng, and D. Dahlmeier*
14:37	An End-to-End Model for Question Answering over Knowledge Base with Cross-Attention Combining Global Knowledge *Y. Hao, Y. Zhang, K. Liu, S. He, Z. Liu, H. Wu, and J. Zhao*	Verb Physics: Relative Physical Knowledge of Actions and Objects *M. Forbes and Y. Choi*	[TACL] Fine-Grained Prediction of Syntactic Typology: Discovering Latent Structure with Supervised Learning *D. Wang and J. Eisner*	Deep Learning in Semantic Kernel Spaces *D. Croce, S. Filice, G. Castellucci, and R. Basili*	Other Topics You May Also Agree or Disagree: Modeling Inter-Topic Preferences using Tweets and Matrix Factorization *A. Sasaki, K. Hanawa, N. Okazaki, and K. Inui*
14:56	[TACL] Domain-Targeted, High Precision Knowledge Extraction *B. Dalvi, N. Tandon, and P. Clark*	[TACL] Visually Grounded and Textual Semantic Models Differentially Decode Brain Activity Associated with Concrete and Abstract Nouns *A. J. Anderson, D. Kiela, S. Clark, and M. Poesio*	[TACL] Learning to Prune: Exploring the Frontier of Fast and Accurate Parsing *T. Vieira and J. Eisner*	Topically Driven Neural Language Model *J. H. Lau, T. Baldwin, and T. Cohn*	[TACL] Overcoming Language Variation in Sentiment Analysis with Social Attention *Y. Yang and J. Eisenstein*

Parallel Session 2

Session Session 2A: Question Answering 1

Gated Self-Matching Networks for Reading Comprehension and Question Answering
Wenhui Wang, Nan Yang, Furu Wei, Baobao Chang, and Ming Zhou 13:40–13:58

In this paper, we present the gated self-matching networks for reading comprehension style question answering, which aims to answer questions from a given passage. We first match the question and passage with gated attention-based recurrent networks to obtain the question-aware passage representation. Then we propose a self-matching attention mechanism to refine the representation by matching the passage against itself, which effectively encodes information from the whole passage. We finally employ the pointer networks to locate the positions of answers from the passages. We conduct extensive experiments on the SQuAD dataset. The single model achieves 71.3% on the evaluation metrics of exact match on the hidden test set, while the ensemble model further boosts the results to 75.9%. At the time of submission of the paper, our model holds the first place on the SQuAD leaderboard for both single and ensemble model.

Generating Natural Answers by Incorporating Copying and Retrieving Mechanisms in Sequence-to-Sequence Learning
Shizhu He, Cao Liu, Kang Liu, and Jun Zhao 13:59–14:17

Generating answer with natural language sentence is very important in real-world question answering systems, which needs to obtain a right answer as well as a coherent natural response. In this paper, we propose an end-to-end question answering system called COREQA in sequence-to-sequence learning, which incorporates copying and retrieving mechanisms to generate natural answers within an encoder-decoder framework. Specifically, in COREQA, the semantic units (words, phrases and entities) in a natural answer are dynamically predicted from the vocabulary, copied from the given question and/or retrieved from the corresponding knowledge base jointly. Our empirical study on both synthetic and real-world datasets demonstrates the efficiency of COREQA, which is able to generate correct, coherent and natural answers for knowledge inquired questions.

Coarse-to-Fine Question Answering for Long Documents
Eunsol Choi, Daniel Hewlett, Jakob Uszkoreit, Illia Polosukhin, Alexandre Lacoste, and Jonathan Berant 14:18–14:36

We present a framework for question answering that can efficiently scale to longer documents while maintaining or even improving performance of state-of-the-art models. While most successful approaches for reading comprehension rely on recurrent neural networks (RNNs), running them over long documents is prohibitively slow because it is difficult to parallelize over sequences. Inspired by how people first skim the document, identify relevant parts, and carefully read these parts to produce an answer, we combine a coarse, fast model for selecting relevant sentences and a more expensive RNN for producing the answer from those sentences. We treat sentence selection as a latent variable trained jointly from the answer only using reinforcement learning. Experiments demonstrate state-of-the-art performance on a challenging subset of the WikiReading dataset and on a new dataset, while speeding up the model by 3.5x-6.7x.

An End-to-End Model for Question Answering over Knowledge Base with Cross-Attention Combining Global Knowledge
Yanchao Hao, Yuanzhe Zhang, Kang Liu, Shizhu He, Zhanyi Liu, Hua Wu, and Jun Zhao 14:37–14:55

With the rapid growth of knowledge bases (KBs) on the web, how to take full advantage of them becomes increasingly important. Question answering over knowledge base (KB-QA) is one of the promising approaches to access the substantial knowledge. Meanwhile, as the neural network-based (NN-based) methods develop, NN-based KB-QA has already achieved impressive results. However, previous work did not put more emphasis on question representation, and the question is converted into a fixed vector regardless of its candidate answers. This simple representation strategy is not easy to express the proper information in the question. Hence, we present an end-to-end neural network model to represent the questions and their corresponding scores dynamically according to the various candidate answer aspects via cross-attention mechanism. In addition, we leverage the global knowledge inside the underlying KB, aiming at integrating the rich KB information into the representation of the answers. As a result, it could alleviates the out-of-vocabulary (OOV) problem, which helps the cross-attention model to represent the question more precisely. The experimental results on WebQuestions demonstrate the effectiveness of the proposed approach.

[TACL] Domain-Targeted, High Precision Knowledge Extraction
Bhavana Dalvi, Niket Tandon, and Peter Clark 14:56–15:14

Our goal is to construct a domain-targeted, high precision knowledge base (KB), containing general (subject,predicate,object) statements about the world, in support of a downstream question-answering (QA) application. Despite recent advances in information extraction (IE) techniques, no suitable resource for our task already exists; existing resources are either too noisy, too named-entity centric, or too incomplete, and typically have not been constructed with a clear scope or purpose. To address these, we have created a domain-targeted, high precision knowledge extraction pipeline, leveraging Open IE, crowdsourcing, and a novel canonical schema learning algorithm (called CASI), that produces high precision knowledge targeted to a particular domain - in our case, elementary science. To measure the KB's coverage of the target domain's knowledge (it's "comprehensiveness" with respect to science) we measure recall with respect to an independent corpus of domain text, and show that our pipeline produces output with over 80% precision and 23% recall with respect to that target, a substantially higher coverage of tuple-expressible science knowledge than other comparable resources. We have made the KB publicly available at http://allenai.org/data/aristo-tuple-kb.

Session Session 2B: Vision 1

Translating Neuralese
Jacob Andreas, Anca Dragan, and Dan Klein 13:40–13:58

Several approaches have recently been proposed for learning decentralized deep multiagent policies that coordinate via a differentiable communication channel. While these policies are effective for many tasks, interpretation of their induced communication strategies has remained a challenge. Here we propose to interpret agents' messages by translating them. Unlike in typical machine translation problems, we have no parallel data to learn from. Instead we develop a translation model based on the insight that agent messages and natural language strings mean the same thing if they induce the same belief about the world in a listener. We present theoretical guarantees and empirical evidence that our approach preserves both the semantics and pragmatics of messages by ensuring that players communicating through a translation layer do not suffer a substantial loss in reward relative to players with a common language.

Obtaining referential word meanings from visual and distributional information: Experiments on object naming
Sina Zarrieß and David Schlangen 13:59–14:17

We investigate object naming, which is an important sub-task of referring expression generation on real-world images. As opposed to mutually exclusive labels used in object recognition, object names are more flexible, subject to communicative preferences and semantically related to each other. Therefore, we investigate models of referential word meaning that link visual to lexical information which we assume to be given through distributional word embeddings. We present a model that learns individual predictors for object names that link visual and distributional aspects of word meaning during training. We show that this is particularly beneficial for zero-shot learning, as compared to projecting visual objects directly into the distributional space. In a standard object naming task, we find that different ways of combining lexical and visual information achieve very similar performance, though experiments on model combination suggest that they capture complementary aspects of referential meaning.

FOIL it! Find One mismatch between Image and Language caption
Ravi Shekhar, Sandro Pezzelle, Yauhen Klimovich, Aurélie Herbelot, Moin Nabi, Enver Sangineto, and Raffaella Bernardi 14:18–14:36

In this paper, we aim to understand whether current language and vision (LaVi) models truly grasp the interaction between the two modalities. To this end, we propose an extension of the MS-COCO dataset, FOIL-COCO, which associates images with both correct and 'foil' captions, that is, descriptions of the image that are highly similar to the original ones, but contain one single mistake ('foil word'). We show that current LaVi models fall into the traps of this data and perform badly on three tasks: a) caption classification (correct vs. foil); b) foil word detection; c) foil word correction. Humans, in contrast, have near-perfect performance on those tasks. We demonstrate that merely utilising language cues is not enough to model FOIL-COCO and that it challenges the state-of-the-art by requiring a fine-grained understanding of the relation between text and image.

Verb Physics: Relative Physical Knowledge of Actions and Objects
Maxwell Forbes and Yejin Choi 14:37–14:55

Learning commonsense knowledge from natural language text is nontrivial due to reporting bias: people rarely state the obvious, e.g., "My house is bigger than me." However, while rarely stated explicitly, this trivial everyday knowledge does influence the way people talk about the world, which provides indirect clues to reason about the world. For example, a statement like, "Tyler entered his house" implies that his house is bigger than Tyler. In this paper, we present an approach to infer relative physical knowledge of actions and objects along five dimensions (e.g., size, weight, and strength) from unstructured natural language text. We frame knowledge acquisition as joint inference over two closely related problems: learning (1) relative physical knowledge of object pairs and (2) physical implications of actions when applied to those object pairs. Empirical results demonstrate that it is possible to extract knowledge of actions and objects from language and that joint inference over different types of knowledge improves performance.

[TACL] Visually Grounded and Textual Semantic Models Differentially Decode Brain Activity Associated with Concrete and Abstract Nouns
Andrew J. Anderson, Douwe Kiela, Stephen Clark, and Massimo Poesio 14:56–15:14

Important advances have recently been made using computational semantic models to decode brain activity patterns associated with concepts; however, this work has almost exclusively focused on concrete nouns. How well these models extend to decoding abstract nouns is largely unknown. We address this question by applying

state-of-the-art computational models to decode functional Magnetic Resonance Imaging (fMRI) activity patterns, elicited by participants reading and imagining a diverse set of both concrete and abstract nouns. One of the models we use is linguistic, exploiting the recent word2vec skipgram approach trained on Wikipedia. The second is visually grounded, using deep convolutional neural networks trained on Google Images. Dual coding theory considers concrete concepts to be encoded in the brain both linguistically and visually, and abstract concepts only linguistically. Splitting the fMRI data according to human concreteness ratings, we indeed observe that both models significantly decode the most concrete nouns; however, accuracy is significantly greater using the text-based models for the most abstract nouns. More generally this confirms that current computational models are sufficiently advanced to assist in investigating the representational structure of abstract concepts in the brain.

Session Session 2C: Syntax 1

A* CCG Parsing with a Supertag and Dependency Factored Model
Masashi Yoshikawa, Hiroshi Noji, and Yuji Matsumoto 13:40–13:58

We propose a new A* CCG parsing model in which the probability of a tree is decomposed into factors of CCG categories and its syntactic dependencies both defined on bi-directional LSTMs. Our factored model allows the precomputation of all probabilities and runs very efficiently, while modeling sentence structures explicitly via dependencies. Our model achieves the state-of-the-art results on English and Japanese CCG parsing.

A Full Non-Monotonic Transition System for Unrestricted Non-Projective Parsing
Daniel Fernández-González and Carlos Gómez-Rodríguez 13:59–14:17

Restricted non-monotonicity has been shown beneficial for the projective arc-eager dependency parser in previous research, as posterior decisions can repair mistakes made in previous states due to the lack of information. In this paper, we propose a novel, fully non-monotonic transition system based on the non-projective Covington algorithm. As a non-monotonic system requires exploration of erroneous actions during the training process, we develop several non-monotonic variants of the recently defined dynamic oracle for the Covington parser, based on tight approximations of the loss. Experiments on datasets from the CoNLL-X and CoNLL-XI shared tasks show that a non-monotonic dynamic oracle outperforms the monotonic version in the majority of languages.

Aggregating and Predicting Sequence Labels from Crowd Annotations
An Thanh Nguyen, Byron Wallace, Junyi Jessy Li, Ani Nenkova, and Matthew Lease 14:18–14:36

Despite sequences being core to NLP, scant work has considered how to handle noisy sequence labels from multiple annotators for the same text. Given such annotations, we consider two complementary tasks: (1) aggregating sequential crowd labels to infer a best single set of consensus annotations; and (2) using crowd annotations as training data for a model that can predict sequences in unannotated text. For aggregation, we propose a novel Hidden Markov Model variant. To predict sequences in unannotated text, we propose a neural approach using Long Short Term Memory. We evaluate a suite of methods across two different applications and text genres: Named-Entity Recognition in news articles and Information Extraction from biomedical abstracts. Results show improvement over strong baselines. Our source code and data are available online.

[TACL] Fine-Grained Prediction of Syntactic Typology: Discovering Latent Structure with Supervised Learning
Dingquan Wang and Jason Eisner 14:37–14:55

We show how to predict the basic word-order facts of a novel language given only a corpus of part-of-speech (POS) sequences. We predict how often direct objects follow their verbs, how often adjectives follow their nouns, and in general the directionalities of all dependency relations. Such typological properties could be helpful in grammar induction. While such a problem is usually regarded as unsupervised learning, our innovation is to treat it as supervised learning, using a large collection of realistic synthetic languages as training data. The supervised learner must identify surface features of a language's POS sequence (hand-engineered or neural features) that correlate with the language's deeper structure (latent trees). In the experiment, we show: 1) Given a small set of real languages, it helps to add many synthetic languages to the training data. 2) Our system is robust even when the POS sequences include noise. 3) Our system on this task outperforms a grammar induction baseline by a large margin.

[TACL] Learning to Prune: Exploring the Frontier of Fast and Accurate Parsing
Tim Vieira and Jason Eisner 14:56–15:14

Pruning hypotheses during dynamic programming is commonly used to speed up inference in settings such as parsing. Unlike prior work, we train a pruning policy under an objective that measures end-to-end performance: we search for a fast and accurate policy. This poses a difficult machine learning problem, which we tackle with the LOLS algorithm. LOLS training must continually compute the effects of changing pruning decisions: we show how to make this efficient in the constituency parsing setting, via dynamic programming and change propagation algorithms. We find that optimizing end-to-end performance in this way leads to a better Pareto frontier—i.e., parsers which are more accurate for a given runtime.

Session Session 2D: Machine Learning 1 (NN)

Multi-space Variational Encoder-Decoders for Semi-supervised Labeled Sequence Transduction

Chunting Zhou and Graham Neubig 13:40–13:58

Labeled sequence transduction is a task of transforming one sequence into another sequence that satisfies desiderata specified by a set of labels. In this paper we propose multi-space variational encoder-decoders, a new model for labeled sequence transduction with semi-supervised learning. The generative model can use neural networks to handle both discrete and continuous latent variables to exploit various features of data. Experiments show that our model provides not only a powerful supervised framework but also can effectively take advantage of the unlabeled data. On the SIGMORPHON morphological inflection benchmark, our model outperforms single-model state-of-art results by a large margin for the majority of languages.

Scalable Bayesian Learning of Recurrent Neural Networks for Language Modeling

Zhe Gan, Chunyuan Li, Changyou Chen, Yunchen Pu, Qinliang Su, and Lawrence Carin 13:59–14:17

Recurrent neural networks (RNNs) have shown promising performance for language modeling. However, traditional training of RNNs using back-propagation through time often suffers from overfitting. One reason for this is that stochastic optimization (used for large training sets) does not provide good estimates of model uncertainty. This paper leverages recent advances in stochastic gradient Markov Chain Monte Carlo (also appropriate for large training sets) to learn weight uncertainty in RNNs. It yields a principled Bayesian learning algorithm, adding gradient noise during training (enhancing exploration of the model-parameter space) and model averaging when testing. Extensive experiments on various RNN models and across a broad range of applications demonstrate the superiority of the proposed approach relative to stochastic optimization.

Learning attention for historical text normalization by learning to pronounce

Marcel Bollmann, Joachim Bingel, and Anders Søgaard 14:18–14:36

Automated processing of historical texts often relies on pre-normalization to modern word forms. Training encoder-decoder architectures to solve such problems typically requires a lot of training data, which is not available for the named task. We address this problem by using several novel encoder-decoder architectures, including a multi-task learning (MTL) architecture using a grapheme-to-phoneme dictionary as auxiliary data, pushing the state-of-the-art by an absolute 2% increase in performance. We analyze the induced models across 44 different texts from Early New High German. Interestingly, we observe that, as previously conjectured, multi-task learning can learn to focus attention during decoding, in ways remarkably similar to recently proposed attention mechanisms. This, we believe, is an important step toward understanding how MTL works.

Deep Learning in Semantic Kernel Spaces

Danilo Croce, Simone Filice, Giuseppe Castellucci, and Roberto Basili 14:37–14:55

Kernel methods enable the direct usage of structured representations of textual data during language learning and inference tasks. Expressive kernels, such as Tree Kernels, achieve excellent performance in NLP. On the other side, deep neural networks have been demonstrated effective in automatically learning feature representations during training. However, their input is tensor data, i.e., they can not manage rich structured information. In this paper, we show that expressive kernels and deep neural networks can be combined in a common framework in order to (i) explicitly model structured information and (ii) learn non-linear decision functions. We show that the input layer of a deep architecture can be pre-trained through the application of the Nystrom low-rank approximation of kernel spaces. The resulting "kernelized" neural network achieves state-of-the-art accuracy in three different tasks.

Topically Driven Neural Language Model

Jey Han Lau, Timothy Baldwin, and Trevor Cohn 14:56–15:14

Language models are typically applied at the sentence level, without access to the broader document context. We present a neural language model that incorporates document context in the form of a topic model-like architecture, thus providing a succinct representation of the broader document context outside of the current sentence. Experiments over a range of datasets demonstrate that our model outperforms a pure sentence-based model in terms of language model perplexity, and leads to topics that are potentially more coherent than those produced by a standard LDA topic model. Our model also has the ability to generate related sentences for a topic, providing another way to interpret topics.

Session Session 2E: Sentiment 1 (NN)

Handling Cold-Start Problem in Review Spam Detection by Jointly Embedding Texts and Behaviors
Xuepeng Wang, Kang Liu, and Jun Zhao 13:40–13:58

Solving cold-start problem in review spam detection is an urgent and significant task. It can help the online review websites to relieve the damage of spammers in time, but has never been investigated by previous work. This paper proposes a novel neural network model to detect review spam for cold-start problem, by learning to represent the new reviewers' review with jointly embedded textual and behavioral information. Experimental results prove the proposed model achieves an effective performance and possesses preferable domain-adaptability. It is also applicable to a large scale dataset in an unsupervised way.

Learning Cognitive Features from Gaze Data for Sentiment and Sarcasm Classification using Convolutional Neural Network
Abhijit Mishra, Kuntal Dey, and Pushpak Bhattacharyya 13:59–14:17

Cognitive NLP systems- i.e., NLP systems that make use of behavioral data - augment traditional text-based features with cognitive features extracted from eye-movement patterns, EEG signals, brain-imaging etc. Such extraction of features is typically manual. We contend that manual extraction of features may not be the best way to tackle text subtleties that characteristically prevail in complex classification tasks like Sentiment Analysis and Sarcasm Detection, and that even the extraction and choice of features should be delegated to the learning system. We introduce a framework to automatically extract cognitive features from the eye-movement/gaze data of human readers reading the text and use them as features along with textual features for the tasks of sentiment polarity and sarcasm detection. Our proposed framework is based on Convolutional Neural Network (CNN). The CNN learns features from both gaze and text and uses them to classify the input text. We test our technique on published sentiment and sarcasm labeled datasets, enriched with gaze information, to show that using a combination of automatically learned text and gaze features often yields better classification performance over (i) CNN based systems that rely on text input alone and (ii) existing systems that rely on handcrafted gaze and textual features.

An Unsupervised Neural Attention Model for Aspect Extraction
Ruidan He, Wee Sun Lee, Hwee Tou Ng, and Daniel Dahlmeier 14:18–14:36

Aspect extraction is an important and challenging task in aspect-based sentiment analysis. Existing works tend to apply variants of topic models on this task. While fairly successful, these methods usually do not produce highly coherent aspects. In this paper, we present a novel neural approach with the aim of discovering coherent aspects. The model improves coherence by exploiting the distribution of word co-occurrences through the use of neural word embeddings. Unlike topic models which typically assume independently generated words, word embedding models encourage words that appear in similar contexts to be located close to each other in the embedding space. In addition, we use an attention mechanism to de-emphasize irrelevant words during training, further improving the coherence of aspects. Experimental results on real-life datasets demonstrate that our approach discovers more meaningful and coherent aspects, and substantially outperforms baseline methods on several evaluation tasks.

Other Topics You May Also Agree or Disagree: Modeling Inter-Topic Preferences using Tweets and Matrix Factorization
Akira Sasaki, Kazuaki Hanawa, Naoaki Okazaki, and Kentaro Inui 14:37–14:55

We presents in this paper our approach for modeling inter-topic preferences of Twitter users: for example, "those who agree with the Trans-Pacific Partnership (TPP) also agree with free trade". This kind of knowledge is useful not only for stance detection across multiple topics but also for various real-world applications including public opinion survey, electoral prediction, electoral campaigns, and online debates. In order to extract users' preferences on Twitter, we design linguistic patterns in which people agree and disagree about specific topics (e.g., "A is completely wrong"). By applying these linguistic patterns to a collection of tweets, we extract statements agreeing and disagreeing with various topics. Inspired by previous work on item recommendation, we formalize the task of modeling inter-topic preferences as matrix factorization: representing users' preference as a user-topic matrix and mapping both users and topics onto a latent feature space that abstracts the preferences. Our experimental results demonstrate both that our presented approach is useful in predicting missing preferences of users and that the latent vector representations of topics successfully encode inter-topic preferences.

[TACL] Overcoming Language Variation in Sentiment Analysis with Social Attention
Yi Yang and Jacob Eisenstein 14:56–15:14

Variation in language is ubiquitous, and is particularly evident in newer forms of writing such as social media. Fortunately, variation is not random, but is usually linked to social factors. By exploiting linguistic homophily — the tendency of socially linked individuals to use language similarly — it is possible to build models that are more robust to variation. In this paper, we focus on social network structures, which make it possible to generalize sociolinguistic properties from authors in the training set to authors in the test sets, without requiring demographic author metadata. We explore the social information by leveraging author embeddings, training with social relations between authors. We introduce an attention based neural architecture — the prediction is based on several basis models with emphasis on different regions in the network, where socially connected users share similar attention weights on these models. We are able to improve the overall accuracies of Twitter sentiment analysis and review sentiment analysis by significant margins over competitive prior work.

Session 3 Overview – Monday, July 31, 2017

	Track A	**Track B**	**Track C**	**Track D**	**Track E**
	Information Extraction 2 / Biomedical 1	*Semantics 2 (NN)*	*Speech 1 / Dialogue 1*	*Multilingual 1*	*Phonology 1*
	Salons B/C	Salons E/F	Salon D	Salon 1	Salons 2/3
15:45	Automatically Labeled Data Generation for Large Scale Event Extraction *Y. Chen, S. Liu, X. Zhang, K. Liu, and J. Zhao*	A Syntactic Neural Model for General-Purpose Code Generation *P. Yin and G. Neubig*	Towards End-to-End Reinforcement Learning of Dialogue Agents for Information Access *B. Dhingra, L. Li, X. Li, J. Gao, Y.-N. Chen, F. Ahmed, and L. Deng*	Found in Translation: Reconstructing Phylogenetic Language Trees from Translations *E. Rabinovich, N. Ordan, and S. Wintner*	MORSE: Semantically Drive-n MORpheme SEgment-er *T. Sakakini, S. Bhat, and P. Viswanath*
16:04	Time Expression Analysis and Recognition Using Syntactic Token Types and General Heuristic Rules *X. Zhong, A. Sun, and E. Cambria*	Learning bilingual word embeddings with (almost) no bilingual data *M. Artetxe, G. Labaka, and E. Agirre*	Sequential Matching Network: A New Architecture for Multi-turn Response Selection in Retrieval-Based Chatbots *Y. Wu, W. Wu, C. Xing, M. Zhou, and Z. Li*	Predicting Native Language from Gaze *Y. Berzak, C. Nakamura, S. Flynn, and B. Katz*	[TACL] A Generative Model of Phonotactics *R. Futrell, A. Albright, P. Graff, and T. J. O'Donnell*
16:23	Learning with Noise: Enhance Distantly Supervised Relation Extraction with Dynamic Transition Matrix *B. Luo, Y. Feng, Z. Wang, Z. Zhu, S. Huang, R. Yan, and D. Zhao*	Abstract Meaning Representation Parsing using LSTM Recurrent Neural Networks *W. Foland and J. H. Martin*	Learning Word-Like Units from Joint Audio-Visual Analysis *D. Harwath and J. Glass*	[TACL] Decoding Anagrammed Texts Written in an Unknown Language and Script *B. Hauer and G. Kondrak*	[TACL] Joint Semantic Synthesis and Morphological Analysis of the Derived Word *R. Cotterell and H. Schütze*
16:42	[TACL] Ordinal Common-sense Inference *S. Zhang, R. Rudinger, K. Duh, and B. Van Durme*	Deep Semantic Role Labeling: What Works and What's Next *L. He, K. Lee, M. Lewis, and L. Zettlemoyer*	Joint CTC/attention decoding for end-to-end speech recognition *T. Hori, S. Watanabe, and J. Hershey*	[TACL] Sparse Coding of Neural Word Embeddings for Multilingual Sequence Labeling *G. Berend*	[TACL] Unsupervised Learning of Morphological Forests *J. Luo and K. Narasimhan*
17:01	Vector space models for evaluating semantic fluency in autism *E. Prud'hommeaux, J. van Santen, and D. Gliner*	Neural Architectures for Multilingual Semantic Parsing *R. H. Susanto and W. Lu*	Incorporating Uncertainty into Deep Learning for Spoken Language Assessment *A. Malinin, A. Ragni, K. Knill, and M. Gales*	Incorporating Dialectal Variability for Socially Equitable Language Identification *D. Jurgens, Y. Tsvetkov, and D. Jurafsky*	Evaluating Compound Splitters Extrinsically with Textual Entailment *G. Jagfeld, P. Ziering, and L. van der Plas*

Parallel Session 3

Session Session 3A: Information Extraction 2 / Biomedical 1

Automatically Labeled Data Generation for Large Scale Event Extraction
Yubo Chen, Shulin Liu, Xiang Zhang, Kang Liu, and Jun Zhao 15:45–16:03

Modern models of event extraction for tasks like ACE are based on supervised learning of events from small hand-labeled data. However, hand-labeled training data is expensive to produce, in low coverage of event types, and limited in size, which makes supervised methods hard to extract large scale of events for knowledge base population. To solve the data labeling problem, we propose to automatically label training data for event extraction via world knowledge and linguistic knowledge, which can detect key arguments and trigger words for each event type and employ them to label events in texts automatically. The experimental results show that the quality of our large scale automatically labeled data is competitive with elaborately human-labeled data. And our automatically labeled data can incorporate with human-labeled data, then improve the performance of models learned from these data.

Time Expression Analysis and Recognition Using Syntactic Token Types and General Heuristic Rules
Xiaoshi Zhong, Aixin Sun, and Erik Cambria 16:04–16:22

Extracting time expressions from free text is a fundamental task for many applications. We analyze the time expressions from four datasets and find that only a small group of words are used to express time information, and the words in time expressions demonstrate similar syntactic behaviour. Based on the findings, we propose a type-based approach, named SynTime, to recognize time expressions. Specifically, we define three main syntactic token types, namely time token, modifier, and numeral, to group time-related regular expressions over tokens. On the types we design general heuristic rules to recognize time expressions. In recognition, SynTime first identifies the time tokens from raw text, then searches their surroundings for modifiers and numerals to form time segments, and finally merges the time segments to time expressions. As a light-weight rule-based tagger, SynTime runs in real time, and can be easily expanded by simply adding keywords for the text of different types and of different domains. Experiment on benchmark datasets and tweets data shows that SynTime outperforms state-of-the-art methods.

Learning with Noise: Enhance Distantly Supervised Relation Extraction with Dynamic Transition Matrix
Bingfeng Luo, Yansong Feng, Zheng Wang, Zhanxing Zhu, Songfang Huang, Rui Yan, and Dongyan Zhao 16:23–16:41

Distant supervision significantly reduces human efforts in building training data for many classification tasks. While promising, this technique often introduces noise to the generated training data, which can severely affect the model performance. In this paper, we take a deep look at the application of distant supervision in relation extraction. We show that the dynamic transition matrix can effectively characterize the noise in the training data built by distant supervision. The transition matrix can be effectively trained using a novel curriculum learning based method without any direct supervision about the noise. We thoroughly evaluate our approach under a wide range of extraction scenarios. Experimental results show that our approach consistently improves the extraction results and outperforms the state-of-the-art in various evaluation scenarios.

[TACL] Ordinal Common-sense Inference
Sheng Zhang, Rachel Rudinger, Kevin Duh, and Benjamin Van Durme 16:42–17:00

Humans have the capacity to draw commonsense inferences from natural language: various things that are likely but not certain to hold based on established discourse, and are rarely stated explicitly. We propose an evaluation of automated common-sense inference based on an extension of recognizing textual entailment: predicting ordinal human responses on the subjective likelihood of an inference holding in a given context. We describe a framework for extracting common-sense knowledge for corpora, which is then used to construct a dataset for this ordinal entailment task. We train a neural sequence-to-sequence model on this dataset, which we use to score and generate possible inferences. Further, we annotate subsets of previously established datasets via our ordinal annotation protocol in order to then analyze the distinctions between these and what we have constructed.

Vector space models for evaluating semantic fluency in autism
Emily Prud'hommeaux, Jan van Santen, and Douglas Gliner 17:01–17:12

A common test administered during neurological examination is the semantic fluency test, in which the patient must list as many examples of a given semantic category as possible under timed conditions. Poor performance is associated with neurological conditions characterized by impairments in executive function, such as dementia, schizophrenia, and autism spectrum disorder (ASD). Methods for analyzing semantic fluency responses at the level of detail necessary to uncover these differences have typically relied on subjective manual annotation. In this paper, we explore automated approaches for scoring semantic fluency responses that leverage ontological resources and distributional semantic models to characterize the semantic fluency responses produced by young children with and without ASD. Using these methods, we find significant differences in the semantic fluency responses of children with ASD, demonstrating the utility of using objective methods for clinical language analysis.

Session Session 3B: Semantics 2 (NN)

A Syntactic Neural Model for General-Purpose Code Generation
Pengcheng Yin and Graham Neubig 15:45–16:03

We consider the problem of parsing natural language descriptions into source code written in a general-purpose programming language like Python. Existing data-driven methods treat this problem as a language generation task without considering the underlying syntax of the target programming language. Informed by previous work in semantic parsing, in this paper we propose a novel neural architecture powered by a grammar model to explicitly capture the target syntax as prior knowledge. Experiments find this an effective way to scale up to generation of complex programs from natural language descriptions, achieving state-of-the-art results that well outperform previous code generation and semantic parsing approaches.

Learning bilingual word embeddings with (almost) no bilingual data
Mikel Artetxe, Gorka Labaka, and Eneko Agirre 16:04–16:22

Most methods to learn bilingual word embeddings rely on large parallel corpora, which is difficult to obtain for most language pairs. This has motivated an active research line to relax this requirement, with methods that use document-aligned corpora or bilingual dictionaries of a few thousand words instead. In this work, we further reduce the need of bilingual resources using a very simple self-learning approach that can be combined with any dictionary-based mapping technique. Our method exploits the structural similarity of embedding spaces, and works with as little bilingual evidence as a 25 word dictionary or even an automatically generated list of numerals, obtaining results comparable to those of systems that use richer resources.

Abstract Meaning Representation Parsing using LSTM Recurrent Neural Networks
William Foland and James H. Martin 16:23–16:41

We present a system which parses sentences into Abstract Meaning Representations, improving state-of-the-art results for this task by more than 5%. AMR graphs represent semantic content using linguistic properties such as semantic roles, coreference, negation, and more. The AMR parser does not rely on a syntactic pre-parse, or heavily engineered features, and uses five recurrent neural networks as the key architectural components for inferring AMR graphs.

Deep Semantic Role Labeling: What Works and What's Next
Luheng He, Kenton Lee, Mike Lewis, and Luke Zettlemoyer 16:42–17:00

We introduce a new deep learning model for semantic role labeling (SRL) that significantly improves the state of the art, along with detailed analyses to reveal its strengths and limitations. We use a deep highway BiLSTM architecture with constrained decoding, while observing a number of recent best practices for initialization and regularization. Our 8-layer ensemble model achieves 83.2 F1 on theCoNLL 2005 test set and 83.4 F1 on CoNLL 2012, roughly a 10% relative error reduction over the previous state of the art. Extensive empirical analysis of these gains show that (1) deep models excel at recovering long-distance dependencies but can still make surprisingly obvious errors, and (2) that there is still room for syntactic parsers to improve these results.

Neural Architectures for Multilingual Semantic Parsing
Raymond Hendy Susanto and Wei Lu 17:01–17:12

In this paper, we address semantic parsing in a multilingual context. We train one multilingual model that is capable of parsing natural language sentences from multiple different languages into their corresponding formal semantic representations. We extend an existing sequence-to-tree model to a multi-task learning framework which shares the decoder for generating semantic representations. We report evaluation results on the multilingual GeoQuery corpus and introduce a new multilingual version of the ATIS corpus.

Session Session 3C: Speech 1 / Dialogue 1

Towards End-to-End Reinforcement Learning of Dialogue Agents for Information Access
Bhuwan Dhingra, Lihong Li, Xiujun Li, Jianfeng Gao, Yun-Nung Chen, Faisal Ahmed, and Li Deng 15:45–16:03

This paper proposes KB-InfoBot - a multi-turn dialogue agent which helps users search Knowledge Bases (KBs) without composing complicated queries. Such goal-oriented dialogue agents typically need to interact with an external database to access real-world knowledge. Previous systems achieved this by issuing a symbolic query to the KB to retrieve entries based on their attributes. However, such symbolic operations break the differentiability of the system and prevent end-to-end training of neural dialogue agents. In this paper, we address this limitation by replacing symbolic queries with an induced "soft" posterior distribution over the KB that indicates which entities the user is interested in. Integrating the soft retrieval process with a reinforcement learner leads to higher task success rate and reward in both simulations and against real users. We also present a fully neural end-to-end agent, trained entirely from user feedback, and discuss its application towards personalized dialogue agents.

Sequential Matching Network: A New Architecture for Multi-turn Response Selection in Retrieval-Based Chatbots
Yu Wu, Wei Wu, Chen Xing, Ming Zhou, and Zhoujun Li 16:04–16:22

We study response selection for multi-turn conversation in retrieval based chatbots. Existing work either concatenates utterances in context or matches a response with a highly abstract context vector finally, which may lose relationships among the utterances or important information in the context. We propose a sequential matching network (SMN) to address both problems. SMN first matches a response with each utterance in the context on multiple levels of granularity, and distills important matching information from each pair as a vector with convolution and pooling operations. The vectors are then accumulated in a chronological order through a recurrent neural network (RNN) which models relationships among the utterances. The final matching score is calculated with the hidden states of the RNN. Empirical study on two public data sets shows that SMN can significantly outperform state-of-the-art methods for response selection in multi-turn conversation.

Learning Word-Like Units from Joint Audio-Visual Analysis
David Harwath and James Glass 16:23–16:41

Given a collection of images and spoken audio captions, we present a method for discovering word-like acoustic units in the continuous speech signal and grounding them to semantically relevant image regions. For example, our model is able to detect spoken instances of the word 'lighthouse' within an utterance and associate them with image regions containing lighthouses. We do not use any form of conventional automatic speech recognition, nor do we use any text transcriptions or conventional linguistic annotations. Our model effectively implements a form of spoken language acquisition, in which the computer learns not only to recognize word categories by sound, but also to enrich the words it learns with semantics by grounding them in images.

Joint CTC/attention decoding for end-to-end speech recognition
Takaaki Hori, Shinji Watanabe, and John Hershey 16:42–17:00

End-to-end automatic speech recognition (ASR) has become a popular alternative to conventional DNN/HMM systems because it avoids the need for linguistic resources such as pronunciation dictionary, tokenization, and context-dependency trees, leading to a greatly simplified model-building process. There are two major types of end-to-end architectures for ASR: attention-based methods use an attention mechanism to perform alignment between acoustic frames and recognized symbols, and connectionist temporal classification (CTC), uses Markov assumptions to efficiently solve sequential problems by dynamic programming. This paper proposes joint decoding algorithm for end-to-end ASR with a hybrid CTC/attention architecture, which effectively utilizes both advantages in decoding. We have applied the proposed method to two ASR benchmarks (spontaneous Japanese and Mandarin Chinese), and showing the comparable performance to conventional state-of-the-art DNN/HMM ASR systems without linguistic resources.

Incorporating Uncertainty into Deep Learning for Spoken Language Assessment
Andrey Malinin, Anton Ragni, Kate Knill, and Mark Gales 17:01–17:12

There is a growing demand for automatic assessment of spoken English proficiency. These systems need to handle large variations in input data owing to the wide range of candidate skill levels and L1s, and errors from ASR. Some candidates will be a poor match to the training data set, undermining the validity of the predicted grade. For high stakes tests it is essential for such systems not only to grade well, but also to

provide a measure of their uncertainty in their predictions, enabling rejection to human graders. Previous work examined Gaussian Process (GP) graders which, though successful, do not scale well with large data sets. Deep Neural Network (DNN) may also be used to provide uncertainty using Monte-Carlo Dropout (MCD). This paper proposes a novel method to yield uncertainty and compares it to GPs and DNNs with MCD. The proposed approach explicitly teaches a DNN to have low uncertainty on training data and high uncertainty on generated artificial data. On experiments conducted on data from the Business Language Testing Service (BULATS), the proposed approach is found to outperform GPs and DNNs with MCD in uncertainty-based rejection whilst achieving comparable grading performance.

Session Session 3D: Multilingual 1

Found in Translation: Reconstructing Phylogenetic Language Trees from Translations
Ella Rabinovich, Noam Ordan, and Shuly Wintner 15:45–16:03

Translation has played an important role in trade, law, commerce, politics, and literature for thousands of years. Translators have always tried to be invisible; ideal translations should look as if they were written originally in the target language. We show that traces of the source language remain in the translation product to the extent that it is possible to uncover the history of the source language by looking only at the translation. Specifically, we automatically reconstruct phylogenetic language trees from monolingual texts (translated from several source languages). The signal of the source language is so powerful that it is retained even after two phases of translation. This strongly indicates that source language interference is the most dominant characteristic of translated texts, overshadowing the more subtle signals of universal properties of translation.

Predicting Native Language from Gaze
Yevgeni Berzak, Chie Nakamura, Suzanne Flynn, and Boris Katz 16:04–16:22

A fundamental question in language learning concerns the role of a speaker's first language in second language acquisition. We present a novel methodology for studying this question: analysis of eye-movement patterns in second language reading of free-form text. Using this methodology, we demonstrate for the first time that the native language of English learners can be predicted from their gaze fixations when reading English. We provide analysis of classifier uncertainty and learned features, which indicates that differences in English reading are likely to be rooted in linguistic divergences across native languages. The presented framework complements production studies and offers new ground for advancing research on multilingualism.

[TACL] Decoding Anagrammed Texts Written in an Unknown Language and Script
Bradley Hauer and Grzegorz Kondrak 16:23–16:41

Algorithmic decipherment is a prime example of a truly unsupervised problem. The first step in the decipherment process is the identification of the encrypted language. We propose three methods for determining the source language of a document enciphered with a monoalphabetic substitution cipher. The best method achieves 97% accuracy on 380 languages. We then present an approach to decoding anagrammed substitution ciphers, in which the letters within words have been arbitrarily transposed. It obtains the average decryption word accuracy of 93% on a set of 50 ciphertexts in 5 languages. Finally, we report the results on the Voynich manuscript, an unsolved fifteenth century cipher, which suggest Hebrew as the language of the document.

[TACL] Sparse Coding of Neural Word Embeddings for Multilingual Sequence Labeling
Gábor Berend 16:42–17:00

In this paper we propose and carefully evaluate a sequence labeling framework which solely utilizes sparse indicator features derived from dense distributed word representations. The proposed model obtains (near) state-of-the art performance for both part-of-speech tagging and named entity recognition for a variety of languages. Our model relies only on a few thousand sparse coding-derived features, without applying any modification of the word representations employed for the different tasks. The proposed model has favorable generalization properties as it retains over 89.8% of its average POS tagging accuracy when trained at 1.2% of the total available training data, i.e. 150 sentences per language.

Incorporating Dialectal Variability for Socially Equitable Language Identification
David Jurgens, Yulia Tsvetkov, and Dan Jurafsky 17:01–17:12

Language identification (LID) is a critical first step for processing multilingual text. Yet most LID systems are not designed to handle the linguistic diversity of global platforms like Twitter, where local dialects and rampant code-switching lead language classifiers to systematically miss minority dialect speakers and multilingual speakers. We propose a new dataset and a character-based sequence-to-sequence model for LID designed to support dialectal and multilingual language varieties. Our model achieves state-of-the-art performance on multiple LID benchmarks. Furthermore, in a case study using Twitter for health tracking, our method substantially increases the availability of texts written by underrepresented populations, enabling the development of "socially inclusive" NLP tools.

Session Session 3E: Phonology 1

MORSE: Semantic-ally Drive-n MORpheme SEgment-er
Tarek Sakakini, Suma Bhat, and Pramod Viswanath 15:45–16:03

We present in this paper a novel framework for morpheme segmentation which uses the morpho-syntactic regularities preserved by word representations, in addition to orthographic features, to segment words into morphemes. This framework is the first to consider vocabulary-wide syntactico-semantic information for this task. We also analyze the deficiencies of available benchmarking datasets and introduce our own dataset that was created on the basis of compositionality. We validate our algorithm across datasets and present state-of-the-art results.

[TACL] A Generative Model of Phonotactics
Richard Futrell, Adam Albright, Peter Graff, and Timothy J. O'Donnell 16:04–16:22

We present a probabilistic model of phonotactics, the set of well-formed phoneme sequences in a language. Unlike most computational models of phonotactics (Hayes and Wilson, 2008; Goldsmith and Riggle, 2012), we take a fully generative approach, modeling a process where forms are built up out of subparts by phonologically-informed structure building operations. We learn an inventory of subparts by applying stochastic memoization (Johnson et al., 2006; Goodman et al., 2008) to a generative process for phonemes structured as an and-or graph, based on concepts of feature hierarchy from generative phonology (Clements, 1985; Dresher, 2009). Subparts are combined in a way that allows tier-based feature interactions. We evaluate our models' ability to capture phonotactic distributions in the lexicons of 14 languages drawn from the WOLEX corpus (Graff, 2012). Our full model robustly assigns higher probabilities to held-out forms than a sophisticated N-gram model for all languages. We also present novel analyses that probe model behavior in more detail.

[TACL] Joint Semantic Synthesis and Morphological Analysis of the Derived Word
Ryan Cotterell and Hinrich Schütze 16:23–16:41

Much like sentences are composed of words, words themselves are composed of smaller units. For example, the English word questionably can be analyzed as question+able+ly. However, this structural decomposition of the word does not directly give us a semantic representation of the word's meaning. Since morphology obeys the principle of compositionality, the semantics of the word can be systematically derived from the meaning of its parts. In this work, we propose a novel probabilistic model of word formation that captures both the analysis of a word w into its constituents segments and the synthesis of the meaning of w from the meanings of those segments. Our model jointly learns to segment words into morphemes and compose distributional semantic vectors of those morphemes. We experiment with the model on English CELEX data and German DerivBase (Zeller et al., 2013) data. We show that jointly modeling semantics increases both segmentation accuracy and morpheme F1 by between 3% and 5%. Additionally, we investigate different models of vector composition, showing that recurrent neural networks yield an improvement over simple additive models. Finally, we study the degree to which the representations correspond to a linguist's notion of morphological productivity.

[TACL] Unsupervised Learning of Morphological Forests
Jiaming Luo and Karthik Narasimhan 16:42–17:00

This paper focuses on unsupervised modeling of morphological families, collectively comprising a forest over the language vocabulary. This formulation enables us to capture edgewise properties reflecting single-step morphological derivations, along with global distributional properties of the entire forest. These global properties constrain the size of the affix set and encourage formation of tight morphological families. The resulting objective is solved using Integer Linear Programming (ILP) paired with contrastive estimation. We train the model by alternating between optimizing the local log-linear model and the global ILP objective. We evaluate our system on three tasks: root detection, clustering of morphological families and segmentation. Our experiments demonstrate that our model yields consistent gains in all three tasks compared with the best published results.

Evaluating Compound Splitters Extrinsically with Textual Entailment
Glorianna Jagfeld, Patrick Ziering, and Lonneke van der Plas 17:01–17:12

Traditionally, compound splitters are evaluated intrinsically on gold-standard data or extrinsically on the task of statistical machine translation. We explore a novel way for the extrinsic evaluation of compound splitters, namely recognizing textual entailment. Compound splitting has great potential for this novel task that is both transparent and well-defined. Moreover, we show that it addresses certain aspects that are either ignored in intrinsic evaluations or compensated for by taskinternal mechanisms in statistical machine translation. We show significant improvements using different compound splitting methods on a German textual entailment

dataset.

Poster Session P1 (Long Papers)

Time: *18:00–21:30* Location: Bayshore Grand Ballroom/Foyer

Enriching Complex Networks with Word Embeddings for Detecting Mild Cognitive Impairment from Speech Transcripts
Leandro Santos, Edilson Anselmo Corrêa Júnior, Osvaldo Oliveira Jr, Diego Raphael Amancio, Letícia Mansur, and Sandra Aluísio

Mild Cognitive Impairment (MCI) is a mental disorder difficult to diagnose. Linguistic features, mainly from parsers, have been used to detect MCI, but this is not suitable for large-scale assessments. MCI disfluencies produce non-grammatical speech that requires manual or high precision automatic correction of transcripts. In this paper, we modeled transcripts into complex networks and enriched them with word embedding (CNE) to better represent short texts produced in neuropsychological assessments. The network measurements were applied with well-known classifiers to automatically identify MCI in transcripts, in a binary classification task. A comparison was made with the performance of traditional approaches using Bag of Words (BoW) and linguistic features for three datasets: DementiaBank in English, and Cinderella and Arizona-Battery in Portuguese. Overall, CNE provided higher accuracy than using only complex networks, while Support Vector Machine was superior to other classifiers. CNE provided the highest accuracies for DementiaBank and Cinderella, but BoW was more efficient for the Arizona-Battery dataset probably owing to its short narratives. The approach using linguistic features yielded higher accuracy if the transcriptions of the Cinderella dataset were manually revised. Taken together, the results indicate that complex networks enriched with embedding is promising for detecting MCI in large-scale assessments.

Adversarial Adaptation of Synthetic or Stale Data
Young-Bum Kim, Karl Stratos, and Dongchan Kim

Two types of data shift common in practice are 1. transferring from synthetic data to live user data (a deployment shift), and 2. transferring from stale data to current data (a temporal shift). Both cause a distribution mismatch between training and evaluation, leading to a model that overfits the flawed training data and performs poorly on the test data. We propose a solution to this mismatch problem by framing it as domain adaptation, treating the flawed training dataset as a source domain and the evaluation dataset as a target domain. To this end, we use and build on several recent advances in neural domain adaptation such as adversarial training (Ganinet al., 2016) and domain separation network (Bousmalis et al., 2016), proposing a new effective adversarial training scheme. In both supervised and unsupervised adaptation scenarios, our approach yields clear improvement over strong baselines.

Chat Detection in an Intelligent Assistant: Combining Task-oriented and Non-task-oriented Spoken Dialogue Systems
Satoshi Akasaki and Nobuhiro Kaji

Recently emerged intelligent assistants on smartphones and home electronics (e.g., Siri and Alexa) can be seen as novel hybrids of domain-specific task-oriented spoken dialogue systems and open-domain non-task-oriented ones. To realize such hybrid dialogue systems, this paper investigates determining whether or not a user is going to have a chat with the system. To address the lack of benchmark datasets for this task, we construct a new dataset consisting of 15,160 utterances collected from the real log data of a commercial intelligent assistant (and will release the dataset to facilitate future research activity). In addition, we investigate using tweets and Web search queries for handling open-domain user utterances, which characterize the task of chat detection. Experimental experiments demonstrated that, while simple supervised methods are effective, the use of the tweets and search queries further improves the F$_1$-score from 86.21 to 87.53.

A Neural Local Coherence Model
Dat Tien Nguyen and Shafiq Joty

We propose a local coherence model based on a convolutional neural network that operates over the entity grid representation of a text. The model captures long range en- tity transitions along with entity-specific features without loosing generalization, thanks to the power of distributed representation. We present a pairwise ranking method to train the model in an end-to-end fashion on a task and learn task-specific high level features. Our evaluation on three different coherence assessment tasks demonstrates that our model achieves state of the art results outperforming existing models by a good margin.

Data-Driven Broad-Coverage Grammars for Opinionated Natural Language Generation (ONLG)
Tomer Cagan, Stefan L. Frank, and Reut Tsarfaty

Opinionated Natural Language Generation (ONLG) is a new, challenging, task that aims to automatically generate human-like, subjective, responses to opinionated articles online. We present a data-driven architecture for ONLG that generates subjective responses triggered by users' agendas, consisting of topics and sentiments, and based on wide-coverage automatically-acquired generative grammars. We compare three types of grammatical representations that we design for ONLG, which interleave different layers of linguistic information and are induced from a new, enriched dataset we developed. Our evaluation shows that generation with Relational-Realizational (Tsarfaty and Sima'an, 2008) inspired grammar gets better language model scores than lexicalized grammars 'a la Collins (2003), and that the latter gets better human-evaluation scores. We also show that conditioning the generation on topic models makes generated responses more relevant to the document content.

Learning to Ask: Neural Question Generation for Reading Comprehension
Xinya Du, Junru Shao, and Claire Cardie

We study automatic question generation for sentences from text passages in reading comprehension. We introduce an attention-based sequence learning model for the task and investigate the effect of encoding sentence- vs. paragraph-level information. In contrast to all previous work, our model does not rely on hand-crafted rules or a sophisticated NLP pipeline; it is instead trainable end-to-end via sequence-to-sequence learning. Automatic evaluation results show that our system significantly outperforms the state-of-the-art rule-based system. In human evaluations, questions generated by our system are also rated as being more natural (i.e., grammaticality, fluency) and as more difficult to answer (in terms of syntactic and lexical divergence from the original text and reasoning needed to answer).

Joint Optimization of User-desired Content in Multi-document Summaries by Learning from User Feedback
Avinesh PVS and Christian M. Meyer

In this paper, we propose an extractive multi-document summarization (MDS) system using joint optimization and active learning for content selection grounded in user feedback. Our method interactively obtains user feedback to gradually improve the results of a state-of-the-art integer linear programming (ILP) framework for MDS. Our methods complement fully automatic methods in producing high-quality summaries with a minimum number of iterations and feedbacks. We conduct multiple simulation-based experiments and analyze the effect of feedback-based concept selection in the ILP setup in order to maximize the user-desired content in the summary.

Flexible and Creative Chinese Poetry Generation Using Neural Memory
Jiyuan Zhang, Yang Feng, Dong Wang, Yang Wang, Andrew Abel, Shiyue Zhang, and Andi Zhang

It has been shown that Chinese poems can be successfully generated by sequence-to-sequence neural models, particularly with the attention mechanism. A potential problem of this approach, however, is that neural models can only learn abstract rules, while poem generation is a highly creative process that involves not only rules but also innovations for which pure statistical models are not appropriate in principle. This work proposes a memory augmented neural model for Chinese poem generation, where the neural model and the augmented memory work together to balance the requirements of linguistic accordance and aesthetic innovation, leading to innovative generations that are still rule-compliant. In addition, it is found that the memory mechanism provides interesting flexibility that can be used to generate poems with different styles.

Learning to Generate Market Comments from Stock Prices
Soichiro Murakami, Akihiko Watanabe, Akira Miyazawa, Keiichi Goshima, Toshihiko Yanase, Hiroya Takamura, and Yusuke Miyao

This paper presents a novel encoder-decoder model for automatically generating market comments from stock prices. The model first encodes both short- and long-term series of stock prices so that it can mention short- and long-term changes in stock prices. In the decoding phase, our model can also generate a numerical value by selecting an appropriate arithmetic operation such as subtraction or rounding, and applying it to the input stock prices. Empirical experiments show that our best model generates market comments at the fluency and the informativeness approaching human-generated reference texts.

Can Syntax Help? Improving an LSTM-based Sentence Compression Model for New Domains
Liangguo Wang, Jing Jiang, Hai Leong Chieu, Chen Hui Ong, Dandan Song, and Lejian Liao

In this paper, we study how to improve the domain adaptability of a deletion-based Long Short-Term Memory (LSTM) neural network model for sentence compression. We hypothesize that syntactic information helps in making such models more robust across domains. We propose two major changes to the model: using

explicit syntactic features and introducing syntactic constraints through Integer Linear Programming (ILP). Our evaluation shows that the proposed model works better than the original model as well as a traditional non-neural-network-based model in a cross-domain setting.

Transductive Non-linear Learning for Chinese Hypernym Prediction
Chengyu Wang, Junchi Yan, Aoying Zhou, and Xiaofeng He

Finding the correct hypernyms for entities is essential for taxonomy learning, fine-grained entity categorization, query understanding, etc. Due to the flexibility of the Chinese language, it is challenging to identify hypernyms in Chinese accurately. Rather than extracting hypernyms from texts, in this paper, we present a transductive learning approach to establish mappings from entities to hypernyms in the embedding space directly. It combines linear and non-linear embedding projection models, with the capacity of encoding arbitrary language-specific rules. Experiments on real-world datasets illustrate that our approach outperforms previous methods for Chinese hypernym prediction.

A Constituent-Centric Neural Architecture for Reading Comprehension
Pengtao Xie and Eric Xing

Reading comprehension (RC), aiming to understand natural texts and answer questions therein, is a challenging task. In this paper, we study the RC problem on the Stanford Question Answering Dataset (SQuAD). Observing from the training set that most correct answers are centered around constituents in the parse tree, we design a constituent-centric neural architecture where the generation of candidate answers and their representation learning are both based on constituents and guided by the parse tree. Under this architecture, the search space of candidate answers can be greatly reduced without sacrificing the coverage of correct answers and the syntactic, hierarchical and compositional structure among constituents can be well captured, which contributes to better representation learning of the candidate answers. On SQuAD, our method achieves the state of the art performance and the ablation study corroborates the effectiveness of individual modules.

Cross-lingual Distillation for Text Classification
Ruochen Xu and Yiming Yang

Cross-lingual text classification(CLTC) is the task of classifying documents written in different languages into the same taxonomy of categories. This paper presents a novel approach to CLTC that builds on model distillation, which adapts and extends a framework originally proposed for model compression. Using soft probabilistic predictions for the documents in a label-rich language as the (induced) supervisory labels in a parallel corpus of documents, we train classifiers successfully for new languages in which labeled training data are not available. An adversarial feature adaptation technique is also applied during the model training to reduce distribution mismatch. We conducted experiments on two benchmark CLTC datasets, treating English as the source language and German, French, Japan and Chinese as the unlabeled target languages. The proposed approach had the advantageous or comparable performance of the other state-of-art methods.

Understanding and Predicting Empathic Behavior in Counseling Therapy
Verónica Pérez-Rosas, Rada Mihalcea, Kenneth Resnicow, Satinder Singh, and Lawrence An

Counselor empathy is associated with better outcomes in psychology and behavioral counseling. In this paper, we explore several aspects pertaining to counseling interaction dynamics and their relation to counselor empathy during motivational interviewing encounters. Particularly, we analyze aspects such as participants' engagement, participants' verbal and nonverbal accommodation, as well as topics being discussed during the conversation, with the final goal of identifying linguistic and acoustic markers of counselor empathy. We also show how we can use these findings alongside other raw linguistic and acoustic features to build accurate counselor empathy classifiers with accuracies of up to 80%.

Leveraging Knowledge Bases in LSTMs for Improving Machine Reading
Bishan Yang and Tom M. Mitchell

This paper focuses on how to take advantage of external knowledge bases (KBs) to improve recurrent neural networks for machine reading. Traditional methods that exploit knowledge from KBs encode knowledge as discrete indicator features. Not only do these features generalize poorly, but they require task-specific feature engineering to achieve good performance. We propose KBLSTM, a novel neural model that leverages continuous representations of KBs to enhance the learning of recurrent neural networks for machine reading. To effectively integrate background knowledge with information from the currently processed text, our model employs an attention mechanism with a sentinel to adaptively decide whether to attend to background knowledge and which information from KBs is useful. Experimental results show that our model achieves accuracies that surpass the previous state-of-the-art results for both entity extraction and event extraction on the widely used

ACE2005 dataset.

Prerequisite Relation Learning for Concepts in MOOCs
Liangming Pan, Chengjiang Li, Juanzi Li, and Jie Tang

What prerequisite knowledge should students achieve a level of mastery before moving forward to learn subsequent coursewares? We study the extent to which the prerequisite relation between knowledge concepts in Massive Open Online Courses (MOOCs) can be inferred automatically. In particular, what kinds of information can be leverage to uncover the potential prerequisite relation between knowledge concepts. We first propose a representation learning-based method for learning latent representations of course concepts, and then investigate how different features capture the prerequisite relations between concepts. Our experiments on three datasets form Coursera show that the proposed method achieves significant improvements (+5.9-48.0% by F1-score) comparing with existing methods.

Unsupervised Text Segmentation Based on Native Language Characteristics
Shervin Malmasi, Mark Dras, Mark Johnson, Lan Du, and Magdalena Wolska

Most work on segmenting text does so on the basis of topic changes, but it can be of interest to segment by other, stylistically expressed characteristics such as change of authorship or native language. We propose a Bayesian unsupervised text segmentation approach to the latter. While baseline models achieve essentially random segmentation on our task, indicating its difficulty, a Bayesian model that incorporates appropriately compact language models and alternating asymmetric priors can achieve scores on the standard metrics around halfway to perfect segmentation.

Weakly Supervised Cross-Lingual Named Entity Recognition via Effective Annotation and Representation Projection
Jian Ni, Georgiana Dinu, and Radu Florian

The state-of-the-art named entity recognition (NER) systems are supervised machine learning models that require large amounts of manually annotated data to achieve high accuracy. However, annotating NER data by human is expensive and time-consuming, and can be quite difficult for a new language. In this paper, we present two weakly supervised approaches for cross-lingual NER with no human annotation in a target language. The first approach is to create automatically labeled NER data for a target language via annotation projection on comparable corpora, where we develop a heuristic scheme that effectively selects good-quality projection-labeled data from noisy data. The second approach is to project distributed representations of words (word embeddings) from a target language to a source language, so that the source-language NER system can be applied to the target language without re-training. We also design two co-decoding schemes that effectively combine the outputs of the two projection-based approaches. We evaluate the performance of the proposed approaches on both in-house and open NER data for several target languages. The results show that the combined systems outperform three other weakly supervised approaches on the CoNLL data.

Context Sensitive Lemmatization Using Two Successive Bidirectional Gated Recurrent Networks
Abhisek Chakrabarty, Onkar Arun Pandit, and Utpal Garain

We introduce a composite deep neural network architecture for supervised and language independent context sensitive lemmatization. The proposed method considers the task as to identify the correct edit tree representing the transformation between a word-lemma pair. To find the lemma of a surface word, we exploit two successive bidirectional gated recurrent structures - the first one is used to extract the character level dependencies and the next one captures the contextual information of the given word. The key advantages of our model compared to the state-of-the-art lemmatizers such as Lemming and Morfette are - (i) it is independent of human decided features (ii) except the gold lemma, no other expensive morphological attribute is required for joint learning. We evaluate the lemmatizer on nine languages - Bengali, Catalan, Dutch, Hindi, Hungarian, Italian, Latin, Romanian and Spanish. It is found that except Bengali, the proposed method outperforms Lemming and Morfette on the other languages. To train the model on Bengali, we develop a gold lemma annotated dataset (having 1,702 sentences with a total of 20,257 word tokens), which is an additional contribution of this work.

Learning to Create and Reuse Words in Open-Vocabulary Neural Language Modeling
Kazuya Kawakami, Chris Dyer, and Phil Blunsom

Fixed-vocabulary language models fail to account for one of the most characteristic statistical facts of natural language: the frequent creation and reuse of new word types. Although character-level language models offer a partial solution in that they can create word types not attested in the training corpus, they do not capture the "bursty" distribution of such words. In this paper, we augment a hierarchical LSTM language model that generates sequences of word tokens character by character with a caching mechanism that learns to reuse

previously generated words. To validate our model we construct a new open-vocabulary language modeling corpus (the Multilingual Wikipedia Corpus; MWC) from comparable Wikipedia articles in 7 typologically diverse languages and demonstrate the effectiveness of our model across this range of languages.

Bandit Structured Prediction for Neural Sequence-to-Sequence Learning
Julia Kreutzer, Artem Sokolov, and Stefan Riezler

Bandit structured prediction describes a stochastic optimization framework where learning is performed from partial feedback. This feedback is received in the form of a task loss evaluation to a predicted output structure, without having access to gold standard structures. We advance this framework by lifting linear bandit learning to neural sequence-to-sequence learning problems using attention-based recurrent neural networks. Furthermore, we show how to incorporate control variates into our learning algorithms for variance reduction and improved generalization. We present an evaluation on a neural machine translation task that shows improvements of up to 5.89 BLEU points for domain adaptation from simulated bandit feedback.

Prior Knowledge Integration for Neural Machine Translation using Posterior Regularization
Jiacheng Zhang, Yang Liu, Huanbo Luan, Jingfang Xu, and Maosong Sun

Although neural machine translation has made significant progress recently, how to integrate multiple overlapping, arbitrary prior knowledge sources remains a challenge. In this work, we propose to use posterior regularization to provide a general framework for integrating prior knowledge into neural machine translation. We represent prior knowledge sources as features in a log-linear model, which guides the learning processing of the neural translation model. Experiments on Chinese-English dataset show that our approach leads to significant improvements.

Incorporating Word Reordering Knowledge into Attention-based Neural Machine Translation
Jinchao Zhang, Mingxuan Wang, Qun Liu, and Jie Zhou

This paper proposes three distortion models to explicitly incorporate the word reordering knowledge into attention-based Neural Machine Translation (NMT) for further improving translation performance. Our proposed models enable attention mechanism to attend to source words regarding both the semantic requirement and the word reordering penalty. Experiments on Chinese-English translation show that the approaches can improve word alignment quality and achieve significant translation improvements over a basic attention-based NMT by large margins. Compared with previous works on identical corpora, our system achieves the state-of-the-art performance on translation quality.

Lexically Constrained Decoding for Sequence Generation Using Grid Beam Search
Chris Hokamp and Qun Liu

We present Grid Beam Search (GBS), an algorithm which extends beam search to allow the inclusion of pre-specified lexical constraints. The algorithm can be used with any model which generates sequences token by token. Lexical constraints take the form of phrases or words that must be present in the output sequence. This is a very general way to incorporate auxillary knowledge into a model's output without requiring any modification of the parameters or training data. We demonstrate the feasibility and flexibility of Lexically Constrained Decoding by conducting experiments on Neural Interactive-Predictive Translation, as well as Domain Adaptation for Neural Machine Translation. Experiments show that GBS can provide large improvements in translation quality in interactive scenarios, and that, even without any user input, GBS can be used to achieve significant gains in performance in domain adaptation scenarios.

Combating Human Trafficking with Multimodal Deep Models
Edmund Tong, Amir Zadeh, Cara Jones, and Louis-Philippe Morency

Human trafficking is a global epidemic affecting millions of people across the planet. Sex trafficking, the dominant form of human trafficking, has seen a significant rise mostly due to the abundance of escort websites, where human traffickers can openly advertise among at-will escort advertisements. In this paper, we take a major step in the automatic detection of advertisements suspected to pertain to human trafficking. We present a novel dataset called Trafficking-10k, with more than 10,000~advertisements annotated for this task. The dataset contains two sources of information per advertisement: text and images. For the accurate detection of trafficking advertisements, we designed and trained a deep multimodal model called the Human Trafficking Deep Network (HTDN).

MalwareTextDB: A Database for Annotated Malware Articles
Swee Kiat Lim, Aldrian Obaja Muis, Wei Lu, and Chen Hui Ong

Cybersecurity risks and malware threats are becoming increasingly dangerous and common. Despite the severity of the problem, there has been few NLP efforts focused on tackling cybersecurity. In this paper, we discuss the construction of a new database for annotated malware texts. An annotation framework is introduced based on the MAEC vocabulary for defining malware characteristics, along with a database consisting of 39 annotated APT reports with a total of 6,819 sentences. We also use the database to construct models that can potentially help cybersecurity researchers in their data collection and analytics efforts.

A Corpus of Annotated Revisions for Studying Argumentative Writing
Fan Zhang, Homa B. Hashemi, Rebecca Hwa, and Diane Litman

This paper presents ArgRewrite, a corpus of between-draft revisions of argumentative essays. Drafts are manually aligned at the sentence level, and the writer's purpose for each revision is annotated with categories analogous to those used in argument mining and discourse analysis. The corpus should enable advanced research in writing comparison and revision analysis, as demonstrated via our own studies of student revision behavior and of automatic revision purpose prediction.

Automatic Induction of Synsets from a Graph of Synonyms
Dmitry Ustalov, Alexander Panchenko, and Chris Biemann

This paper presents a new graph-based approach that induces synsets using synonymy dictionaries and word embeddings. First, we build a weighted graph of synonyms extracted from commonly available resources, such as Wiktionary. Second, we apply word sense induction to deal with ambiguous words. Finally, we cluster the disambiguated version of the ambiguous input graph into synsets. Our meta-clustering approach lets us use an efficient hard clustering algorithm to perform a fuzzy clustering of the graph. Despite its simplicity, our approach shows excellent results, outperforming five competitive state-of-the-art methods in terms of F-score on three gold standard datasets for English and Russian derived from large-scale manually constructed lexical resources.

Neural Modeling of Multi-Predicate Interactions for Japanese Predicate Argument Structure Analysis
Hiroki Ouchi, Hiroyuki Shindo, and Yuji Matsumoto

The performance of Japanese predicate argument structure (PAS) analysis has improved in recent years thanks to the joint modeling of interactions between multiple predicates. However, this approach relies heavily on syntactic information predicted by parsers, and suffers from errorpropagation. To remedy this problem, we introduce a model that uses grid-type recurrent neural networks. The proposed model automatically induces features sensitive to multi-predicate interactions from the word sequence information of a sentence. Experiments on the NAIST Text Corpus demonstrate that without syntactic information, our model outperforms previous syntax-dependent models.

TriviaQA: A Large Scale Distantly Supervised Challenge Dataset for Reading Comprehension
Mandar Joshi, Eunsol Choi, Daniel Weld, and Luke Zettlemoyer

We present TriviaQA, a challenging reading comprehension dataset containing over 650K question-answer-evidence triples. TriviaQA includes 95K question-answer pairs authored by trivia enthusiasts and independently gathered evidence documents, six per question on average, that provide high quality distant supervision for answering the questions. We show that, in comparison to other recently introduced large-scale datasets, TriviaQA (1) has relatively complex, compositional questions, (2) has considerable syntactic and lexical variability between questions and corresponding answer-evidence sentences, and (3) requires more cross sentence reasoning to find answers. We also present two baseline algorithms: a feature-based classifier and a state-of-the-art neural network, that performs well on SQuAD reading comprehension. Neither approach comes close to human performance (23% and 40% vs. 80%), suggesting that TriviaQA is a challenging testbed that is worth significant future study.

Learning Semantic Correspondences in Technical Documentation
Kyle Richardson and Jonas Kuhn

We consider the problem of translating high-level textual descriptions to formal representations in technical documentation as part of an effort to model the meaning of such documentation. We focus specifically on the problem of learning translational correspondences between text descriptions and grounded representations in the target documentation, such as formal representation of functions or code templates. Our approach exploits the parallel nature of such documentation, or the tight coupling between high-level text and the low-level representations we aim to learn. Data is collected by mining technical documents for such parallel text-representation pairs, which we use to train a simple semantic parsing model. We report new baseline

results on sixteen novel datasets, including the standard library documentation for nine popular programming languages across seven natural languages, and a small collection of Unix utility manuals.

Bridge Text and Knowledge by Learning Multi-Prototype Entity Mention Embedding
Yixin Cao, Lifu Huang, Heng Ji, Xu Chen, and Juanzi Li

Integrating text and knowledge into a unified semantic space has attracted significant research interests recently. However, the ambiguity in the common space remains a challenge, namely that the same mention phrase usually refers to various entities. In this paper, to deal with the ambiguity of entity mentions, we propose a novel Multi-Prototype Mention Embedding model, which learns multiple sense embeddings for each mention by jointly modeling words from textual contexts and entities derived from a knowledge base. In addition, we further design an efficient language model based approach to disambiguate each mention to a specific sense. In experiments, both qualitative and quantitative analysis demonstrate the high quality of the word, entity and multi-prototype mention embeddings. Using entity linking as a study case, we apply our disambiguation method as well as the multi-prototype mention embeddings on the benchmark dataset, and achieve state-of-the-art performance.

Interactive Learning of Grounded Verb Semantics towards Human-Robot Communication
Lanbo She and Joyce Chai

To enable human-robot communication and collaboration, previous works represent grounded verb semantics as the potential change of state to the physical world caused by these verbs. Grounded verb semantics are acquired mainly based on the parallel data of the use of a verb phrase and its corresponding sequences of primitive actions demonstrated by humans. The rich interaction between teachers and students that is considered important in learning new skills has not yet been explored. To address this limitation, this paper presents a new interactive learning approach that allows robots to proactively engage in interaction with human partners by asking good questions to learn models for grounded verb semantics. The proposed approach uses reinforcement learning to allow the robot to acquire an optimal policy for its question-asking behaviors by maximizing the long-term reward. Our empirical results have shown that the interactive learning approach leads to more reliable models for grounded verb semantics, especially in the noisy environment which is full of uncertainties. Compared to previous work, the models acquired from interactive learning result in a 48% to 145% performance gain when applied in new situations.

Multimodal Word Distributions
Ben Athiwaratkun and Andrew Wilson

Word embeddings provide point representations of words containing useful semantic information. We introduce multimodal word distributions formed from Gaussian mixtures, for multiple word meanings, entailment, and rich uncertainty information. To learn these distributions, we propose an energy-based max-margin objective. We show that the resulting approach captures uniquely expressive semantic information, and outperforms alternatives, such as word2vec skip-grams, and Gaussian embeddings, on benchmark datasets such as word similarity and entailment.

Enhanced LSTM for Natural Language Inference
Qian Chen, Xiaodan Zhu, Zhen-Hua Ling, Si Wei, Hui Jiang, and Diana Inkpen

Reasoning and inference are central to human and artificial intelligence. Modeling inference in human language is very challenging. With the availability of large annotated data (Bowman et al., 2015), it has recently become feasible to train neural network based inference models, which have shown to be very effective. In this paper, we present a new state-of-the-art result, achieving the accuracy of 88.6% on the Stanford Natural Language Inference Dataset. Unlike the previous top models that use very complicated network architectures, we first demonstrate that carefully designing sequential inference models based on chain LSTMs can outperform all previous models. Based on this, we further show that by explicitly considering recursive architectures in both local inference modeling and inference composition, we achieve additional improvement. Particularly, incorporating syntactic parsing information contributes to our best result—it further improves the performance even when added to the already very strong model.

Linguistic analysis of differences in portrayal of movie characters
Anil Ramakrishna, Victor R. Martínez, Nikolaos Malandrakis, Karan Singla, and Shrikanth Narayanan

We examine differences in portrayal of characters in movies using psycholinguistic and graph theoretic measures computed directly from screenplays. Differences are examined with respect to characters' gender, race, age and other metadata. Psycholinguistic metrics are extrapolated to dialogues in movies using a linear regres-

sion model built on a set of manually annotated seed words. Interesting patterns are revealed about relationships between genders of production team and the gender ratio of characters. Several correlations are noted between gender, race, age of characters and the linguistic metrics.

Linguistically Regularized LSTM for Sentiment Classification
Qiao Qian, Minlie Huang, Jinhao Lei, and Xiaoyan Zhu

This paper deals with sentence-level sentiment classification. Though a variety of neural network models have been proposed recently, however, previous models either depend on expensive phrase-level annotation, most of which has remarkably degraded performance when trained with only sentence-level annotation; or do not fully employ linguistic resources (e.g., sentiment lexicons, negation words, intensity words). In this paper, we propose simple models trained with sentence-level annotation, but also attempt to model the linguistic role of sentiment lexicons, negation words, and intensity words. Results show that our models are able to capture the linguistic role of sentiment words, negation words, and intensity words in sentiment expression.

Sarcasm SIGN: Interpreting Sarcasm with Sentiment Based Monolingual Machine Translation
Lotem Peled and Roi Reichart

Sarcasm is a form of speech in which speakers say the opposite of what they truly mean in order to convey a strong sentiment. In other words, "Sarcasm is the giant chasm between what I say, and the person who doesn't get it.". In this paper we present the novel task of sarcasm interpretation, defined as the generation of a non-sarcastic utterance conveying the same message as the original sarcastic one. We introduce a novel dataset of 3000 sarcastic tweets, each interpreted by five human judges. Addressing the task as monolingual machine translation (MT), we experiment with MT algorithms and evaluation measures. We then present SIGN: an MT based sarcasm interpretation algorithm that targets sentiment words, a defining element of textual sarcasm. We show that while the scores of n-gram based automatic measures are similar for all interpretation models, SIGN's interpretations are scored higher by humans for adequacy and sentiment polarity. We conclude with a discussion on future research directions for our new task.

Active Sentiment Domain Adaptation
Fangzhao Wu, Yongfeng Huang, and Jun Yan

Domain adaptation is an important technology to handle domain dependence problem in sentiment analysis field. Existing methods usually rely on sentiment classifiers trained in source domains. However, their performance may heavily decline if the distributions of sentiment features in source and target domains have significant difference. In this paper, we propose an active sentiment domain adaptation approach to handle this problem. Instead of the source domain sentiment classifiers, our approach adapts the general-purpose sentiment lexicons to target domain with the help of a small number of labeled samples which are selected and annotated in an active learning mode, as well as the domain-specific sentiment similarities among words mined from unlabeled samples of target domain. A unified model is proposed to fuse different types of sentiment information and train sentiment classifier for target domain. Extensive experiments on benchmark datasets show that our approach can train accurate sentiment classifier with less labeled samples.

Volatility Prediction using Financial Disclosures Sentiments with Word Embedding-based IR Models
Navid Rekabsaz, Mihai Lupu, Artem Baklanov, Alexander Dür, Linda Andersson, and Allan Hanbury

Volatility prediction—an essential concept in financial markets—has recently been addressed using sentiment analysis methods. We investigate the sentiment of annual disclosures of companies in stock markets to forecast volatility. We specifically explore the use of recent Information Retrieval (IR) term weighting models that are effectively extended by related terms using word embeddings. In parallel to textual information, factual market data have been widely used as the mainstream approach to forecast market risk. We therefore study different fusion methods to combine text and market data resources. Our word embedding-based approach significantly outperforms state-of-the-art methods. In addition, we investigate the characteristics of the reports of the companies in different financial sectors.

CANE: Context-Aware Network Embedding for Relation Modeling
Cunchao Tu, Han Liu, Zhiyuan Liu, and Maosong Sun

Network embedding (NE) is playing a critical role in network analysis, due to its ability to represent vertices with efficient low-dimensional embedding vectors. However, existing NE models aim to learn a fixed context-free embedding for each vertex and neglect the diverse roles when interacting with other vertices. In this paper, we assume that one vertex usually shows different aspects when interacting with different neighbor

vertices, and should own different embeddings respectively. Therefore, we present Context-Aware Network Embedding (CANE), a novel NE model to address this issue. CANE learns context-aware embeddings for vertices with mutual attention mechanism and is expected to model the semantic relationships between vertices more precisely. In experiments, we compare our model with existing NE models on three real-world datasets. Experimental results show that CANE achieves significant improvement than state-of-the-art methods on link prediction and comparable performance on vertex classification. The source code and datasets can be obtained from `https://github.com/thunlp/CANE`.

Universal Dependencies Parsing for Colloquial Singaporean English
Hongmin Wang, Yue Zhang, GuangYong Leonard Chan, Jie Yang, and Hai Leong Chieu

Singlish can be interesting to the ACL community both linguistically as a major creole based on English, and computationally for information extraction and sentiment analysis of regional social media. We investigate dependency parsing of Singlish by constructing a dependency treebank under the Universal Dependencies scheme, and then training a neural network model by integrating English syntactic knowledge into a state-of-the-art parser trained on the Singlish treebank. Results show that English knowledge can lead to 25% relative error reduction, resulting in a parser of 84.47% accuracies. To the best of our knowledge, we are the first to use neural stacking to improve cross-lingual dependency parsing on low-resource languages. We make both our annotation and parser available for further research.

Generic Axiomatization of Families of Noncrossing Graphs in Dependency Parsing
Anssi Yli-Jyrä and Carlos Gómez-Rodríguez

We present a simple encoding for unlabeled noncrossing graphs and show how its latent counterpart helps us to represent several families of directed and undirected graphs used in syntactic and semantic parsing of natural language as context-free languages. The families are separated purely on the basis of forbidden patterns in latent encoding, eliminating the need to differentiate the families of non-crossing graphs in inference algorithms: one algorithm works for all when the search space can be controlled in parser input.

Semi-supervised sequence tagging with bidirectional language models
Matthew Peters, Waleed Ammar, Chandra Bhagavatula, and Russell Power

Pre-trained word embeddings learned from unlabeled text have become a stan- dard component of neural network archi- tectures for NLP tasks. However, in most cases, the recurrent network that oper- ates on word-level representations to pro- duce context sensitive representations is trained on relatively little labeled data. In this paper, we demonstrate a general semi-supervised approach for adding pre- trained context embeddings from bidi- rectional language models to NLP sys- tems and apply it to sequence labeling tasks. We evaluate our model on two stan- dard datasets for named entity recognition (NER) and chunking, and in both cases achieve state of the art results, surpassing previous systems that use other forms of transfer or joint learning with additional labeled data and task specific gazetteers.

Poster Session P1 (Short Papers)

Time: *18:00–21:30* Location: Bayshore Grand Ballroom/Foyer

Neural Architecture for Temporal Relation Extraction: A Bi-LSTM Approach for Detecting Narrative Containers
Julien Tourille, Olivier Ferret, Aurelie Neveol, and Xavier Tannier

We present a neural architecture for containment relation identification between medical events and/or temporal expressions. We experiment on a corpus of de-identified clinical notes in English from the Mayo Clinic, namely the THYME corpus. Our model achieves an F-measure of 0.613 and outperforms the best result reported on this corpus to date.

How to Make Context More Useful? An Empirical Study on Context-Aware Neural Conversational Models
Zhiliang Tian, Rui Yan, Lili Mou, Yiping Song, Yansong Feng, and Dongyan Zhao

Generative conversational systems are attracting increasing attention in natural language processing (NLP). Recently, researchers have noticed the importance of context information in dialog processing, and built various models to utilize context. However, there is no systematic comparison to analyze how to use context effectively. In this paper, we conduct an empirical study to compare various models and investigate the effect of context information in dialog systems. We also propose a variant that explicitly weights context vectors by

context-query relevance, outperforming the other baselines.

Cross-lingual and cross-domain discourse segmentation of entire documents
Chloé Braud, Ophélie Lacroix, and Anders Søgaard

Discourse segmentation is a crucial step in building end-to-end discourse parsers. However, discourse segmenters only exist for a few languages and domains. Typically they only detect intra-sentential segment boundaries, assuming gold standard sentence and token segmentation, and relying on high-quality syntactic parses and rich heuristics that are not generally available across languages and domains. In this paper, we propose statistical discourse segmenters for five languages and three domains that do not rely on gold pre-annotations. We also consider the problem of learning discourse segmenters when no labeled data is available for a language. Our fully supervised system obtains 89.5% F1 for English newswire, with slight drops in performance on other domains, and we report supervised and unsupervised (cross-lingual) results for five languages in total.

Detecting Good Arguments in a Non-Topic-Specific Way: An Oxymoron?
Beata Beigman Klebanov, Binod Gyawali, and Yi Song

Automatic identification of good arguments on a controversial topic has applications in civics and education, to name a few. While in the civics context it might be acceptable to create separate models for each topic, in the context of scoring of students' writing there is a preference for a single model that applies to all responses. Given that good arguments for one topic are likely to be irrelevant for another, is a single model for detecting good arguments a contradiction in terms? We investigate the extent to which it is possible to close the performance gap between topic-specific and across-topics models for identification of good arguments.

Argumentation Quality Assessment: Theory vs. Practice
Henning Wachsmuth, Nona Naderi, Ivan Habernal, Yufang Hou, Graeme Hirst, Iryna Gurevych, and Benno Stein

Argumentation quality is viewed differently in argumentation theory and in practical assessment approaches. This paper studies to what extent the views match empirically. We find that most observations on quality phrased spontaneously are in fact adequately represented by theory. Even more, relative comparisons of arguments in practice correlate with absolute quality ratings based on theory. Our results clarify how the two views can learn from each other.

A Recurrent Neural Model with Attention for the Recognition of Chinese Implicit Discourse Relations
Samuel Rönnqvist, Niko Schenk, and Christian Chiarcos

We introduce an attention-based Bi-LSTM for Chinese implicit discourse relations and demonstrate that modeling argument pairs as a joint sequence can outperform word order-agnostic approaches. Our model benefits from a partial sampling scheme and is conceptually simple, yet achieves state-of-the-art performance on the Chinese Discourse Treebank. We also visualize its attention activity to illustrate the model's ability to selectively focus on the relevant parts of an input sequence.

Discourse Annotation of Non-native Spontaneous Spoken Responses Using the Rhetorical Structure Theory Framework
Xinhao Wang, James Bruno, Hillary Molloy, Keelan Evanini, and Klaus Zechner

The availability of the Rhetorical Structure Theory (RST) Discourse Treebank has spurred substantial research into discourse analysis of written texts; however, limited research has been conducted to date on RST annotation and parsing of spoken language, in particular, non-native spontaneous speech. Considering that the measurement of discourse coherence is typically a key metric in human scoring rubrics for assessments of spoken language, we initiated a research effort to obtain RST annotations of a large number of non-native spoken responses from a standardized assessment of academic English proficiency. The resulting inter-annotator kappa agreements on the three different levels of Span, Nuclearity, and Relation are 0.848, 0.766, and 0.653, respectively. Furthermore, a set of features was explored to evaluate the discourse structure of non-native spontaneous speech based on these annotations; the highest performing feature resulted in a correlation of 0.612 with scores of discourse coherence provided by expert human raters.

Improving Implicit Discourse Relation Recognition with Discourse-specific Word Embeddings
Changxing Wu, Xiaodong Shi, Yidong Chen, Jinsong Su, and Boli Wang

We introduce a simple and effective method to learn discourse-specific word embeddings (DSWE) for implicit discourse relation recognition. Specifically, DSWE is learned by performing connective classification on massive explicit discourse data, and capable of capturing discourse relationships between words. On the PDTB

data set, using DSWE as features achieves significant improvements over baselines.

Oracle Summaries of Compressive Summarization
Tsutomu Hirao, Masaaki Nishino, and Masaaki Nagata

This paper derives an Integer Linear Programming (ILP) formulation to obtain an oracle summary of the compressive summarization paradigm in terms of ROUGE. The oracle summary is essential to reveal the upper bound performance of the paradigm. Experimental results on the DUC dataset showed that ROUGE scores of compressive oracles are significantly higher than those of extractive oracles and state-of-the-art summarization systems. These results reveal that compressive summarization is a promising paradigm and encourage us to continue with the research to produce informative summaries.

Japanese Sentence Compression with a Large Training Dataset
Shun Hasegawa, Yuta Kikuchi, Hiroya Takamura, and Manabu Okumura

In English, high-quality sentence compression models by deleting words have been trained on automatically created large training datasets. We work on Japanese sentence compression by a similar approach. To create a large Japanese training dataset, a method of creating English training dataset is modified based on the characteristics of the Japanese language. The created dataset is used to train Japanese sentence compression models based on the recurrent neural network.

A Neural Architecture for Generating Natural Language Descriptions from Source Code Changes
Pablo Loyola, Edison Marrese-Taylor, and Yutaka Matsuo

We propose a model to automatically describe changes introduced in the source code of a program using natural language. Our method receives as input a set of code commits, which contains both the modifications and message introduced by an user. These two modalities are used to train an encoder-decoder architecture. We evaluated our approach on twelve real world open source projects from four different programming languages. Quantitative and qualitative results showed that the proposed approach can generate feasible and semantically sound descriptions not only in standard in-project settings, but also in a cross-project setting.

English Event Detection With Translated Language Features
Sam Wei, Igor Korostil, Joel Nothman, and Ben Hachey

We propose novel radical features from automatic translation for event extraction. Event detection is a complex language processing task for which it is expensive to collect training data, making generalisation challenging. We derive meaningful subword features from automatic translations into target language. Results suggest this method is particularly useful when using languages with writing systems that facilitate easy decomposition into subword features, e.g., logograms and Cangjie. The best result combines logogram features from Chinese and Japanese with syllable features from Korean, providing an additional 3.0 points f-score when added to state-of-the-art generalisation features on the TAC KBP 2015 Event Nugget task.

EviNets: Neural Networks for Combining Evidence Signals for Factoid Question Answering
Denis Savenkov and Eugene Agichtein

A critical task for question answering is the final answer selection stage, which has to combine multiple signals available about each answer candidate. This paper proposes EviNets: a novel neural network architecture for factoid question answering. EviNets scores candidate answer entities by combining the available supporting evidence, e.g., structured knowledge bases and unstructured text documents. EviNets represents each piece of evidence with a dense embeddings vector, scores their relevance to the question, and aggregates the support for each candidate to predict their final scores. Each of the components is generic and allows plugging in a variety of models for semantic similarity scoring and information aggregation. We demonstrate the effectiveness of EviNets in experiments on the existing TREC QA and WikiMovies benchmarks, and on the new Yahoo! Answers dataset introduced in this paper. EviNets can be extended to other information types and could facilitate future work on combining evidence signals for joint reasoning in question answering.

Pocket Knowledge Base Population
Travis Wolfe, Mark Dredze, and Benjamin Van Durme

Existing Knowledge Base Population methods extract relations from a closed relational schema with limited coverage leading to sparse KBs. We propose Pocket Knowledge Base Population (PKBP), the task of dynamically constructing a KB of entities related to a query and finding the best characterization of relationships between entities. We describe novel Open Information Extraction methods which leverage the PKB to find informative trigger words. We evaluate using existing KBP shared-task data as well anew annotations collected

for this work. Our methods produce high quality KB from just text with many more entities and relationships than existing KBP systems.

Answering Complex Questions Using Open Information Extraction
Tushar Khot, Ashish Sabharwal, and Peter Clark

While there has been substantial progress in factoid question-answering (QA), answering complex questions remains challenging, typically requiring both a large body of knowledge and inference techniques. Open Information Extraction (Open IE) provides a way to generate semi-structured knowledge for QA, but to date such knowledge has only been used to answer simple questions with retrieval-based methods. We overcome this limitation by presenting a method for reasoning with Open IE knowledge, allowing more complex questions to be handled. Using a recently proposed support graph optimization framework for QA, we develop a new inference model for Open IE, in particular one that can work effectively with multiple short facts, noise, and the relational structure of tuples. Our model significantly outperforms a state-of-the-art structured solver on complex questions of varying difficulty, while also removing the reliance on manually curated knowledge.

Bootstrapping for Numerical Open IE
Swarnadeep Saha, Harinder Pal, and Mausam

We design and release BONIE, the first open numerical relation extractor, for extracting Open IE tuples where one of the arguments is a number or a quantity-unit phrase. BONIE uses bootstrapping to learn the specific dependency patterns that express numerical relations in a sentence. BONIE's novelty lies in task-specific customizations, such as inferring implicit relations, which are clear due to context such as units (for e.g., 'square kilometers' suggests area, even if the word 'area' is missing in the sentence). BONIE obtains 1.5x yield and 15 point precision gain on numerical facts over a state-of-the-art Open IE system.

Feature-Rich Networks for Knowledge Base Completion
Alexandros Komninos and Suresh Manandhar

We propose jointly modelling Knowledge Bases and aligned text with Feature-Rich Networks. Our models perform Knowledge Base Completion by learning to represent and compose diverse feature types from partially aligned and noisy resources. We perform experiments on Freebase utilizing additional entity type information and syntactic textual relations. Our evaluation suggests that the proposed models can better incorporate side information than previously proposed combinations of bilinear models with convolutional neural networks, showing large improvements when scoring the plausibility of unobserved facts with associated textual mentions.

Fine-Grained Entity Typing with High-Multiplicity Assignments
Maxim Rabinovich and Dan Klein

As entity type systems become richer and more fine-grained, we expect the number of types assigned to a given entity to increase. However, most fine-grained typing work has focused on datasets that exhibit a low degree of type multiplicity. In this paper, we consider the high-multiplicity regime inherent in data sources such as Wikipedia that have semi-open type systems. We introduce a set-prediction approach to this problem and show that our model outperforms unstructured baselines on a new Wikipedia-based fine-grained typing corpus.

Group Sparse CNNs for Question Classification with Answer Sets
Mingbo Ma, Liang Huang, Bing Xiang, and Bowen Zhou

Question classification is an important task with wide applications. However, traditional techniques treat questions as general sentences, ignoring the corresponding answer data. In order to consider answer information into question modeling, we first introduce novel group sparse autoencoders which refine question representation by utilizing group information in the answer set. We then propose novel group sparse CNNs which naturally learn question representation with respect to their answers by implanting group sparse autoencoders into traditional CNNs. The proposed model significantly outperform strong baselines on four datasets.

Multi-Task Learning of Keyphrase Boundary Classification
Isabelle Augenstein and Anders Søgaard

Keyphrase boundary classification (KBC) is the task of detecting keyphrases in scientific articles and labelling them with respect to predefined types. Although important in practice, this task is so far underexplored, partly due to the lack of labelled data. To overcome this, we explore several auxiliary tasks, including semantic supersense tagging and identification of multi-word expressions, and cast the task as a multi-task learning problem with deep recurrent neural networks. Our multi-task models perform significantly better than previous state of the art approaches on two scientific KBC datasets, particularly for long keyphrases.

Cardinal Virtues: Extracting Relation Cardinalities from Text
Paramita Mirza, Simon Razniewski, Fariz Darari, and Gerhard Weikum

Information extraction (IE) from text has largely focused on relations between individual entities, such as who has won which award. However, some facts are never fully mentioned, and no IE method has perfect recall. Thus, it is beneficial to also tap contents about the cardinalities of these relations, for example, how many awards someone has won. We introduce this novel problem of extracting cardinalities and discusses the specific challenges that set it apart from standard IE. We present a distant supervision method using conditional random fields. A preliminary evaluation results in precision between 3% and 55%, depending on the difficulty of relations.

Integrating Deep Linguistic Features in Factuality Prediction over Unified Datasets
Gabriel Stanovsky, Judith Eckle-Kohler, Yevgeniy Puzikov, Ido Dagan, and Iryna Gurevych

Previous models for the assessment of commitment towards a predicate in a sentence (also known as factuality prediction) were trained and tested against a specific annotated dataset, subsequently limiting the generality of their results. In this work we propose an intuitive method for mapping three previously annotated corpora onto a single factuality scale, thereby enabling models to be tested across these corpora. In addition, we design a novel model for factuality prediction by first extending a previous rule-based factuality prediction system and applying it over an abstraction of dependency trees, and then using the output of this system in a supervised classifier. We show that this model outperforms previous methods on all three datasets. We make both the unified factuality corpus and our new model publicly available.

Question Answering on Knowledge Bases and Text using Universal Schema and Memory Networks
Rajarshi Das, Manzil Zaheer, Siva Reddy, and Andrew McCallum

Existing question answering methods infer answers either from a knowledge base or from raw text. While knowledge base (KB) methods are good at answering compositional questions, their performance is often affected by the incompleteness of the KB. Au contraire, web text contains millions of facts that are absent in the KB, however in an unstructured form. Universal schema can support reasoning on the union of both structured KBs and unstructured text by aligning them in a common embedded space. In this paper we extend universal schema to natural language question answering, employing Memory networks to attend to the large body of facts in the combination of text and KB. Our models can be trained in an end-to-end fashion on question-answer pairs. Evaluation results on Spades fill-in-the-blank question answering dataset show that exploiting universal schema for question answering is better than using either a KB or text alone. This model also outperforms the current state-of-the-art by 8.5 F1 points.

Differentiable Scheduled Sampling for Credit Assignment
Kartik Goyal, Chris Dyer, and Taylor Berg-Kirkpatrick

We demonstrate that a continuous relaxation of the argmax operation can be used to create a differentiable approximation to greedy decoding in sequence-to-sequence (seq2seq) models. By incorporating this approximation into the scheduled sampling training procedure–a well-known technique for correcting exposure bias–we introduce a new training objective that is continuous and differentiable everywhere and can provide informative gradients near points where previous decoding decisions change their value. By using a related approximation, we also demonstrate a similar approach to sampled-based training. We show that our approach outperforms both standard cross-entropy training and scheduled sampling procedures in two sequence prediction tasks: named entity recognition and machine translation.

A Deep Network with Visual Text Composition Behavior
Hongyu Guo

While natural languages are compositional, how state-of-the-art neural models achieve compositionality is still unclear. We propose a deep network, which not only achieves competitive accuracy for text classification, but also exhibits compositional behavior. That is, while creating hierarchical representations of a piece of text, such as a sentence, the lower layers of the network distribute their layer-specific attention weights to individual words. In contrast, the higher layers compose meaningful phrases and clauses, whose lengths increase as the networks get deeper until fully composing the sentence.

Neural System Combination for Machine Translation
Long Zhou, Wenpeng Hu, Jiajun Zhang, and Chengqing Zong

Neural machine translation (NMT) becomes a new approach to machine translation and generates much more fluent results compared to statistical machine translation (SMT). However, SMT is usually better than NMT in translation adequacy. It is therefore a promising direction to combine the advantages of both NMT and SMT.

In this paper, we propose a neural system combination framework leveraging multi-source NMT, which takes as input the outputs of NMT and SMT systems and produces the final translation. Extensive experiments on the Chinese-to-English translation task show that our model archives significant improvement by 5.3 BLEU points over the best single system output and 3.4 BLEU points over the state-of-the-art traditional system combination methods.

An Empirical Comparison of Domain Adaptation Methods for Neural Machine Translation
Chenhui Chu, Raj Dabre, and Sadao Kurohashi

In this paper, we propose a novel domain adaptation method named "mixed fine tuning" for neural machine translation (NMT). We combine two existing approaches namely fine tuning and multi domain NMT. We first train an NMT model on an out-of-domain parallel corpus, and then fine tune it on a parallel corpus which is a mix of the in-domain and out-of-domain corpora. All corpora are augmented with artificial tags to indicate specific domains. We empirically compare our proposed method against fine tuning and multi domain methods and discuss its benefits and shortcomings.

Efficient Extraction of Pseudo-Parallel Sentences from Raw Monolingual Data Using Word Embeddings
Benjamin Marie and Atsushi Fujita

We propose a new method for extracting pseudo-parallel sentences from a pair of large monolingual corpora, without relying on any document-level information. Our method first exploits word embeddings in order to efficiently evaluate trillions of candidate sentence pairs and then a classifier to find the most reliable ones. We report significant improvements in domain adaptation for statistical machine translation when using a translation model trained on the sentence pairs extracted from in-domain monolingual corpora.

Feature Hashing for Language and Dialect Identification
Shervin Malmasi and Mark Dras

We evaluate feature hashing for language identification (LID), a method not previously used for this task. Using a standard dataset, we first show that while feature performance is high, LID data is highly dimensional and mostly sparse (>99.5%) as it includes large vocabularies for many languages; memory requirements grow as languages are added. Next we apply hashing using various hash sizes, demonstrating that there is no performance loss with dimensionality reductions of up to 86%. We also show that using an ensemble of low-dimension hash-based classifiers further boosts performance. Feature hashing is highly useful for LID and holds great promise for future work in this area.

Detection of Chinese Word Usage Errors for Non-Native Chinese Learners with Bidirectional LSTM
Yow-Ting Shiue, Hen-Hsen Huang, and Hsin-Hsi Chen

Selecting appropriate words to compose a sentence is one common problem faced by non-native Chinese learners. In this paper, we propose (bidirectional) LSTM sequence labeling models and explore various features to detect word usage errors in Chinese sentences. By combining CWINDOW word embedding features and POS information, the best bidirectional LSTM model achieves accuracy 0.5138 and MRR 0.6789 on the HSK dataset. For 80.79% of the test data, the model ranks the ground-truth within the top two at position level.

Automatic Compositor Attribution in the First Folio of Shakespeare
Maria Ryskina, Hannah Alpert-Abrams, Dan Garrette, and Taylor Berg-Kirkpatrick

Compositor attribution, the clustering of pages in a historical printed document by the individual who set the type, is a bibliographic task that relies on analysis of orthographic variation and inspection of visual details of the printed page. In this paper, we introduce a novel unsupervised model that jointly describes the textual and visual features needed to distinguish compositors. Applied to images of Shakespeare's First Folio, our model predicts attributions that agree with the manual judgements of bibliographers with an accuracy of 87%, even on text that is the output of OCR.

STAIR Captions: Constructing a Large-Scale Japanese Image Caption Dataset
Yuya Yoshikawa, Yutaro Shigeto, and Akikazu Takeuchi

In recent years, automatic generation of image descriptions (captions), that is, image captioning, has attracted a great deal of attention. In this paper, we particularly consider generating Japanese captions for images. Since most available caption datasets have been constructed for English language, there are few datasets for Japanese. To tackle this problem, we construct a large-scale Japanese image caption dataset based on images from MS-COCO, which is called STAIR Captions. STAIR Captions consists of 820,310 Japanese captions for 164,062 images. In the experiment, we show that a neural network trained using STAIR Captions can generate more

natural and better Japanese captions, compared to those generated using English-Japanese machine translation after generating English captions.

"Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection
William Yang Wang

Automatic fake news detection is a challenging problem in deception detection, and it has tremendous real-world political and social impacts. However, statistical approaches to combating fake news has been dramatically limited by the lack of labeled benchmark datasets. In this paper, we present LIAR: a new, publicly available dataset for fake news detection. We collected a decade-long, 12.8K manually labeled short statements in various contexts from PolitiFact.com, which provides detailed analysis report and links to source documents for each case. This dataset can be used for fact-checking research as well. Notably, this new dataset is an order of magnitude larger than previously largest public fake news datasets of similar type. Empirically, we investigate automatic fake news detection based on surface-level linguistic patterns. We have designed a novel, hybrid convolutional neural network to integrate meta-data with text. We show that this hybrid approach can improve a text-only deep learning model.

English Multiword Expression-aware Dependency Parsing Including Named Entities
Akihiko Kato, Hiroyuki Shindo, and Yuji Matsumoto

Because syntactic structures and spans of multiword expressions (MWEs) are independently annotated in many English syntactic corpora, they are generally inconsistent with respect to one another, which is harmful to the implementation of an aggregate system. In this work, we construct a corpus that ensures consistency between dependency structures and MWEs, including named entities. Further, we explore models that predict both MWE-spans and an MWE-aware dependency structure. Experimental results show that our joint model using additional MWE-span features achieves an MWE recognition improvement of 1.35 points over a pipeline model.

Improving Semantic Composition with Offset Inference
Thomas Kober, Julie Weeds, Jeremy Reffin, and David Weir

Count-based distributional semantic models suffer from sparsity due to unobserved but plausible co-occurrences in any text collection. This problem is amplified for models like Anchored Packed Trees (APTs), that take the grammatical type of a co-occurrence into account. We therefore introduce a novel form of distributional inference that exploits the rich type structure in APTs and infers missing data by the same mechanism that is used for semantic composition.

Learning Topic-Sensitive Word Representations
Marzieh Fadaee, Arianna Bisazza, and Christof Monz

Distributed word representations are widely used for modeling words in NLP tasks. Most of the existing models generate one representation per word and do not consider different meanings of a word. We present two approaches to learn multiple topic-sensitive representations per word by using Hierarchical Dirichlet Process. We observe that by modeling topics and integrating topic distributions for each document we obtain representations that are able to distinguish between different meanings of a given word. Our models yield statistically significant improvements for the lexical substitution task indicating that commonly used single word representations, even when combined with contextual information, are insufficient for this task.

Temporal Word Analogies: Identifying Lexical Replacement with Diachronic Word Embeddings
Terrence Szymanski

This paper introduces the concept of temporal word analogies: pairs of words which occupy the same semantic space at different points in time. One well-known property of word embeddings is that they are able to effectively model traditional word analogies ("word w1 is to word w2 as word w3 is to word w4") through vector addition. Here, I show that temporal word analogies ("word w1 at time t1 is like word w2 at time t2") can effectively be modeled with diachronic word embeddings, provided that the independent embedding spaces from each time period are appropriately transformed into a common vector space. When applied to a diachronic corpus of news articles, this method is able to identify temporal word analogies such as "Ronald Reagan in 1987 is like Bill Clinton in 1997", or "Walkman in 1987 is like iPod in 2007".

Methodical Evaluation of Arabic Word Embeddings
Mohammed Elrazzaz, Shady Elbassuoni, Khaled Shaban, and Chadi Helwe

Many unsupervised learning techniques have been proposed to obtain meaningful representations of words from text. In this study, we evaluate these various techniques when used to generate Arabic word embeddings.

We first build a benchmark for the Arabic language that can be utilized to perform intrinsic evaluation of different word embeddings. We then perform additional extrinsic evaluations of the embeddings based on two NLP tasks.

Multilingual Connotation Frames: A Case Study on Social Media for Targeted Sentiment Analysis and Forecast
Hannah Rashkin, Eric Bell, Yejin Choi, and Svitlana Volkova

People around the globe respond to major real world events through social media. To study targeted public sentiments across many languages and geographic locations, we introduce multilingual connotation frames: an extension from English connotation frames of Rashkin et al. (2016) with 10 additional European languages, focusing on the implied sentiments among event participants engaged in a frame. As a case study, we present large scale analysis on targeted public sentiments toward salient events and entities using 1.2 million multilingual connotation frames extracted from Twitter.

Best-Worst Scaling More Reliable than Rating Scales: A Case Study on Sentiment Intensity Annotation
Svetlana Kiritchenko and Saif M. Mohammad

Rating scales are a widely used method for data annotation; however, they present several challenges, such as difficulty in maintaining inter- and intra-annotator consistency. Best—worst scaling (BWS) is an alternative method of annotation that is claimed to produce high-quality annotations while keeping the required number of annotations similar to that of rating scales. However, the veracity of this claim has never been systematically established. Here for the first time, we set up an experiment that directly compares the rating scale method with BWS. We show that with the same total number of annotations, BWS produces significantly more reliable results than the rating scale.

Demographic Inference on Twitter using Recursive Neural Networks
Sunghwan Mac Kim, Qiongkai Xu, Lizhen Qu, Stephen Wan, and Cecile Paris

In social media, demographic inference is a critical task in order to gain a better understanding of a cohort and to facilitate interacting with one's audience. Most previous work has made independence assumptions over topological, textual and label information on social networks. In this work, we employ recursive neural networks to break down these independence assumptions to obtain inference about demographic characteristics on Twitter. We show that our model performs better than existing models including the state-of-the-art.

Twitter Demographic Classification Using Deep Multi-modal Multi-task Learning
Prashanth Vijayaraghavan, Soroush Vosoughi, and Deb Roy

Twitter should be an ideal place to get a fresh read on how different issues are playing with the public, one that's potentially more reflective of democracy in this new media age than traditional polls. Pollsters typically ask people a fixed set of questions, while in social media people use their own voices to speak about whatever is on their minds. However, the demographic distribution of users on Twitter is not representative of the general population. In this paper, we present a demographic classifier for gender, age, political orientation and location on Twitter. We collected and curated a robust Twitter demographic dataset for this task. Our classifier uses a deep multi-modal multi-task learning architecture to reach a state-of-the-art performance, achieving an F1-score of 0.89, 0.82, 0.86, and 0.68 for gender, age, political orientation, and location respectively.

A Network Framework for Noisy Label Aggregation in Social Media
Xueying Zhan, Yaowei Wang, Yanghui Rao, Haoran Xie, Qing Li, Fu Lee Wang, and Tak-Lam Wong

This paper focuses on the task of noisy label aggregation in social media, where users with different social or culture backgrounds may annotate invalid or malicious tags for documents. To aggregate noisy labels at a small cost, a network framework is proposed by calculating the matching degree of a document's topics and the annotators' meta-data. Unlike using the back-propagation algorithm, a probabilistic inference approach is adopted to estimate network parameters. Finally, a new simulation method is designed for validating the effectiveness of the proposed framework in aggregating noisy labels.

Parser Adaptation for Social Media by Integrating Normalization
Rob van der Goot and Gertjan van Noord

This work explores different approaches of using normalization for parser adaptation. Traditionally, normalization is used as separate pre-processing step. We show that integrating the normalization model into the parsing algorithm is more beneficial. This way, multiple normalization candidates can be leveraged, which improves parsing performance on social media. We test this hypothesis by modifying the Berkeley parser;

out-of-the-box it achieves an F1 score of 66.52. Our integrated approach reaches a significant improvement with an F1 score of 67.36, while using the best normalization sequence results in an F1 score of only 66.94.

Poster Session P1 (SRW Papers)

Time: *18:00–21:30* Location: Bayshore Grand Ballroom/Foyer

Computational Characterization of Mental States: A Natural Language Processing Approach
Facundo Carrillo
Psychiatry is an area of medicine that strongly bases its diagnoses on the psychiatrist's subjective appreciation. More precisely, speech is used almost exclusively as a window to the patient's mind. Few other cues are available to objectively justify a diagnostic, unlike what happens in other disciplines which count on laboratory tests or imaging procedures, such as X-rays. Daily practice is based on the use of semi-structured interviews and standardized tests to build the diagnoses, heavily relying on her personal experience. This methodology has a big problem: diagnoses are commonly validated a posteriori in function of how the pharmacological treatment works. This validation cannot be done until months after the start of the treatment and, if the patient condition does not improve, the psychiatrist often changes the diagnosis and along with the pharmacological treatment. This delay prolongs the patient's suffering until the correct diagnosis is found. According to NIMH, more than 1% and 2% of US population is affected by Schizophrenia and Bipolar Disorder, respectively. Moreover, the WHO reported that the global cost of mental illness reached $2.5T in 2010 [1] . The task of diagnosis, largely simplified, mainly consists of understanding the mind state through the extraction of patterns from the patient's speech and finding the best matching pathology in the standard diagnostic literature. This pipeline, consisting of extracting patterns and then classifying them, loosely resembles the common pipelines used in supervised learning schema. Therefore, we propose to augment the psychiatrists' diagnosis toolbox with an artificial intelligence system based on natural language processing and machine learning algorithms. The proposed system would assist in the diagnostic using a patient's speech as input. The understanding and insights obtained from customizing these systems to specific pathologies is likely to be more broadly applicable to other NLP tasks, therefore we expect to make contributions not only for psychiatry but also within the computer science community. We intend to develop these ideas and evaluate them beyond the lab setting. Our end goal is to make it possible for a practitioner to integrate our tools into her daily practice with minimal effort

Improving Distributed Representations of Tweets - Present and Future
Ganesh Jawahar
Unsupervised representation learning for tweets is an important research field which helps in solving several business applications such as sentiment analysis, hashtag prediction, paraphrase detection and microblog ranking. A good tweet representation learning model must handle the idiosyncratic nature of tweets which poses several challenges such as short length, informal words, unusual grammar and misspellings. However, there is a lack of prior work which surveys the representation learning models with a focus on tweets. In this work, we organize the models based on its objective function which aids the understanding of the literature. We also provide interesting future directions, which we believe are fruitful in advancing this field by building high-quality tweet representation learning models.

Bilingual Word Embeddings with Bucketed CNN for Parallel Sentence Extraction
Jeenu Grover and Pabitra Mitra
We propose a novel model which can be used to align the sentences of two different languages using neural architectures. First, we train our model to get the bilingual word embeddings and then, we create a similarity matrix between the words of the two sentences. Because of different lengths of the sentences involved, we get a matrix of varying dimension. We dynamically pool the similarity matrix into a matrix of fixed dimension and use Convolutional Neural Network (CNN) to classify the sentences as aligned or not. To further improve upon this technique, we bucket the sentence pairs to be classified into different groups and train CNN's separately. Our approach not only solves sentence alignment problem but our model can be regarded as a generic bag-of-words similarity measure for monolingual or bilingual corpora.

nQuery - A Natural Language Statement to SQL Query Generator
Nandan Sukthankar, Sanket Maharnawar, Pranay Deshmukh, Yashodhara Haribhakta, and Vibhavari Kamble

In this research, an intelligent system is designed between the user and the database system which accepts natural language input and then converts it into an SQL query. The research focuses on incorporating complex queries along with simple queries irrespective of the database. The system accommodates aggregate functions, multiple conditions in WHERE clause, advanced clauses like ORDER BY, GROUP BY and HAVING. The system handles single sentence natural language inputs, which are with respect to selected database. The research currently concentrates on MySQL database system. The natural language statement goes through various stages of Natural Language Processing like morphological, lexical, syntactic and semantic analysis resulting in SQL query formation.

V for Vocab: An Intelligent Flashcard Application
Nihal V. Nayak, Tanmay Chinchore, Aishwarya Hanumanth Rao, Shane Michael Martin, Sagar Nagaraj Simha, G. M. Lingaraju, and H. S. Jamadagni

Students choose to use flashcard applications available on the Internet to help memorize word-meaning pairs. This is helpful for tests such as GRE, TOEFL or IELTS, which emphasize on verbal skills. However, monotonous nature of flashcard applications can be diminished with the help of Cognitive Science through Testing Effect. Experimental evidences have shown that memory tests are an important tool for long term retention (Roediger and Karpicke, 2006). Based on these evidences, we developed a novel flashcard application called "V for Vocab" that implements short answer based tests for learning new words. Furthermore, we aid this by implementing our short answer grading algorithm which automatically scores the user's answer. The algorithm makes use of an alternate thesaurus instead of traditional Wordnet and delivers state-of-the-art performance on popular word similarity datasets. We also look to lay the foundation for analysis based on implicit data collected from our application.

Are You Asking the Right Questions? Teaching Machines to Ask Clarification Questions
Sudha Rao

Inquiry is fundamental to communication, and machines cannot effectively collaborate with humans unless they can ask questions. In this thesis work, we explore how can we teach machines to ask clarification questions when faced with uncertainty, a goal of increasing importance in today's automated society. We do a preliminary study using data from StackExchange, a plentiful online resource where people routinely ask clarifying questions to posts so that they can better offer assistance to the original poster. We build neural network models inspired by the idea of the expected value of perfect information: a good question is one whose expected answer is going to be most useful. To build generalizable systems, we propose two future research directions: a template-based model and a sequence-to-sequence based neural generative model.

Building a Non-Trivial Paraphrase Corpus Using Multiple Machine Translation Systems
Yui Suzuki, Tomoyuki Kajiwara, and Mamoru Komachi

We propose a novel sentential paraphrase acquisition method. To build a well-balanced corpus for Paraphrase Identification, we especially focus on acquiring both non-trivial positive and negative instances. We use multiple machine translation systems to generate positive candidates and a monolingual corpus to extract negative candidates. To collect non-trivial instances, the candidates are uniformly sampled by word overlap rate. Finally, annotators judge whether the candidates are either positive or negative. Using this method, we built and released the first evaluation corpus for Japanese paraphrase identification, which comprises 655 sentence pairs.

Segmentation Guided Attention Networks for Visual Question Answering
Vasu Sharma, Ankita Bishnu, and Labhesh Patel

In this paper we propose to solve the problem of Visual Question answering by using a novel segmentation guided attention based networks which we call SegAttendNet. We use image segmentation maps, generated by a Fully Convolutional Deep Neural Network to refine our attention maps and use these refined attention maps to make the model focus on the relevant parts of the image to answer a question. The refined attention maps are used by the LSTM network to learn to produce the answer. We presently train our model on the visual7W dataset and do a category wise evaluation of the 7 question categories. We achieve state of the art results on this dataset and beat the previous benchmark on this dataset by a 1.5% margin improving the question answering accuracy from 54.1% to 55.6% and demonstrate improvements in each of the question categories. We also visualize our generated attention maps and note their improvement over the attention maps generated by the previous best approach.

Text-based Speaker Identification on Multiparty Dialogues Using Multi-document Convolutional Neural Networks
Kaixin Ma, Catherine Xiao, and Jinho D. Choi

We propose a convolutional neural network model for text-based speaker identification on multiparty dialogues extracted from the TV show, *Friends*. While most previous works on this task rely heavily on acoustic features, our approach attempts to identify speakers in dialogues using their speech patterns as captured by transcriptions to the TV show. It has been shown that different individual speakers exhibit distinct idiolectal styles. Several convolutional neural network models are developed to discriminate between differing speech patterns. Our results confirm the promise of text-based approaches, with the best performing model showing an accuracy improvement of over 6% upon the baseline CNN model.

Variation Autoencoder Based Network Representation Learning for Classification
Hang Li, Haozheng Wang, Zhenglu Yang, and Masato Odagaki

Network representation is the basis of many applications and of extensive interest in various fields, such as information retrieval, social network analysis, and recommendation systems. Most previous methods for network representation only consider the incomplete aspects of a problem, including link structure, node information, and partial integration. The present study introduces a deep network representation model that seamlessly integrates the text information and structure of a network. The model captures highly non-linear relationships between nodes and complex features of a network by exploiting the variational autoencoder (VAE), which is a deep unsupervised generation algorithm. The representation learned with a paragraph vector model is merged with that learned with the VAE to obtain the network representation, which preserves both structure and text information. Comprehensive experiments is conducted on benchmark datasets and find that the introduced model performs better than state-of-the-art techniques.

Blind Phoneme Segmentation With Temporal Prediction Errors
Paul Michel, Okko Räsänen, Roland Thiolliere, and Emmanuel Dupoux

Phonemic segmentation of speech is a critical step of speech recognition systems. We propose a novel unsupervised algorithm based on sequence prediction models such as Markov chains and recurrent neural networks. Our approach consists in analyzing the error profile of a model trained to predict speech features frame-by-frame. Specifically, we try to learn the dynamics of speech in the MFCC space and hypothesize boundaries from local maxima in the prediction error. We evaluate our system on the TIMIT dataset, with improvements over similar methods

Automatic Generation of Jokes in Hindi
Srishti Aggarwal and Radhika Mamidi

When it comes to computational language processing systems, humour is a relatively unexplored domain, especially more so for Hindi (or rather, most languages other than English). Most researchers agree that a joke consists of two main parts - the setup and the punchline, which the humour being encoded in the incongruity between the two. In this paper, we look at *Dur se Dekha* jokes, a restricted domain of humorous three liner poetry in Hindi. We analyze their structure to understand how humour is encoded in them and formalize it. We then develop a system which is successfully able to generate a basic form of these jokes.

Word Embedding for Response-To-Text Assessment of Evidence
Haoran Zhang and Diane Litman

Manually grading the Response to Text Assessment (RTA) is labor intensive. Therefore, an automatic method is being developed for scoring analytical writing when the RTA is administered in large numbers of classrooms. Our long-term goal is to also use this scoring method to provide formative feedback to students and teachers about students' writing quality. As a first step towards this goal, interpretable features for automatically scoring the evidence rubric of the RTA have been developed. In this paper, we present a simple but promising method for improving evidence scoring by employing the word embedding model. We evaluate our method on corpora of responses written by upper elementary students.

Domain Specific Automatic Question Generation from Text
Katira Soleymanzadeh

The goal of my doctoral thesis is to automatically generate interrogative sentences from descriptive sentences of Turkish biology text. We employ syntactic and semantic approaches to parse descriptive sentences. Syntactic and semantic approaches utilize syntactic (constituent or dependency) parsing and semantic role labeling systems respectively. After parsing step, question statements whose answers are embedded in the descriptive sentences are going to be formulated by using some predefined rules and templates. Syntactic parsing is done using an open source dependency parser called MaltParser (Nivre et al. 2007). Whereas to accomplish semantic parsing, we will construct a biological proposition bank (BioPropBank) and a corpus annotated with semantic roles. Then we will employ supervised methods to automatic label the semantic roles of a sentence.

SoccEval: An Annotation Schema for Rating Soccer Players
Jose Ramirez, Matthew Garber, and Xinhao Wang

This paper describes the SoccEval Annotation Project, an annotation schema designed to support machine-learning classification efforts to evaluate the performance of soccer players based on match reports taken from online news sources. In addition to factual information about player attributes and actions, the schema annotates subjective opinions about them. After explaining the annotation schema and annotation process, we describe a machine learning experiment. Classifiers trained on features derived from annotated data performed better than a baseline trained on unigram features. Initial results suggest that improvements can be made to the annotation scheme and guidelines as well as the amount of data annotated. We believe our schema could be potentially expanded to extract more information about soccer players and teams.

Accent Adaptation for the Air Traffic Control Domain
Matthew Garber, Meital Singer, and Christopher Ward

Automated speech recognition (ASR) plays a significant role in training and simulation systems for air traffic controllers. However, because English is the default language used in air traffic control (ATC), ASR systems often encounter difficulty with speakers' non-native accents, for which there is a paucity of data. This paper examines the effects of accent adaptation on the recognition of non-native English speech in the ATC domain. Accent adaptation has been demonstrated to be an effective way to model under-resourced speech, and can be applied to a variety of models. We use Subspace Gaussian Mixture Models (SGMMs) with the Kaldi Speech Recognition Toolkit to adapt acoustic models from American English to German-accented English, and compare it against other adaptation methods. Our results provide additional evidence that SGMMs can be an efficient and effective way to approach this problem, particularly with smaller amounts of accented training data.

Generating Steganographic Text with LSTMs
Tina Fang, Martin Jaggi, and Katerina Argyraki

Motivated by concerns for user privacy, we design a steganographic system ("stegosystem") that enables two users to exchange encrypted messages without an adversary detecting that such an exchange is taking place. In this paper, we propose a novel linguistic stegosystem based on a Long-Short Term Memory (LSTM) neural network. We demonstrate our approach on the Twitter and Enron email datasets and show that it yields high-quality steganographic text while significantly improving capacity (encrypted bits per word) relative to the state-of-the-art.

Predicting Depression for Japanese Blog Text
Misato Hiraga

This study aims to predict clinical depression, a prevalent mental disorder, from blog posts written in Japanese by using machine learning approaches. The study focuses on how data quality and various types of linguistic features (characters, tokens, and lemmas) affect prediction outcome. Depression prediction achieved 95.5% accuracy using selected lemmas as features.

Fast Forward Through Opportunistic Incremental Meaning Representation Construction
Petr Babkin and Sergei Nirenburg

One of the challenges semantic parsers face involves upstream errors originating from pre-processing modules such as ASR and syntactic parsers, which undermine the end result from the get go. We report the work in progress on a novel incremental semantic parsing algorithm that supports simultaneous application of independent heuristics and facilitates the construction of partial but potentially actionable meaning representations to overcome this problem. Our contribution to this point is mainly theoretical. In future work we intend to evaluate the algorithm as part of a dialogue understanding system on state of the art benchmarks.

Modeling Situations in Neural Chat Bots
Shoetsu Sato, Naoki Yoshinaga, Masashi Toyoda, and Masaru Kitsuregawa

Social media accumulates vast amounts of online conversations that enable data-driven modeling of chat dialogues. It is, however, still hard to utilize the neural network-based SEQ2SEQ model for dialogue modeling in spite of its acknowledged success in machine translation. The main challenge comes from the high degrees of freedom of responses in dialogues. In this study, we explore neural conversational models that have general mechanisms for handling a variety of situations that affect our response. In our experiments, we confirmed the effectiveness of the proposed method in a response selection test by using massive dialogue data we have collected from Twitter.

An Empirical Study on End-to-End Sentence Modelling
Kurt Junshean Espinosa

Accurately representing the meaning of a piece of text, otherwise known as sentence modelling, is an important component in many natural language inference tasks. We survey the spectrum of these methods, which lie along two dimensions: input representation granularity and composition model complexity. Using this framework, we reveal in our quantitative and qualitative experiments the limitations of the current state-of-the-art model in the context of sentence similarity tasks.

Varying Linguistic Purposes of Emoji in (Twitter) Context
Noa Naaman, Hannah Provenza, and Orion Montoya

Research into emoji in textual communication has, thus far, focused on high-frequency usages and the ambiguity of interpretations. Investigation of emoji uses across a wide range of uses can divide them into different linguistic functions: function and content words, or multimodal affective markers. Identifying where an emoji is merely replacing part of the text allows NLP tools the possibility of parsing them as any other word or phrase. Smiling emoticons are usually left out of data sets, but if they are used as the noun "smile" or the verb "smiling", we should be able to predict their part of speech. We report on an annotation task on English Twitter data with the goal of classifying emoji usage by these categories, and on the effectiveness of a classifier trained on these annotations. We find that it is possible to train a classifier to tell the difference between those emoji used as linguistic content words and those used as paralinguistic or affective multimodal markers even with a small amount of training data, but that accurate sub-classification of these multimodal emoji into specific classes like attitude, topic, or gesture will require more data and more feature engineering.

Negotiation of Antibiotic Treatment in Medical Consultations: A Corpus Based Study
Nan Wang

Doctor-patient conversation is considered a contributing factor to antibiotic over-prescription. Some language practices have been identified as parent pressuring doctors for prescribing; other practices are considered as likely to engender parent resistance to non-antibiotic treatment recommendations. In social science studies, approaches such as conversation analysis have been applied to identify those language practices. Current research for dialogue systems offer an alternative approach. Past research proved that corpus-based approaches have been effectively used for research involving modeling dialogue acts and sequential relations. In this proposal, we propose a corpus-based study of doctor-patient conversations of antibiotic treatment negotiation in pediatric consultations. Based on findings from conversation analysis studies, we use a computational linguistic approach to assist annotating and modeling of doctor-patient language practices, and analyzing their influence on antibiotic over-prescribing.

4

Main Conference: Tuesday, August 1

Overview

7:30 – 9:00 **Breakfast** *Bayshore Grand Foyer*

9:00 – 10:10 **Plenary Session** *Bayshore Grand Ballroom*
Invited Talk: Noah Smith – "Squashing Computational Linguistics"

10:10 – 10:30 **Coffee break** *Bayshore Grand Foyer*

Session 4

10:30 – 12:05					
	Information Extraction 3 (NN)	Cognitive Modelling 1 / Vision 2	Dialogue 2	Machine Translation 2	Social Media 1
	Salons B/C	*Salons E/F*	*Salon D*	*Salon 1*	*Salons 2/3*

12:05 – 13:30 **Lunch break**

Session 5

13:30 – 15:05					
	Multi-disciplinary 1	Language and Resources 1	Syntax 2 (NN)	Machine Translation 3 (NN)	Sentiment 2
	Salons B/C	*Salons E/F*	*Salon D*	*Salon 1*	*Salons 2/3*

15:05 – 15:25 **Coffee break** *Bayshore Grand Foyer*

Session 6

15:25 – 17:00					
	Information Extraction 4	Semantics 2 (NN)	Discourse 2 / Dialogue 3	Machine Learning 2	Summarization 1
	Salons B/C	*Salons E/F*	*Salon D*	*Salon 1*	*Salons 2/3*

17:45 – 19:40 **Poster Session P2 (includes systems demonstrations)**
Bayshore Grand Ballroom/Foyer, Stanley Park Foyer/Cypress

19:00 – 22:00 **Social event at the Vancouver Aquarium** *Vancouver Aquarium*

Keynote Address: Noah Smith

Squashing Computational Linguistics

Tuesday, August 1, 2017, 9:00–10:10

Bayshore Grand Ballroom

Abstract: The computational linguistics and natural language processing community is experiencing an episode of deep fascination with representation learning. Like many other presenters at this conference, I will describe new ways to use representation learning in models of natural language. Noting that a data-driven model always assumes a theory (not necessarily a good one), I will argue for the benefits of language-appropriate inductive bias for representation-learning-infused models of language. Such bias often comes in the form of assumptions baked into a model, constraints on an inference algorithm, or linguistic analysis applied to data. Indeed, many decades of research in linguistics (including computational linguistics) put our community in a strong position to identify promising inductive biases. The new models, in turn, may allow us to explore previously unavailable forms of bias, and to produce findings of interest to linguistics. I will focus on new models of documents and of sentential semantic structures, and I will emphasize abstract, reusable components and their assumptions rather than applications.

Biography: Noah Smith is an Associate Professor in the Paul G. Allen School of Computer Science and Engineering at the University of Washington. Previously, he was an Associate Professor in the School of Computer Science at Carnegie Mellon University. He received his Ph.D. in Computer Science from Johns Hopkins University and his B.S. in Computer Science and B.A. in Linguistics from the University of Maryland. His research spans many topics in natural language processing, machine learning, and computational social science. He has served on the editorial boards of CL, JAIR, and TACL, as the secretary-treasurer of SIGDAT (2012–2015), and as program co-chair of ACL 2016. Alumni of his research group, Noah's ARK, are international leaders in NLP in academia and industry. Smith's work has been recognized with a UW Innovation award, a Finmeccanica career development chair at CMU, an NSF CAREER award, a Hertz Foundation graduate fellowship, numerous best paper nominations and awards, and coverage by NPR, BBC, CBC, the New York Times, the Washington Post, and Time.

Session 4 Overview – Tuesday, August 1, 2017

	Track A	**Track B**	**Track C**	**Track D**	**Track E**
	Information Extraction 3 (NN)	*Cognitive Modelling 1 / Vision 2*	*Dialogue 2*	*Machine Translation 2*	*Social Media 1*
	Salons B/C	Salons E/F	Salon D	Salon 1	Salons 2/3
10:30	Deep Pyramid Convolutional Neural Networks for Text Categorization *R. Johnson and T. Zhang*	Alignment at Work: Using Language to Distinguish the Internalization and Self-Regulation Components of Cultural Fit in Organizations *G. Doyle, A. Goldberg, S. Srivastava, and M. Frank*	Affect-LM: A Neural Language Model for Customizable Affective Text Generation *S. Ghosh, M. Chollet, E. Laksana, L.-P. Morency, and S. Scherer*	Modeling Source Syntax for Neural Machine Translation *J. Li, D. Xiong, Z. Tu, M. Zhu, M. Zhang, and G. Zhou*	Detect Rumors in Microblog Posts Using Propagation Structure via Kernel Learning *J. Ma, W. Gao, and K.-F. Wong*
10:49	Improved Neural Relation Detection for Knowledge Base Question Answering *M. Yu, W. Yin, K. S. Hasan, C. dos Santos, B. Xiang, and B. Zhou*	Representations of language in a model of visually grounded speech signal *G. Chrupała, L. Gelderloos, and A. Alishahi*	Domain Attention with an Ensemble of Experts *Y.-B. Kim, K. Stratos, and D. Kim*	Sequence-to-Dependency Neural Machine Translation *S. Wu, D. Zhang, N. Yang, M. Li, and M. Zhou*	EmoNet: Fine-Grained Emotion Detection with Gated Recurrent Neural Networks *M. Abdul-Mageed and L. Ungar*
11:08	Deep Keyphrase Generation *R. Meng, S. Zhao, S. Han, D. He, P. Brusilovsky, and Y. Chi*	Spectral Analysis of Information Density in Dialogue Predicts Collaborative Task Performance *Y. Xu and D. Reitter*	Learning Discourse-level Diversity for Neural Dialog Models using Conditional Variational Autoencoders *T. Zhao, R. Zhao, and M. Eskenazi*	[TACL] Head-Lexicalized Bidirectional Tree LSTMs *Y. Teng Zhiyang and Zhang*	Beyond Binary Labels: Political Ideology Prediction of Twitter Users *D. Preoţiuc-Pietro, Y. Liu, D. Hopkins, and L. Ungar*
11:27	Attention-over-Attention Neural Networks for Reading Comprehension *Y. Cui, Z. Chen, S. Wei, S. Wang, T. Liu, and G. Hu*	[TACL] Modelling Semantic Expectation: Using Script Knowledge for Referent Prediction *A. Modi, I. Titov, V. Demberg, A. Sayeed, and M. Pinkal*	Hybrid Code Networks: practical and efficient end-to-end dialog control with supervised and reinforcement learning *J. D. Williams, K. Asadi, and G. Zweig*	[TACL] Pushing the Limits of Translation Quality Estimation *A. F. T. Martins, M. Junczys-Dowmunt, F. N. Kepler, R. Astudillo, C. Hokamp, and R. Grundkiewicz*	Leveraging Behavioral and Social Information for Weakly Supervised Collective Classification of Political Discourse on Twitter *K. M. Johnson, D. Jin, and D. Goldwasser*
11:46	[TACL] Cross-Sentence N-ary Relation Extraction with Graph LSTMs *N. Peng, H. Poon, C. Quirk, K. Toutanova, and W.-t. Yih*	An Analysis of Action Recognition Datasets for Language and Vision Tasks *S. Gella and F. Keller*	Generating Contrastive Referring Expressions *M. Villalba, C. Teichmann, and A. Koller*	Learning to Parse and Translate Improves Neural Machine Translation *A. Eriguchi, Y. Tsuruoka, and K. Cho*	On the Distribution of Lexical Features at Multiple Levels of Analysis *F. Almodaresi, L. Ungar, V. Kulkarni, M. Zakeri, S. Giorgi, and H. A. Schwartz*

Parallel Session 4

Session Session 4A: Information Extraction 3 (NN)

Deep Pyramid Convolutional Neural Networks for Text Categorization
Rie Johnson and Tong Zhang 10:30–10:48

This paper proposes a low-complexity word-level deep convolutional neural network (CNN) architecture for text categorization that can efficiently represent long-range associations in text. In the literature, several deep and complex neural networks have been proposed for this task, assuming availability of relatively large amounts of training data. However, the associated computational complexity increases as the networks go deeper, which poses serious challenges in practical applications. Moreover, it was shown recently that shallow word-level CNNs are more accurate and much faster than the state-of-the-art very deep nets such as character-level CNNs even in the setting of large training data. Motivated by these findings, we carefully studied deepening of word-level CNNs to capture global representations of text, and found a simple network architecture with which the best accuracy can be obtained by increasing the network depth without increasing computational cost by much. We call it deep pyramid CNN. The proposed model with 15 weight layers outperforms the previous best models on six benchmark datasets for sentiment classification and topic categorization.

Improved Neural Relation Detection for Knowledge Base Question Answering
Mo Yu, Wenpeng Yin, Kazi Saidul Hasan, Cicero dos Santos, Bing Xiang, and Bowen Zhou 10:49–11:07

Relation detection is a core component of many NLP applications including Knowledge Base Question Answering (KBQA). In this paper, we propose a hierarchical recurrent neural network enhanced by residual learning which detects KB relations given an input question. Our method uses deep residual bidirectional LSTMs to compare questions and relation names via different levels of abstraction. Additionally, we propose a simple KBQA system that integrates entity linking and our proposed relation detector to make the two components enhance each other. Our experimental results show that our approach not only achieves outstanding relation detection performance, but more importantly, it helps our KBQA system achieve state-of-the-art accuracy for both single-relation (SimpleQuestions) and multi-relation (WebQSP) QA benchmarks.

Deep Keyphrase Generation
Rui Meng, Sanqiang Zhao, Shuguang Han, Daqing He, Peter Brusilovsky, and Yu Chi 11:08–11:26

Keyphrase provides highly-summative information that can be effectively used for understanding, organizing and retrieving text content. Though previous studies have provided many workable solutions for automated keyphrase extraction, they commonly divided the to-be-summarized content into multiple text chunks, then ranked and selected the most meaningful ones. These approaches could neither identify keyphrases that do not appear in the text, nor capture the real semantic meaning behind the text. We propose a generative model for keyphrase prediction with an encoder-decoder framework, which can effectively overcome the above drawbacks. We name it as *deep keyphrase generation* since it attempts to capture the deep semantic meaning of the content with a deep learning method. Empirical analysis on six datasets demonstrates that our proposed model not only achieves a significant performance boost on extracting keyphrases that appear in the source text, but also can generate absent keyphrases based on the semantic meaning of the text. Code and dataset are available at https://github.com/memray/seq2seq-keyphrase.

Attention-over-Attention Neural Networks for Reading Comprehension
Yiming Cui, Zhipeng Chen, Si Wei, Shijin Wang, Ting Liu, and Guoping Hu 11:27–11:45

Cloze-style reading comprehension is a representative problem in mining relationship between document and query. In this paper, we present a simple but novel model called attention-over-attention reader for better solving cloze-style reading comprehension task. The proposed model aims to place another attention mechanism over the document-level attention and induces "attended attention" for final answer predictions. One advantage of our model is that it is simpler than related works while giving excellent performance. In addition to the primary model, we also propose an N-best re-ranking strategy to double check the validity of the candidates and further improve the performance. Experimental results show that the proposed methods significantly outperform various state-of-the-art systems by a large margin in public datasets, such as CNN and Children's Book Test.

[TACL] Cross-Sentence N-ary Relation Extraction with Graph LSTMs
Nanyun Peng, Hoifung Poon, Chris Quirk, Kristina Toutanova, and Wen-tau Yih 11:46–12:04

Past work in relation extraction focuses on binary relations in single sentences. Recent NLP inroads in high-

valued domains have kindled strong interest in the more general setting of extracting n-ary relations that span multiple sentences. In this paper, we explore a general relation extraction framework based on graph long short-term memory networks (graph LSTMs), which can be easily extended to cross-sentence n-ary relation extraction. The graph formulation provides a unifying way to explore different LSTM approaches and incorporate various intra-sentential and inter-sentential dependencies, such as sequential, syntactic, and discourse relations. A robust contextual representation is learned for the entities, which serves as input to the relation classifier, making it easy for scaling to arbitrary relation arity n, as well as for multi-task learning with related relations. We evaluated this framework in two important domains in precision medicine and demonstrated its effectiveness with both supervised learning and distant supervision. Cross-sentence extraction produced far more knowledge, and multi-task learning significantly improved extraction accuracy. A thorough analysis comparing various LSTM approaches yielded interesting insight on how linguistic analysis impacts the performance.

Session Session 4B: Cognitive Modelling 1 / Vision 2

Alignment at Work: Using Language to Distinguish the Internalization and Self-Regulation Components of Cultural Fit in Organizations
Gabriel Doyle, Amir Goldberg, Sameer Srivastava, and Michael Frank 10:30–10:48

Cultural fit is widely believed to affect the success of individuals and the groups to which they belong. Yet it remains an elusive, poorly measured construct. Recent research draws on computational linguistics to measure cultural fit but overlooks asymmetries in cultural adaptation. By contrast, we develop a directed, dynamic measure of cultural fit based on linguistic alignment, which estimates the influence of one person's word use on another's and distinguishes between two enculturation mechanisms: internalization and self-regulation. We use this measure to trace employees' enculturation trajectories over a large, multi-year corpus of corporate emails and find that patterns of alignment in the first six months of employment are predictive of individuals' downstream outcomes, especially involuntary exit. Further predictive analyses suggest referential alignment plays an overlooked role in linguistic alignment.

Representations of language in a model of visually grounded speech signal
Grzegorz Chrupała, Lieke Gelderloos, and Afra Alishahi 10:49–11:07

We present a visually grounded model of speech perception which projects spoken utterances and images to a joint semantic space. We use a multi-layer recurrent highway network to model the temporal nature of spoken speech, and show that it learns to extract both form and meaning-based linguistic knowledge from the input signal. We carry out an in-depth analysis of the representations used by different components of the trained model and show that encoding of semantic aspects tends to become richer as we go up the hierarchy of layers, whereas encoding of form-related aspects of the language input tends to initially increase and then plateau or decrease.

Spectral Analysis of Information Density in Dialogue Predicts Collaborative Task Performance
Yang Xu and David Reitter 11:08–11:26

We propose a perspective on dialogue that focuses on relative information contributions of conversation partners as a key to successful communication. We predict the success of collaborative task in English and Danish corpora of task-oriented dialogue. Two features are extracted from the frequency domain representations of the lexical entropy series of each interlocutor, power spectrum overlap (PSO) and relative phase (RP). We find that PSO is a negative predictor of task success, while RP is a positive one. An SVM with these features significantly improved on previous task success prediction models. Our findings suggest that the strategic distribution of information density between interlocutors is relevant to task success.

[TACL] Modelling Semantic Expectation: Using Script Knowledge for Referent Prediction
Ashutosh Modi, Ivan Titov, Vera Demberg, Asad Sayeed, and Manfred Pinkal 11:27–11:45

Recent research in psycholinguistics has provided increasing evidence that humans predict upcoming content. Prediction also affects perception and might be a key to robustness in human language processing. In this paper, we investigate the factors that affect human prediction by building a computational model that can predict upcoming discourse referents based on linguistic knowledge alone vs. lin- guistic knowledge jointly with common-sense knowledge in the form of scripts. We find that script knowledge significantly improves model estimates of human predictions. In a second study, we test the highly controversial hypothesis that predictability influences referring expression type but do not find evidence for such an effect.

An Analysis of Action Recognition Datasets for Language and Vision Tasks
Spandana Gella and Frank Keller 11:46–12:04

A large amount of recent research has focused on tasks that combine language and vision, resulting in a proliferation of datasets and methods. One such task is action recognition, whose applications include image annotation, scene understanding and image retrieval. In this survey, we categorize the existing approaches based on how they conceptualize this problem and provide a detailed review of existing datasets, highlighting their diversity as well as advantages and disadvantages. We focus on recently developed datasets which link visual information with linguistic resources and provide a fine-grained syntactic and semantic analysis of actions in images.

Session Session 4C: Dialogue 2

Affect-LM: A Neural Language Model for Customizable Affective Text Generation
Sayan Ghosh, Mathieu Chollet, Eugene Laksana, Louis-Philippe Morency, and Stefan Scherer
10:30–10:48

Human verbal communication includes affective messages which are conveyed through use of emotionally colored words. There has been a lot of research effort in this direction but the problem of integrating state-of-the-art neural language models with affective information remains an area ripe for exploration. In this paper, we propose an extension to an LSTM (Long Short-Term Memory) language model for generation of conversational text, conditioned on affect categories. Our proposed model, Affect-LM enables us to customize the degree of emotional content in generated sentences through an additional design parameter. Perception studies conducted using Amazon Mechanical Turk show that Affect-LM can generate naturally looking emotional sentences without sacrificing grammatical correctness. Affect-LM also learns affect-discriminative word representations, and perplexity experiments show that additional affective information in conversational text can improve language model prediction.

Domain Attention with an Ensemble of Experts
Young-Bum Kim, Karl Stratos, and Dongchan Kim 10:49–11:07

An important problem in domain adaptation is to quickly generalize to a new domain with limited supervision given K existing domains. One approach is to retrain a global model across all K + 1 domains using standard techniques, for instance Daum´e III (2009). However, it is desirable to adapt without having to re-estimate a global model from scratch each time a new domain with potentially new intents and slots is added. We describe a solution based on attending an ensemble of domain experts. We assume K domain specific intent and slot models trained on respective domains. When given domain K + 1, our model uses a weighted combination of the K domain experts' feedback along with its own opinion to make predictions on the new domain. In experiments, the model significantly outperforms baselines that do not use domain adaptation and also performs better than the full retraining approach.

Learning Discourse-level Diversity for Neural Dialog Models using Conditional Variational Autoencoders
Tiancheng Zhao, Ran Zhao, and Maxine Eskenazi 11:08–11:26

While recent neural encoder-decoder models have shown great promise in modeling open-domain conversations, they often generate dull and generic responses. Unlike past work that has focused on diversifying the output of the decoder from word-level to alleviate this problem, we present a novel framework based on conditional variational autoencoders that capture the discourse-level diversity in the encoder. Our model uses latent variables to learn a distribution over potential conversational intents and generates diverse responses using only greedy decoders. We have further developed a novel variant that is integrated with linguistic prior knowledge for better performance. Finally, the training procedure is improved through introducing a bag-of-word loss. Our proposed models have been validated to generate significantly more diverse responses than baseline approaches and exhibit competence of discourse-level decision-making.

Hybrid Code Networks: practical and efficient end-to-end dialog control with supervised and reinforcement learning
Jason D Williams, Kavosh Asadi, and Geoffrey Zweig 11:27–11:45

End-to-end learning of recurrent neural networks (RNNs) is an attractive solution for dialog systems; however, current techniques are data-intensive and require thousands of dialogs to learn simple behaviors. We introduce Hybrid Code Networks (HCNs), which combine an RNN with domain-specific knowledge encoded as software and system action templates. Compared to existing end-to-end approaches, HCNs considerably reduce the amount of training data required, while retaining the key benefit of inferring a latent representation of dialog state. In addition, HCNs can be optimized with supervised learning, reinforcement learning, or a mixture of both. HCNs attain state-of-the-art performance on the bAbI dialog dataset (Bordes and Weston, 2016), and outperform two commercially deployed customer-facing dialog systems at our company.

Generating Contrastive Referring Expressions
Martin Villalba, Christoph Teichmann, and Alexander Koller 11:46–12:04

The referring expressions (REs) produced by a natural language generation (NLG) system can be misunderstood by the hearer, even when they are semantically correct. In an interactive setting, the NLG system can try to recognize such misunderstandings and correct them. We present an algorithm for generating corrective REs that use contrastive focus ("no, the BLUE button") to emphasize the information the hearer most

likely misunderstood. We show empirically that these contrastive REs are preferred over REs without contrast marking.

Session Session 4D: Machine Translation 2

Modeling Source Syntax for Neural Machine Translation

Junhui Li, Deyi Xiong, Zhaopeng Tu, Muhua Zhu, Min Zhang, and Guodong Zhou 10:30–10:48

Even though a linguistics-free sequence to sequence model in neural machine translation (NMT) has certain capability of implicitly learning syntactic information of source sentences, this paper shows that source syntax can be explicitly incorporated into NMT effectively to provide further improvements. Specifically, we linearize parse trees of source sentences to obtain structural label sequences. On the basis, we propose three different sorts of encoders to incorporate source syntax into NMT: 1) Parallel RNN encoder that learns word and label annotation vectors parallelly; 2) Hierarchical RNN encoder that learns word and label annotation vectors in a two-level hierarchy; and 3) Mixed RNN encoder that stitchingly learns word and label annotation vectors over sequences where words and labels are mixed. Experimentation on Chinese-to-English translation demonstrates that all the three proposed syntactic encoders are able to improve translation accuracy. It is interesting to note that the simplest RNN encoder, i.e., Mixed RNN encoder yields the best performance with an significant improvement of 1.4 BLEU points. Moreover, an in-depth analysis from several perspectives is provided to reveal how source syntax benefits NMT.

Sequence-to-Dependency Neural Machine Translation

Shuangzhi Wu, Dongdong Zhang, Nan Yang, Mu Li, and Ming Zhou 10:49–11:07

Nowadays a typical Neural Machine Translation (NMT) model generates translations from left to right as a linear sequence, during which latent syntactic structures of the target sentences are not explicitly concerned. Inspired by the success of using syntactic knowledge of target language for improving statistical machine translation, in this paper we propose a novel Sequence-to-Dependency Neural Machine Translation (SD-NMT) method, in which the target word sequence and its corresponding dependency structure are jointly constructed and modeled, and this structure is used as context to facilitate word generations. Experimental results show that the proposed method significantly outperforms state-of-the-art baselines on Chinese-English and Japanese-English translation tasks.

[TACL] Head-Lexicalized Bidirectional Tree LSTMs

Yue Teng Zhiyang and Zhang 11:08–11:26

Sequential LSTM has been extended to model tree structures, giving competitive results for a number of tasks. Existing methods model constituent trees by bottom-up combinations of constituent nodes, making direct use of input word information only for leaf nodes. This is different from sequential LSTMs, which contain reference to input words for each node. In this paper, we propose a method for automatic head-lexicalization for tree-structure LSTMs, propagating head words from leaf nodes to every constituent node. In addition, enabled by head lexicalization, we build a tree LSTM in the top-down direction, which corresponds to bidirectional sequential LSTM structurally. Experiments show that both extensions give better representations of tree structures. Our final model gives the best results on the Stanford Sentiment Treebank and highly competitive results on the TREC question type classification task.

[TACL] Pushing the Limits of Translation Quality Estimation

André F. T. Martins, Marcin Junczys-Dowmunt, Fabio N. Kepler, Ramón Astudillo, Chris Hokamp, and Roman Grundkiewicz 11:27–11:45

Translation quality estimation is a task of growing importance in NLP, due to its potential to reduce post-editing human effort in disruptive ways. However, this potential is currently limited by the relatively low accuracy of existing systems. In this paper, we achieve remarkable improvements by exploiting synergies between the related tasks of word-level quality estimation and automatic post-editing. First, we stack a new, carefully engineered, neural model into a rich feature-based word-level quality estimation system. Then, we use the output of an automatic post-editing system as an extra feature, obtaining striking results on WMT16: a word-level F 1 MULT score of 57.47% (an absolute gain of +7.95% over the current state of the art), and a Pearson correlation score of 65.56% for sentence-level HTER prediction (an absolute gain of +13.36%).

Learning to Parse and Translate Improves Neural Machine Translation

Akiko Eriguchi, Yoshimasa Tsuruoka, and Kyunghyun Cho 11:46–12:04

There has been relatively little attention to incorporating linguistic prior to neural machine translation. Much of the previous work was further constrained to considering linguistic prior on the source side. In this paper, we propose a hybrid model, called NMT+RNNG, that learns to parse and translate by combining the recurrent neural network grammar into the attention-based neural machine translation. Our approach encourages the neural machine translation model to incorporate linguistic prior during training, and lets it translate on

its own afterward. Extensive experiments with four language pairs show the effectiveness of the proposed NMT+RNNG.

Session Session 4E: Social Media 1

Detect Rumors in Microblog Posts Using Propagation Structure via Kernel Learning
Jing Ma, Wei Gao, and Kam-Fai Wong 10:30–10:48

How fake news goes viral via social media? How does its propagation pattern differ from real stories? In this paper, we attempt to address the problem of identifying rumors, i.e., fake information, out of microblog posts based on their propagation structure. We firstly model microblog posts diffusion with propagation trees, which provide valuable clues on how an original message is transmitted and developed over time. We then propose a kernel-based method called Propagation Tree Kernel, which captures high-order patterns differentiating different types of rumors by evaluating the similarities between their propagation tree structures. Experimental results on two real-world datasets demonstrate that the proposed kernel-based approach can detect rumors more quickly and accurately than state-of-the-art rumor detection models.

EmoNet: Fine-Grained Emotion Detection with Gated Recurrent Neural Networks
Muhammad Abdul-Mageed and Lyle Ungar 10:49–11:07

Accurate detection of emotion from natural language has applications ranging from building emotional chatbots to better understanding individuals and their lives. However, progress on emotion detection has been hampered by the absence of large labeled datasets. In this work, we build a very large dataset for fine-grained emotions and develop deep learning models on it. We achieve a new state-of-the-art on 24 fine-grained types of emotions (with an average accuracy of 87.58%). We also extend the task beyond emotion types to model Robert Plutick's 8 primary emotion dimensions, acquiring a superior accuracy of 95.68%.

Beyond Binary Labels: Political Ideology Prediction of Twitter Users
Daniel Preoţiuc-Pietro, Ye Liu, Daniel Hopkins, and Lyle Ungar 11:08–11:26

Automatic political orientation prediction from social media posts has to date proven successful only in distinguishing between publicly declared liberals and conservatives in the US. This study examines users' political ideology using a seven-point scale which enables us to identify politically moderate and neutral users — groups which are of particular interest to political scientists and pollsters. Using a novel data set with political ideology labels self-reported through surveys, our goal is two-fold: a) to characterize the groups of politically engaged users through language use on Twitter; b) to build a fine-grained model that predicts political ideology of unseen users. Our results identify differences in both political leaning and engagement and the extent to which each group tweets using political keywords. Finally, we demonstrate how to improve ideology prediction accuracy by exploiting the relationships between the user groups.

Leveraging Behavioral and Social Information for Weakly Supervised Collective Classification of Political Discourse on Twitter
Kristen Marie Johnson, Di Jin, and Dan Goldwasser 11:27–11:45

Framing is a political strategy in which politicians carefully word their statements in order to control public perception of issues. Previous works exploring political framing typically analyze frame usage in longer texts, such as congressional speeches. We present a collection of weakly supervised models which harness collective classification to predict the frames used in political discourse on the microblogging platform, Twitter. Our global probabilistic models show that by combining both lexical features of tweets and network-based behavioral features of Twitter, we are able to increase the average, unsupervised F1 score by 21.52 points over a lexical baseline alone.

On the Distribution of Lexical Features at Multiple Levels of Analysis
Fatemeh Almodaresi, Lyle Ungar, Vivek Kulkarni, Mohsen Zakeri, Salvatore Giorgi, and H. Andrew Schwartz 11:46–12:04

Natural language processing has increasingly moved from modeling documents and words toward studying the people behind the language. This move to working with data at the user or community level has presented the field with different characteristics of linguistic data. In this paper, we empirically characterize various lexical distributions at different levels of analysis, showing that, while most features are decidedly sparse and non-normal at the message-level (as with traditional NLP), they follow the central limit theorem to become much more Log-normal or even Normal at the user- and county-levels. Finally, we demonstrate that modeling lexical features for the correct level of analysis leads to marked improvements in common social scientific prediction tasks.

Session 5 Overview – Tuesday, August 1, 2017

	Track A	**Track B**	**Track C**	**Track D**	**Track E**
	Multidisciplinary 1	*Language and Resources 1*	*Syntax 2 (NN)*	*Machine Translation 3 (NN)*	*Sentiment 2*
	Salons B/C	Salons E/F	Salon D	Salon 1	Salons 2/3
13:30	Exploring Neural Text Simplification Models *S. Nisioi, S. Štajner, S. P. Ponzetto, and L. P. Dinu*	Polish evaluation dataset for compositional distributional semantics models *A. Wróblewska and K. Krasnowska-Kieraś*	A Minimal Span-Based Neural Constituency Parser *M. Stern, J. Andreas, and D. Klein*	Neural Machine Translation via Binary Code Prediction *Y. Oda, P. Arthur, G. Neubig, K. Yoshino, and S. Nakamura*	Context-Dependent Sentiment Analysis in User-Generated Videos *S. Poria, E. Cambria, D. Hazarika, N. Majumder, A. Zadeh, and L.-P. Morency*
13:49	A Nested Attention Neural Hybrid Model for Grammatical Error Correction *J. Ji, Q. Wang, K. Toutanova, Y. Gong, S. Truong, and J. Gao*	Automatic Annotation and Evaluation of Error Types for Grammatical Error Correction *C. Bryant, M. Felice, and T. Briscoe*	Semantic Dependency Parsing via Book Embedding *W. Sun, J. Cao, and X. Wan*	What do Neural Machine Translation Models Learn about Morphology? *Y. Belinkov, N. Durrani, F. Dalvi, H. Sajjad, and J. Glass*	A Multidimensional Lexicon for Interpersonal Stancetaking *U. Pavalanathan, J. Fitzpatrick, S. Kiesling, and J. Eisenstein*
14:08	TextFlow: A Text Similarity Measure based on Continuous Sequences *Y. Mrabet, H. Kilicoglu, and D. Demner-Fushman*	Evaluation Metrics for Machine Reading Comprehension: Prerequisite Skills and Readability *S. Sugawara, Y. Kido, H. Yokono, and A. Aizawa*	Neural Word Segmentation with Rich Pretraining *J. Yang, Y. Zhang, and F. Dong*	[TACL] Fully Character-Level Neural Machine Translation without Explicit Segmentation *J. Lee, K. Cho, and T. Hofmann*	Learning Lexico-Functional Patterns for First-Person Affect *L. Reed, J. Wu, S. Oraby, P. Anand, and M. Walker*
14:27	Friendships, Rivalries, and Trysts: Characterizing Relations between Ideas in Texts *C. Tan, D. Card, and N. A. Smith*	Sentence Alignment Methods for Improving Text Simplification Systems *S. Štajner, M. Franco-Salvador, S. P. Ponzetto, P. Rosso, and H. Stuckenschmidt*	Arc-swift: A Novel Transition System for Dependency Parsing *P. Qi and C. D. Manning*	Hybrid Neural Network Alignment and Lexicon Model in Direct HMM for Statistical Machine Translation *W. Wang, T. Alkhouli, D. Zhu, and H. Ney*	Lifelong Learning CRF for Supervised Aspect Extraction *L. Shu, H. Xu, and B. Liu*
14:40	On the Challenges of Translating NLP Research into Commercial Products *D. Dahlmeier*	Understanding Task Design Trade-offs in Crowdsourced Paraphrase Collection *Y. Jiang, J. K. Kummerfeld, and W. S. Lasecki*	A Generative Parser with a Discriminative Recognition Algorithm *J. Cheng, A. Lopez, and M. Lapata*	Towards String-To-Tree Neural Machine Translation *R. Aharoni and Y. Goldberg*	Exploiting Domain Knowledge via Grouped Weight Sharing with Application to Text Categorization *Y. Zhang, M. Lease, and B. C. Wallace*

Parallel Session 5

Session Session 5A: Multidisciplinary 1

Exploring Neural Text Simplification Models

Sergiu Nisioi, Sanja Štajner, Simone Paolo Ponzetto, and Liviu P. Dinu 13:30–13:48

We present the first attempt at using sequence to sequence neural networks to model text simplification (TS). Unlike the previously proposed automated TS systems, our neural text simplification (NTS) systems are able to simultaneously perform lexical simplification and content reduction. An extensive human evaluation of the output has shown that NTS systems achieve almost perfect grammaticality and meaning preservation of output sentences and higher level of simplification than the state-of-the-art automated TS systems

A Nested Attention Neural Hybrid Model for Grammatical Error Correction

Jianshu Ji, Qinlong Wang, Kristina Toutanova, Yongen Gong, Steven Truong, and Jianfeng Gao 13:49–14:07

Grammatical error correction (GEC) systems strive to correct both global errors inword order and usage, and local errors inspelling and inflection. Further developing upon recent work on neural machine translation, we propose a new hybrid neural model with nested attention layers for GEC.Experiments show that the new model can effectively correct errors of both types by incorporating word and character-level information, and that the model significantly outperforms previous neural models for GEC as measured on the standard CoNLL-14 benchmark dataset.Further analysis also shows that the superiority of the proposed model can be largely attributed to the use of the nested attention mechanism, which has proven particularly effective incorrecting local errors that involve small edits in orthography.

TextFlow: A Text Similarity Measure based on Continuous Sequences

Yassine Mrabet, Halil Kilicoglu, and Dina Demner-Fushman 14:08–14:26

Text similarity measures are used in multiple tasks such as plagiarism detection, information ranking and recognition of paraphrases and textual entailment. While recent advances in deep learning highlighted the relevance of sequential models in natural language generation, existing similarity measures do not fully exploit the sequential nature of language. Examples of such similarity measures include n-grams and skip-grams overlap which rely on distinct slices of the input texts. In this paper we present a novel text similarity measure inspired from a common representation in DNA sequence alignment algorithms. The new measure, called TextFlow, represents input text pairs as continuous curves and uses both the actual position of the words and sequence matching to compute the similarity value. Our experiments on 8 different datasets show very encouraging results in paraphrase detection, textual entailment recognition and ranking relevance.

Friendships, Rivalries, and Trysts: Characterizing Relations between Ideas in Texts

Chenhao Tan, Dallas Card, and Noah A. Smith 14:27–14:39

Understanding how ideas relate to each other is a fundamental question in many domains, ranging from intellectual history to public communication. Because ideas are naturally embedded in texts, we propose the first framework to systematically characterize the relations between ideas based on their occurrence in a corpus of documents, independent of how these ideas are represented. Combining two statistics—cooccurrence within documents and prevalence correlation over time—our approach reveals a number of different ways in which ideas can cooperate and compete. For instance, two ideas can closely track each other's prevalence over time, and yet rarely cooccur, almost like a "cold war" scenario. We observe that pairwise cooccurrence and prevalence correlation exhibit different distributions. We further demonstrate that our approach is able to uncover intriguing relations between ideas through in-depth case studies on news articles and research papers.

On the Challenges of Translating NLP Research into Commercial Products

Daniel Dahlmeier 14:40–15:02

This paper highlights challenges in industrial research related to translating research in natural language processing into commercial products. While the interest in natural language processing from industry is significant, the transfer of research to commercial products is non-trivial and its challenges are often unknown to or underestimated by many researchers. I discuss current obstacles and provide suggestions for increasing the chances for translating research to commercial success based on my experience in industrial research.

Session Session 5B: Language and Resources 1

Polish evaluation dataset for compositional distributional semantics models
Alina Wróblewska and Katarzyna Krasnowska-Kieraś 13:30–13:48

The paper presents a procedure of building an evaluation dataset. for the validation of compositional distributional semantics models estimated for languages other than English. The procedure generally builds on steps designed to assemble the SICK corpus, which contains pairs of English sentences annotated for semantic relatedness and entailment, because we aim at building a comparable dataset. However, the implementation of particular building steps significantly differs from the original SICK design assumptions, which is caused by both lack of necessary extraneous resources for an investigated language and the need for language-specific transformation rules. The designed procedure is verified on Polish, a fusional language with a relatively free word order, and contributes to building a Polish evaluation dataset. The resource consists of 10K sentence pairs which are human-annotated for semantic relatedness and entailment. The dataset may be used for the evaluation of compositional distributional semantics models of Polish.

Automatic Annotation and Evaluation of Error Types for Grammatical Error Correction
Christopher Bryant, Mariano Felice, and Ted Briscoe 13:49–14:07

Until now, error type performance for Grammatical Error Correction (GEC) systems could only be measured in terms of recall because system output is not annotated. To overcome this problem, we introduce ERRANT, a grammatical ERRor ANnotation Toolkit designed to automatically extract edits from parallel original and corrected sentences and classify them according to a new, dataset-agnostic, rule-based framework. This not only facilitates error type evaluation at different levels of granularity, but can also be used to reduce annotator workload and standardise existing GEC datasets. Human experts rated the automatic edits as "Good" or "Acceptable" in at least 95% of cases, so we applied ERRANT to the system output of the CoNLL-2014 shared task to carry out a detailed error type analysis for the first time.

Evaluation Metrics for Machine Reading Comprehension: Prerequisite Skills and Readability
Saku Sugawara, Yusuke Kido, Hikaru Yokono, and Akiko Aizawa 14:08–14:26

Knowing the quality of reading comprehension (RC) datasets is important for the development of natural-language understanding systems. In this study, two classes of metrics were adopted for evaluating RC datasets: prerequisite skills and readability. We applied these classes to six existing datasets, including MCTest and SQuAD, and highlighted the characteristics of the datasets according to each metric and the correlation between the two classes. Our dataset analysis suggests that the readability of RC datasets does not directly affect the question difficulty and that it is possible to create an RC dataset that is easy to read but difficult to answer.

Sentence Alignment Methods for Improving Text Simplification Systems
Sanja Štajner, Marc Franco-Salvador, Simone Paolo Ponzetto, Paolo Rosso, and Heiner Stuckenschmidt 14:27–14:39

We provide several methods for sentence-alignment of texts with different complexity levels. Using the best of them, we sentence-align the Newsela corpora, thus providing large training materials for automatic text simplification (ATS) systems. We show that using this dataset, even the standard phrase-based statistical machine translation models for ATS can outperform the state-of-the-art ATS systems.

Understanding Task Design Trade-offs in Crowdsourced Paraphrase Collection
Youxuan Jiang, Jonathan K. Kummerfeld, and Walter S. Lasecki 14:40–15:02

Linguistically diverse datasets are critical for training and evaluating robust machine learning systems, but data collection is a costly process that often requires experts. Crowdsourcing the process of paraphrase generation is an effective means of expanding natural language datasets, but there has been limited analysis of the trade-offs that arise when designing tasks. In this paper, we present the first systematic study of the key factors in crowdsourcing paraphrase collection. We consider variations in instructions, incentives, data domains, and workflows. We manually analyzed paraphrases for correctness, grammaticality, and linguistic diversity. Our observations provide new insight into the trade-offs between accuracy and diversity in crowd responses that arise as a result of task design, providing guidance for future paraphrase generation procedures.

Session Session 5C: Syntax 2 (NN)

A Minimal Span-Based Neural Constituency Parser
Mitchell Stern, Jacob Andreas, and Dan Klein 13:30–13:48

In this work, we present a minimal neural model for constituency parsing based on independent scoring of labels and spans. We show that this model is not only compatible with classical dynamic programming techniques, but also admits a novel greedy top-down inference algorithm based on recursive partitioning of the input. We demonstrate empirically that both prediction schemes are competitive with recent work, and when combined with basic extensions to the scoring model are capable of achieving state-of-the-art single-model performance on the Penn Treebank (91.79 F1) and strong performance on the French Treebank (82.23 F1).

Semantic Dependency Parsing via Book Embedding
Weiwei Sun, Junjie Cao, and Xiaojun Wan 13:49–14:07

We model a dependency graph as a book, a particular kind of topological space, for semantic dependency parsing. The spine of the book is made up of a sequence of words, and each page contains a subset of noncrossing arcs. To build a semantic graph for a given sentence, we design new Maximum Subgraph algorithms to generate noncrossing graphs on each page, and a Lagrangian Relaxation-based algorithm tocombine pages into a book. Experiments demonstrate the effectiveness of the bookembedding framework across a wide range of conditions. Our parser obtains comparable results with a state-of-the-art transition-based parser.

Neural Word Segmentation with Rich Pretraining
Jie Yang, Yue Zhang, and Fei Dong 14:08–14:26

Neural word segmentation research has benefited from large-scale raw texts by leveraging them for pretraining character and word embeddings. On the other hand, statistical segmentation research has exploited richer sources of external information, such as punctuation, automatic segmentation and POS. We investigate the effectiveness of a range of external training sources for neural word segmentation by building a modular segmentation model, pretraining the most important submodule using rich external sources. Results show that such pretraining significantly improves the model, leading to accuracies competitive to the best methods on six benchmarks.

Arc-swift: A Novel Transition System for Dependency Parsing
Peng Qi and Christopher D. Manning 14:27–14:39

Transition-based dependency parsers often need sequences of local shift and reduce operations to produce certain attachments. Correct individual decisions hence require global information about the sentence context and mistakes cause error propagation. This paper proposes a novel transition system, arc-swift, that enables direct attachments between tokens farther apart with a single transition. This allows the parser to leverage lexical information more directly in transition decisions. Hence, arc-swift can achieve significantly better performance with a very small beam size. Our parsers reduce error by 3.7–7.6% relative to those using existing transition systems on the Penn Treebank dependency parsing task and English Universal Dependencies.

A Generative Parser with a Discriminative Recognition Algorithm
Jianpeng Cheng, Adam Lopez, and Mirella Lapata 14:40–15:02

Generative models defining joint distributions over parse trees and sentences are useful for parsing and language modeling, but impose restrictions on the scope of features and are often outperformed by discriminative models. We propose a framework for parsing and language modeling which marries a generative model with a discriminative recognition model in an encoder-decoder setting. We provide interpretations of the framework based on expectation maximization and variational inference, and show that it enables parsing and language modeling within a single implementation. On the English Penn Treen-bank, our framework obtains competitive performance on constituency parsing while matching the state-of-the-art single- model language modeling score.

Session Session 5D: Machine Translation 3 (NN)

Neural Machine Translation via Binary Code Prediction
Yusuke Oda, Philip Arthur, Graham Neubig, Koichiro Yoshino, and Satoshi Nakamura 13:30–13:48

In this paper, we propose a new method for calculating the output layer in neural machine translation systems. The method is based on predicting a binary code for each word and can reduce computation time/memory requirements of the output layer to be logarithmic in vocabulary size in the best case. In addition, we also introduce two advanced approaches to improve the robustness of the proposed model: using error-correcting codes and combining softmax and binary codes. Experiments on two English-Japanese bidirectional translation tasks show proposed models achieve BLEU scores that approach the softmax, while reducing memory usage to the order of less than 1/10 and improving decoding speed on CPUs by x5 to x10.

What do Neural Machine Translation Models Learn about Morphology?
Yonatan Belinkov, Nadir Durrani, Fahim Dalvi, Hassan Sajjad, and James Glass 13:49–14:07

Neural machine translation (MT) models obtain state-of-the-art performance while maintaining a simple, end-to-end architecture. However, little is known about what these models learn about source and target languages during the training process. In this work, we analyze the representations learned by neural MT models at various levels of granularity and empirically evaluate the quality of the representations for learning morphology through extrinsic part-of-speech and morphological tagging tasks. We conduct a thorough investigation along several parameters: word-based vs. character-based representations, depth of the encoding layer, the identity of the target language, and encoder vs. decoder representations. Our data-driven, quantitative evaluation sheds light on important aspects in the neural MT system and its ability to capture word structure.

[TACL] Fully Character-Level Neural Machine Translation without Explicit Segmentation
Jason Lee, Kyunghyun Cho, and Thomas Hofmann 14:08–14:26

Most existing machine translation systems operate at the level of words, relying on explicit segmentation to extract tokens. We introduce a neural machine translation (NMT) model that maps a source character sequence to a target character sequence without any segmentation. We employ a character-level convolutional network with max-pooling at the encoder to reduce the length of source representation, allowing the model to be trained at a speed comparable to subword-level models while capturing local regularities. Our character-to-character model outperforms a recently proposed baseline with a subword-level encoder on WMT'15 DE-EN and CS-EN, and gives comparable performance on FI-EN and RU-EN. We then demonstrate that it is possible to share a single character-level encoder across multiple languages by training a model on a many-to-one translation task. In this multilingual setting, the character-level encoder significantly outperforms the subword-level encoder on all the language pairs. We observe that on CS-EN, FI-EN and RU-EN, the quality of the multilingual character-level translation even surpasses the models specifically trained on that language pair alone, both in terms of BLEU score and human judgment.

Hybrid Neural Network Alignment and Lexicon Model in Direct HMM for Statistical Machine Translation
Weiyue Wang, Tamer Alkhouli, Derui Zhu, and Hermann Ney 14:27–14:39

Recently, the neural machine translation systems showed their promising performance and surpassed the phrase-based systems for most translation tasks. Retreating into conventional concepts machine translation while utilizing effective neural models is vital for comprehending the leap accomplished by neural machine translation over phrase-based methods. This work proposes a direct HMM with neural network-based lexicon and alignment models, which are trained jointly using the Baum-Welch algorithm. The direct HMM is applied to rerank the n-best list created by a state-of-the-art phrase-based translation system and it provides improvements by up to 1.0% Bleu scores on two different translation tasks.

Towards String-To-Tree Neural Machine Translation
Roee Aharoni and Yoav Goldberg 14:40–15:02

We present a simple method to incorporate syntactic information about the target language in a neural machine translation system by translating into linearized, lexicalized constituency trees. An experiment on the WMT16 German-English news translation task resulted in an improved BLEU score when compared to a syntax-agnostic NMT baseline trained on the same dataset. An analysis of the translations from the syntax-aware system shows that it performs more reordering during translation in comparison to the baseline. A small-scale human evaluation also showed an advantage to the syntax-aware system.

Session Session 5E: Sentiment 2

Context-Dependent Sentiment Analysis in User-Generated Videos
Soujanya Poria, Erik Cambria, Devamanyu Hazarika, Navonil Majumder, Amir Zadeh, and Louis-Philippe Morency 13:30–13:48

Multimodal sentiment analysis is a developing area of research, which involves the identification of sentiments in videos. Current research considers utterances as independent entities, i.e., ignores the interdependencies and relations among the utterances of a video. In this paper, we propose a LSTM-based model that enables utterances to capture contextual information from their surroundings in the same video, thus aiding the classification process. Our method shows 5-10% performance improvement over the state of the art and high robustness to generalizability.

A Multidimensional Lexicon for Interpersonal Stancetaking
Umashanthi Pavalanathan, Jim Fitzpatrick, Scott Kiesling, and Jacob Eisenstein 13:49–14:07

The sociolinguistic construct of stancetaking describes the activities through which discourse participants create and signal relationships to their interlocutors, to the topic of discussion, and to the talk itself. Stancetaking underlies a wide range of interactional phenomena, relating to formality, politeness, affect, and subjectivity. We present a computational approach to stancetaking, in which we build a theoretically-motivated lexicon of stance markers, and then use multidimensional analysis to identify a set of underlying stance dimensions. We validate these dimensions intrinscially and extrinsically, showing that they are internally coherent, match pre-registered hypotheses, and correlate with social phenomena.

Learning Lexico-Functional Patterns for First-Person Affect
Lena Reed, Jiaqi Wu, Shereen Oraby, Pranav Anand, and Marilyn Walker 14:08–14:26

Informal first-person narratives are a unique resource for computational mod- els of everyday events and people's affective reactions to them. People blogging about their day tend not to explicitly say I am happy. Instead they describe situations from which other humans can readily infer their affective reactions. However current sentiment dictionaries are missing much of the information needed to make similar inferences. We build on recent work that models affect in terms of lexical predicate functions and affect on the predicate's arguments. We present a method to learn proxies for these functions from first- person narratives. We construct a novel fine-grained test set, and show that the pat- terns we learn improve our ability to pre- dict first-person affective reactions to everyday events, from a Stanford sentiment baseline of .67F to .75F.

Lifelong Learning CRF for Supervised Aspect Extraction
Lei Shu, Hu Xu, and Bing Liu 14:27–14:39

This paper makes a focused contribution to supervised aspect extraction. It shows that if the system has performed aspect extraction from many past domains and retained their results as knowledge, Conditional Random Fields (CRF) can leverage this knowledge in a lifelong learning manner to extract in a new domain markedly better than the traditional CRF without using this prior knowledge. The key innovation is that even after CRF training, the model can still improve its extraction with experiences in its applications.

Exploiting Domain Knowledge via Grouped Weight Sharing with Application to Text Categorization
Ye Zhang, Matthew Lease, and Byron C. Wallace 14:40–15:02

A fundamental advantage of neural models for NLP is their ability to learn representations from scratch. However, in practice this often means ignoring existing external linguistic resources, e.g., WordNet or domain specific ontologies such as the Unified Medical Language System (UMLS). We propose a general, novel method for exploiting such resources via weight sharing. Prior work on weight sharing in neural networks has considered it largely as a means of model compression. In contrast, we treat weight sharing as a flexible mechanism for incorporating prior knowledge into neural models. We show that this approach consistently yields improved performance on classification tasks compared to baseline strategies that do not exploit weight sharing.

Session 6 Overview – Tuesday, August 1, 2017

	Track A *Information Extraction 4* Salons B/C	Track B *Semantics 2 (NN)* Salons E/F	Track C *Discourse 2 / Dialogue 3* Salon D	Track D *Machine Learning 2* Salon 1	Track E *Summarization 1* Salons 2/3
15:25	Tandem Anchoring: a Multiword Anchor Approach for Interactive Topic Modeling *J. Lund, C. Cook, K. Seppi, and J. Boyd-Graber*	Naturalizing a Programming Language via Interactive Learning *S. I. Wang, S. Ginn, P. Liang, and C. D. Manning*	Joint Modeling of Content and Discourse Relations in Dialogues *K. Qin, L. Wang, and J. Kim*	Bayesian Modeling of Lexical Resources for Low-Resource Settings *N. Andrews, M. Dredze, B. Van Durme, and J. Eisner*	Diversity driven attention model for query-based abstractive summarization *P. Nema, M. M. Khapra, A. Laha, and B. Ravindran*
15:44	Apples to Apples: Learning Semantics of Common Entities Through a Novel Comprehension Task *O. Bakhshandeh and J. Allen*	Semantic Word Clusters Using Signed Spectral Clustering *J. Sedoc, J. Gallier, D. Foster, and L. Ungar*	Argument Mining with Structured SVMs and RNNs *V. Niculae, J. Park, and C. Cardie*	Semi-Supervised QA with Generative Domain-Adaptive Nets *Z. Yang, J. Hu, R. Salakhutdinov, and W. Cohen*	Get To The Point: Summarization with Pointer-Generator Networks *A. See, P. J. Liu, and C. D. Manning*
16:03	Going out on a limb: Joint Extraction of Entity Mentions and Relations without Dependency Trees *A. Katiyar and C. Cardie*	An Interpretable Knowledge Transfer Model for Knowledge Base Completion *Q. Xie, X. Ma, Z. Dai, and E. Hovy*	Neural Discourse Structure for Text Categorization *Y. Ji and N. A. Smith*	From Language to Programs: Bridging Reinforcement Learning and Maximum Marginal Likelihood *K. Guu, P. Pasupat, E. Liu, and P. Liang*	Supervised Learning of Automatic Pyramid for Optimization-Based Multi-Document Summarization *M. Peyrard and J. Eckle-Kohler*
16:22	[TACL] Evaluating Visual Representations for Topic Understanding and Their Effects on Manually Generated Labels *A. Smith, T. Y. Lee, F. Poursabzi-Sangdeh, J. Boyd-Graber, N. Elmqvist, and L. Findlater*	Learning a Neural Semantic Parser from User Feedback *S. Iyer, I. Konstas, A. Cheung, J. Krishnamurthy, and L. Zettlemoyer*	Adversarial Connective-exploiting Networks for Implicit Discourse Relation Classification *L. Qin, Z. Zhang, H. Zhao, Z. Hu, and E. Xing*	Information-Theory Interpretation of the Skip-Gram Negative-Sampling Objective Function *O. Melamud and J. Goldberger*	Selective Encoding for Abstractive Sentence Summarization *Q. Zhou, N. Yang, F. Wei, and M. Zhou*
16:41	Improving Neural Parsing by Disentangling Model Combination and Reranking Effects *D. Fried, M. Stern, and D. Klein*	[TACL] Enriching Word Vectors with Subword Information *P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov*	Don't understand a measure? Learn it: Structured Prediction for Coreference Resolution optimizing its measures *I. Haponchyk and A. Moschitti*	Implicitly-Defined Neural Networks for Sequence Labeling *M. Kazi and B. Thompson*	PositionRank: An Unsupervised Approach to Keyphrase Extraction from Scholarly Documents *C. Florescu and C. Caragea*

Parallel Session 6

Session Session 6A: Information Extraction 4

Tandem Anchoring: a Multiword Anchor Approach for Interactive Topic Modeling

Jeffrey Lund, Connor Cook, Kevin Seppi, and Jordan Boyd-Graber 15:25–15:43

Interactive topic models are powerful tools for those seeking to understand large collections of text. However, existing sampling-based interactive topic modeling approaches scale poorly to large data sets. Anchor methods, which use a single word to uniquely identify a topic, offer the speed needed for interactive work but lack both a mechanism to inject prior knowledge and lack the intuitive semantics needed for user-facing applications. We propose combinations of words as anchors, go- ing beyond existing single word anchor algorithms—an approach we call "Tan- dem Anchors". We begin with a synthetic investigation of this approach then apply the approach to interactive topic modeling in a user study and compare it to interac- tive and non-interactive approaches. Tan- dem anchors are faster and more intuitive than existing interactive approaches.

Apples to Apples: Learning Semantics of Common Entities Through a Novel Comprehension Task

Omid Bakhshandeh and James Allen 15:44–16:02

Understanding common entities and their attributes is a primary requirement for any system that comprehends natural language. In order to enable learning about common entities, we introduce a novel machine comprehension task, GuessTwo: given a short paragraph comparing different aspects of two real-world semantically-similar entities, a system should guess what those entities are. Accomplishing this task requires deep language understanding which enables inference, connecting each comparison paragraph to different levels of knowledge about world entities and their attributes. So far we have crowdsourced a dataset of more than 14K comparison paragraphs comparing entities from a variety of categories such as fruits and animals. We have designed two schemes for evaluation: open-ended, and binary-choice prediction. For benchmarking further progress in the task, we have collected a set of paragraphs as the test set on which human can accomplish the task with an accuracy of 94.2% on open-ended prediction. We have implemented various models for tackling the task, ranging from semantic-driven to neural models. The semantic-driven approach outperforms the neural models, however, the results indicate that the task is very challenging across the models.

Going out on a limb: Joint Extraction of Entity Mentions and Relations without Dependency Trees

Arzoo Katiyar and Claire Cardie 16:03–16:21

We present a novel attention-based recurrent neural network for joint extraction of entity mentions and relations. We show that attention along with long short term memory (LSTM) network can extract semantic relations between entity mentions without having access to dependency trees. Experiments on Automatic Content Extraction (ACE) corpora show that our model significantly outperforms feature-based joint model by Li and Ji (2014). We also compare our model with an end-to-end tree-based LSTM model (SPTree) by Miwa and Bansal (2016) and show that our model performs within 1% on entity mentions and 2% on relations. Our fine-grained analysis also shows that our model performs significantly better on Agent-Artifact relations, while SPTree performs better on Physical and Part-Whole relations.

[TACL] Evaluating Visual Representations for Topic Understanding and Their Effects on Manually Generated Labels

Alison Smith, Tak Yeon Lee, Forough Poursabzi-Sangdeh, Jordan Boyd-Graber, Niklas Elmqvist, and Leah Findlater 16:22–16:40

Probabilistic topic models are important tools for indexing, summarizing, and analyzing large document collections by their themes. However, promoting end-user understanding of topics remains an open research problem. We compare labels generated by users given four topic visualization techniques—word lists, word lists with bars, word clouds, and network graphs—against each other and against automatically generated labels. Our basis of comparison is participant ratings of how well labels describe documents from the topic. Our study has two phases: a labeling phase where participants label visualized topics and a validation phase where different participants select which labels best describe the topics' documents. Although all visualizations produce similar quality labels, simple visualizations such as word lists allow participants to quickly understand topics, while complex visualizations take longer but expose multi-word expressions that simpler visualizations obscure. Automatic labels lag behind user-created labels, but our dataset of manually labeled topics highlights linguistic patterns (e.g., hypernyms, phrases) that can be used to improve automatic topic labeling algorithms.

Improving Neural Parsing by Disentangling Model Combination and Reranking Effects
Daniel Fried, Mitchell Stern, and Dan Klein 16:41–17:00

Recent work has proposed several generative neural models for constituency parsing that achieve state-of-the-art results. Since direct search in these generative models is difficult, they have primarily been used to rescore candidate outputs from base parsers in which decoding is more straightforward. We first present an algorithm for direct search in these generative models. We then demonstrate that the rescoring results are at least partly due to implicit model combination rather than reranking effects. Finally, we show that explicit model combination can improve performance even further, resulting in new state-of-the-art numbers on the PTB of 94.25 F1 when training only on gold data and 94.66 F1 when using external data.

Session Session 6B: Semantics 2 (NN)

Naturalizing a Programming Language via Interactive Learning
Sida I. Wang, Samuel Ginn, Percy Liang, and Christopher D. Manning 15:25–15:43

Our goal is to create a convenient natural language interface for performing well-specified but complex actions such as analyzing data, manipulating text, and querying databases. However, existing natural language interfaces for such tasks are quite primitive compared to the power one wields with a programming language. To bridge this gap, we start with a core programming language and allow users to "naturalize" the core language incrementally by defining alternative, more natural syntax and increasingly complex concepts in terms of compositions of simpler ones. In a voxel world, we show that a community of users can simultaneously teach a common system a diverse language and use it to build hundreds of complex voxel structures. Over the course of three days, these users went from using only the core language to using the naturalized language in 85.9% of the last 10K utterances.

Semantic Word Clusters Using Signed Spectral Clustering
Joao Sedoc, Jean Gallier, Dean Foster, and Lyle Ungar 15:44–16:02

Vector space representations of words capture many aspects of word similarity, but such methods tend to produce vector spaces in which antonyms (as well as synonyms) are close to each other. For spectral clustering using such word embeddings, words are points in a vector space where synonyms are linked with positive weights, while antonyms are linked with negative weights. We present a new signed spectral normalized graph cut algorithm, *signed clustering*, that overlays existing thesauri upon distributionally derived vector representations of words, so that antonym relationships between word pairs are represented by negative weights. Our signed clustering algorithm produces clusters of words that simultaneously capture distributional and synonym relations. By using randomized spectral decomposition (Halko et al., 2011) and sparse matrices, our method is both fast and scalable. We validate our clusters using datasets containing human judgments of word pair similarities and show the benefit of using our word clusters for sentiment prediction.

An Interpretable Knowledge Transfer Model for Knowledge Base Completion
Qizhe Xie, Xuezhe Ma, Zihang Dai, and Eduard Hovy 16:03–16:21

Knowledge bases are important resources for a variety of natural language processing tasks but suffer from incompleteness. We propose a novel embedding model, ITransF, to perform knowledge base completion. Equipped with a sparse attention mechanism, ITransF discovers hidden concepts of relations and transfer statistical strength through the sharing of concepts. Moreover, the learned associations between relations and concepts, which are represented by sparse attention vectors, can be interpreted easily. We evaluate ITransF on two benchmark datasets—WN18 and FB15k for knowledge base completion and obtains improvements on both the mean rank and Hits10 metrics, over all baselines that do not use additional information.

Learning a Neural Semantic Parser from User Feedback
Srinivasan Iyer, Ioannis Konstas, Alvin Cheung, Jayant Krishnamurthy, and Luke Zettlemoyer 16:22–16:40

We present an approach to rapidly and easily build natural language interfaces to databases for new domains, whose performance improves over time based on user feedback, and requires minimal intervention. To achieve this, we adapt neural sequence models to map utterances directly to SQL with its full expressivity, bypassing any intermediate meaning representations. These models are immediately deployed online to solicit feedback from real users to flag incorrect queries. Finally, the popularity of SQL facilitates gathering annotations for incorrect predictions using the crowd, which is directly used to improve our models. This complete feedback loop, without intermediate representations or database specific engineering, opens up new ways of building high quality semantic parsers. Experiments suggest that this approach can be deployed quickly for any new target domain, as we show by learning a semantic parser for an online academic database from scratch.

[TACL] Enriching Word Vectors with Subword Information
Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov 16:41–17:00

Continuous word representations, trained on large unlabeled corpora are useful for many natural language processing tasks. Popular models to learn such representations ignore the morphology of words, by assigning a distinct vector to each word. This is a limitation, especially for languages with large vocabularies and many rare words. In this paper, we propose a new approach based on the skipgram model, where each word is represented as a bag of character n-grams. A vector representation is associated to each character n-gram, words being represented as the sum of these representations. Our method is fast, allowing to train models on large corpora quickly and allows to compute word representations for words that did not appear in the training

data. We evaluate our word representations on nine different languages, both on word similarity and analogy tasks. By comparing to recently proposed morphological word representations, we show that our vectors achieve state-of-the-art performance on these tasks.

Session Session 6C: Discourse 2 / Dialogue 3

Joint Modeling of Content and Discourse Relations in Dialogues
Kechen Qin, Lu Wang, and Joseph Kim 15:25–15:43

We present a joint modeling approach to identify salient discussion points in spoken meetings as well as to label the discourse relations between speaker turns. A variation of our model is also discussed when discourse relations are treated as latent variables. Experimental results on two popular meeting corpora show that our joint model can outperform state-of-the-art approaches for both phrase-based content selection and discourse relation prediction tasks. We also evaluate our model on predicting the consistency among team members' understanding of their group decisions. Classifiers trained with features constructed from our model achieve significant better predictive performance than the state-of-the-art.

Argument Mining with Structured SVMs and RNNs
Vlad Niculae, Joonsuk Park, and Claire Cardie 15:44–16:02

We propose a novel factor graph model for argument mining, designed for settings in which the argumentative relations in a document do not necessarily form a tree structure. (This is the case in over 20% of the web comments dataset we release.) Our model jointly learns elementary unit type classification and argumentative relation prediction. Moreover, our model supports SVM and RNN parametrizations, can enforce structure constraints (e.g., transitivity), and can express dependencies between adjacent relations and propositions. Our approaches outperform unstructured baselines in both web comments and argumentative essay datasets.

Neural Discourse Structure for Text Categorization
Yangfeng Ji and Noah A. Smith 16:03–16:21

We show that discourse structure, as defined by Rhetorical Structure Theory and provided by an existing discourse parser, benefits text categorization. Our approach uses a recursive neural network and a newly proposed attention mechanism to compute a representation of the text that focuses on salient content, from the perspective of both RST and the task. Experiments consider variants of the approach and illustrate its strengths and weaknesses.

Adversarial Connective-exploiting Networks for Implicit Discourse Relation Classification
Lianhui Qin, Zhisong Zhang, Hai Zhao, Zhiting Hu, and Eric Xing 16:22–16:40

Implicit discourse relation classification is of great challenge due to the lack of connectives as strong linguistic cues, which motivates the use of annotated implicit connectives to improve the recognition. We propose a feature imitation framework in which an implicit relation network is driven to learn from another neural network with access to connectives, and thus encouraged to extract similarly salient features for accurate classification. We develop an adversarial model to enable an adaptive imitation scheme through competition between the implicit network and a rival feature discriminator. Our method effectively transfers discriminability of connectives to the implicit features, and achieves state-of-the-art performance on the PDTB benchmark.

Don't understand a measure? Learn it: Structured Prediction for Coreference Resolution optimizing its measures
Iryna Haponchyk and Alessandro Moschitti 16:41–17:00

An interesting aspect of structured prediction is the evaluation of an output structure against the gold standard. Especially in the loss-augmented setting, the need of finding the max-violating constraint has severely limited the expressivity of effective loss functions. In this paper, we trade off exact computation for enabling the use and study of more complex loss functions for coreference resolution. Most interestingly, we show that such functions can be (i) automatically learned also from controversial but commonly accepted coreference measures, e.g., MELA, and (ii) successfully used in learning algorithms. The accurate model comparison on the standard CoNLL-2012 setting shows the benefit of more expressive loss functions.

Session Session 6D: Machine Learning 2

Bayesian Modeling of Lexical Resources for Low-Resource Settings
Nicholas Andrews, Mark Dredze, Benjamin Van Durme, and Jason Eisner 15:25–15:43

Lexical resources such as dictionaries and gazetteers are often used as auxiliary data for tasks such as part-of-speech induction and named-entity recognition. However, discriminative training with lexical features requires annotated data to reliably estimate the lexical feature weights and may result in overfitting the lexical features at the expense of features which generalize better. In this paper, we investigate a more robust approach: we stipulate that the lexicon is the result of an assumed generative process. Practically, this means that we may treat the lexical resources as observations under the proposed generative model. The lexical resources provide training data for the generative model without requiring separate data to estimate lexical feature weights. We evaluate the proposed approach in two settings: part-of-speech induction and low-resource named-entity recognition.

Semi-Supervised QA with Generative Domain-Adaptive Nets
Zhilin Yang, Junjie Hu, Ruslan Salakhutdinov, and William Cohen 15:44–16:02

We study the problem of semi-supervised question answering—-utilizing unlabeled text to boost the performance of question answering models. We propose a novel training framework, the *Generative Domain-Adaptive Nets*. In this framework, we train a generative model to generate questions based on the unlabeled text, and combine model-generated questions with human-generated questions for training question answering models. We develop novel domain adaptation algorithms, based on reinforcement learning, to alleviate the discrepancy between the model-generated data distribution and the human-generated data distribution. Experiments show that our proposed framework obtains substantial improvement from unlabeled text.

From Language to Programs: Bridging Reinforcement Learning and Maximum Marginal Likelihood
Kelvin Guu, Panupong Pasupat, Evan Liu, and Percy Liang 16:03–16:21

Our goal is to learn a semantic parser that maps natural language utterances into executable programs when only indirect supervision is available: examples are labeled with the correct execution result, but not the program itself. Consequently, we must search the space of programs for those that output the correct result, while not being misled by *spurious programs*: incorrect programs that coincidentally output the correct result. We connect two common learning paradigms, reinforcement learning (RL) and maximum marginal likelihood (MML), and then present a new learning algorithm that combines the strengths of both. The new algorithm guards against spurious programs by combining the systematic search traditionally employed in MML with the randomized exploration of RL, and by updating parameters such that probability is spread more evenly across consistent programs. We apply our learning algorithm to a new neural semantic parser and show significant gains over existing state-of-the-art results on a recent context-dependent semantic parsing task.

Information-Theory Interpretation of the Skip-Gram Negative-Sampling Objective Function
Oren Melamud and Jacob Goldberger 16:22–16:40

In this paper we define a measure of dependency between two random variables, based on the Jensen-Shannon (JS) divergence between their joint distribution and the product of their marginal distributions. Then, we show that word2vec's skip-gram with negative sampling embedding algorithm finds the optimal low-dimensional approximation of this JS dependency measure between the words and their contexts. The gap between the optimal score and the low-dimensional approximation is demonstrated on a standard text corpus.

Implicitly-Defined Neural Networks for Sequence Labeling
Michaeel Kazi and Brian Thompson 16:41–17:00

In this work, we propose a novel, implicitly-defined neural network architecture and describe a method to compute its components. The proposed architecture forgoes the causality assumption used to formulate recurrent neural networks and instead couples the hidden states of the network, allowing improvement on problems with complex, long-distance dependencies. Initial experiments demonstrate the new architecture outperforms both the Stanford Parser and baseline bidirectional networks on the Penn Treebank Part-of-Speech tagging task and a baseline bidirectional network on an additional artificial random biased walk task.

Session Session 6E: Summarization 1

Diversity driven attention model for query-based abstractive summarization
Preksha Nema, Mitesh M. Khapra, Anirban Laha, and Balaraman Ravindran 15:25–15:43

Abstractive summarization aims to generate a shorter version of the document covering all the salient points in a compact and coherent fashion. On the other hand, query-based summarization highlights those points that are relevant in the context of a given query. The encode-attend-decode paradigm has achieved notable success in machine translation, extractive summarization, dialog systems, etc. But it suffers from the drawback of generation of repeated phrases. In this work we propose a model for the query-based summarization task based on the encode-attend-decode paradigm with two key additions (i) a query attention model (in addition to document attention model) which learns to focus on different portions of the query at different time steps (instead of using a static representation for the query) and (ii) a new diversity based attention model which aims to alleviate the problem of repeating phrases in the summary. In order to enable the testing of this model we introduce a new query-based summarization dataset building on debatepedia. Our experiments show that with these two additions the proposed model clearly outperforms vanilla encode-attend-decode models with a gain of 28% (absolute) in ROUGE-L scores.

Get To The Point: Summarization with Pointer-Generator Networks
Abigail See, Peter J. Liu, and Christopher D. Manning 15:44–16:02

Neural sequence-to-sequence models have provided a viable new approach for abstractive text summarization (meaning they are not restricted to simply selecting and rearranging passages from the original text). However, these models have two shortcomings: they are liable to reproduce factual details inaccurately, and they tend to repeat themselves. In this work we propose a novel architecture that augments the standard sequence-to-sequence attentional model in two orthogonal ways. First, we use a hybrid pointer-generator network that can copy words from the source text via pointing, which aids accurate reproduction of information, while retaining the ability to produce novel words through the generator. Second, we use coverage to keep track of what has been summarized, which discourages repetition. We apply our model to the CNN / Daily Mail summarization task, outperforming the current abstractive state-of-the-art by at least 2 ROUGE points.

Supervised Learning of Automatic Pyramid for Optimization-Based Multi-Document Summarization
Maxime Peyrard and Judith Eckle-Kohler 16:03–16:21

We present a new supervised framework that learns to estimate automatic Pyramid scores and uses them for optimization-based extractive multi-document summarization. For learning automatic Pyramid scores, we developed a method for automatic training data generation which is based on a genetic algorithm using automatic Pyramid as the fitness function. Our experimental evaluation shows that our new framework significantly outperforms strong baselines regarding automatic Pyramid, and that there is much room for improvement in comparison with the upper-bound for automatic Pyramid.

Selective Encoding for Abstractive Sentence Summarization
Qingyu Zhou, Nan Yang, Furu Wei, and Ming Zhou 16:22–16:40

We propose a selective encoding model to extend the sequence-to-sequence framework for abstractive sentence summarization. It consists of a sentence encoder, a selective gate network, and an attention equipped decoder. The sentence encoder and decoder are built with recurrent neural networks. The selective gate network constructs a second level sentence representation by controlling the information flow from encoder to decoder. The second level representation is tailored for sentence summarization task, which leads to better performance. We evaluate our model on the English Gigaword, DUC 2004 and MSR abstractive sentence summarization datasets. The experimental results show that the proposed selective encoding model outperforms the state-of-the-art baseline models.

PositionRank: An Unsupervised Approach to Keyphrase Extraction from Scholarly Documents
Corina Florescu and Cornelia Caragea 16:41–17:00

The large and growing amounts of online scholarly data present both challenges and opportunities to enhance knowledge discovery. One such challenge is to automatically extract a small set of keyphrases from a document that can accurately describe the document's content and can facilitate fast information processing. In this paper, we propose PositionRank, an unsupervised model for keyphrase extraction from scholarly documents that incorporates information from all positions of a word's occurrences into a biased PageRank. Our model obtains remarkable improvements in performance over PageRank models that do not take into account word

positions as well as over strong baselines for this task. Specifically, on several datasets of research papers, PositionRank achieves improvements as high as 29.09%.

Poster Session P2 (Long Papers)

Time: *17:40–19:40* Location: Bayshore Grand Ballroom/Foyer

Learning Symmetric Collaborative Dialogue Agents with Dynamic Knowledge Graph Embeddings
He He, Anusha Balakrishnan, Mihail Eric, and Percy Liang

We study a *symmetric collaborative dialogue* setting in which two agents, each with private knowledge, must strategically communicate to achieve a common goal. The open-ended dialogue state in this setting poses new challenges for existing dialogue systems. We collected a dataset of 11K human-human dialogues, which exhibits interesting lexical, semantic, and strategic elements. To model both structured knowledge and unstructured language, we propose a neural model with dynamic knowledge graph embeddings that evolve as the dialogue progresses. Automatic and human evaluations show that our model is both more effective at achieving the goal and more human-like than baseline neural and rule-based models.

Neural Belief Tracker: Data-Driven Dialogue State Tracking
Nikola Mrkšić, Diarmuid Ó Séaghdha, Tsung-Hsien Wen, Blaise Thomson, and Steve Young

One of the core components of modern spoken dialogue systems is the belief tracker, which estimates the user's goal at every step of the dialogue. However, most current approaches have difficulty scaling to larger, more complex dialogue domains. This is due to their dependency on either: a) Spoken Language Understanding models that require large amounts of annotated training data; or b) hand-crafted lexicons for capturing some of the linguistic variation in users' language. We propose a novel Neural Belief Tracking (NBT) framework which overcomes these problems by building on recent advances in representation learning. NBT models reason over pre-trained word vectors, learning to compose them into distributed representations of user utterances and dialogue context. Our evaluation on two datasets shows that this approach surpasses past limitations, matching the performance of state-of-the-art models which rely on hand-crafted semantic lexicons and outperforming them when such lexicons are not provided.

Exploiting Argument Information to Improve Event Detection via Supervised Attention Mechanisms
Shulin Liu, Yubo Chen, Kang Liu, and Jun Zhao

This paper tackles the task of event detection (ED), which involves identifying and categorizing events. We argue that arguments provide significant clues to this task, but they are either completely ignored or exploited in an indirect manner in existing detection approaches. In this work, we propose to exploit argument information explicitly for ED via supervised attention mechanisms. In specific, we systematically investigate the proposed model under the supervision of different attention strategies. Experimental results show that our approach advances state-of-the-arts and achieves the best F1 score on ACE 2005 dataset.

Topical Coherence in LDA-based Models through Induced Segmentation
Hesam Amoualian, Wei Lu, Eric Gaussier, Georgios Balikas, Massih R Amini, and Marianne Clausel

This paper presents an LDA-based model that generates topically coherent segments within documents by jointly segmenting documents and assigning topics to their words. The coherence between topics is ensured through a copula, binding the topics associated to the words of a segment. In addition, this model relies on both document and segment specific topic distributions so as to capture fine grained differences in topic assignments. We show that the proposed model naturally encompasses other state-of-the-art LDA-based models designed for similar tasks. Furthermore, our experiments, conducted on six different publicly available datasets, show the effectiveness of our model in terms of perplexity, Normalized Pointwise Mutual Information, which captures the coherence between the generated topics, and the Micro F1 measure for text classification.

Jointly Extracting Relations with Class Ties via Effective Deep Ranking
Hai Ye, Wenhan Chao, Zhunchen Luo, and Zhoujun Li

Connections between relations in relation extraction, which we call class ties, are common. In distantly supervised scenario, one entity tuple may have multiple relation facts. Exploiting class ties between relations of one entity tuple will be promising for distantly supervised relation extraction. However, previous models are not effective or ignore to model this property. In this work, to effectively leverage class ties, we propose to make joint relation extraction with a unified model that integrates convolutional neural network (CNN) with a general pairwise ranking framework, in which three novel ranking loss functions are introduced. Additionally, an effective method is presented to relieve the severe class imbalance problem from NR (not relation)

for model training. Experiments on a widely used dataset show that leveraging class ties will enhance extraction and demonstrate the effectiveness of our model to learn class ties. Our model outperforms the baselines significantly, achieving state-of-the-art performance.

Search-based Neural Structured Learning for Sequential Question Answering
Mohit Iyyer, Wen-tau Yih, and Ming-Wei Chang

Recent work in semantic parsing for question answering has focused on long and complicated questions, many of which would seem unnatural if asked in a normal conversation between two humans. In an effort to explore a conversational QA setting, we present a more realistic task: answering sequences of simple but inter-related questions. We collect a dataset of 6,066 question sequences that inquire about semi-structured tables from Wikipedia, with 17,553 question-answer pairs in total. To solve this sequential question answering task, we propose a novel dynamic neural semantic parsing framework trained using a weakly supervised reward-guided search. Our model effectively leverages the sequential context to outperform state-of-the-art QA systems that are designed to answer highly complex questions.

Gated-Attention Readers for Text Comprehension
Bhuwan Dhingra, Hanxiao Liu, Zhilin Yang, William Cohen, and Ruslan Salakhutdinov

In this paper we study the problem of answering cloze-style questions over documents. Our model, the Gated-Attention (GA) Reader, integrates a multi-hop architecture with a novel attention mechanism, which is based on multiplicative interactions between the query embedding and the intermediate states of a recurrent neural network document reader. This enables the reader to build query-specific representations of tokens in the document for accurate answer selection. The GA Reader obtains state-of-the-art results on three benchmarks for this task–the CNN & Daily Mail news stories and the Who Did What dataset. The effectiveness of multiplicative interaction is demonstrated by an ablation study, and by comparing to alternative compositional operators for implementing the gated-attention.

Determining Gains Acquired from Word Embedding Quantitatively Using Discrete Distribution Clustering
Jianbo Ye, Yanran Li, Zhaohui Wu, James Z. Wang, Wenjie Li, and Jia Li

Word embeddings have become widely-used in document analysis. While a large number of models for mapping words to vector spaces have been developed, it remains undetermined how much net gain can be achieved over traditional approaches based on bag-of-words. In this paper, we propose a new document clustering approach by combining any word embedding with a state-of-the-art algorithm for clustering empirical distributions. By using the Wasserstein distance between distributions, the word-to-word semantic relationship is taken into account in a principled way. The new clustering method is easy to use and consistently outperforms other methods on a variety of data sets. More importantly, the method provides an effective framework for determining when and how much word embeddings contribute to document analysis. Experimental results with multiple embedding models are reported.

Towards a Seamless Integration of Word Senses into Downstream NLP Applications
Mohammad Taher Pilehvar, Jose Camacho-Collados, Roberto Navigli, and Nigel Collier

Lexical ambiguity can impede NLP systems from accurate understanding of semantics. Despite its potential benefits, the integration of sense-level information into NLP systems has remained understudied. By incorporating a novel disambiguation algorithm into a state-of-the-art classification model, we create a pipeline to integrate sense-level information into downstream NLP applications. We show that a simple disambiguation of the input text can lead to consistent performance improvement on multiple topic categorization and polarity detection datasets, particularly when the fine granularity of the underlying sense inventory is reduced and the document is sufficiently large. Our results also point to the need for sense representation research to focus more on in vivo evaluations which target the performance in downstream NLP applications rather than artificial benchmarks.

Reading Wikipedia to Answer Open-Domain Questions
Danqi Chen, Adam Fisch, Jason Weston, and Antoine Bordes

This paper proposes to tackle open-domain question answering using Wikipedia as the unique knowledge source: the answer to any factoid question is a text span in a Wikipedia article. This task of machine reading at scale combines the challenges of document retrieval (finding the relevant articles) with that of machine comprehension of text (identifying the answer spans from those articles). Our approach combines a search component based on bigram hashing and TF-IDF matching with a multi-layer recurrent neural network model trained to detect answers in Wikipedia paragraphs. Our experiments on multiple existing QA datasets indicate

that (1) both modules are highly competitive with respect to existing counterparts and (2) multitask learning using distant supervision on their combination is an effective complete system on this challenging task.

Learning to Skim Text
Adams Wei Yu, Hongrae Lee, and Quoc Le

Recurrent Neural Networks are showing much promise in many sub-areas of natural language processing, ranging from document classification to machine translation to automatic question answering. Despite their promise, many recurrent models have to read the whole text word by word, making it slow to handle long documents. For example, it is difficult to use a recurrent network to read a book and answer questions about it. In this paper, we present an approach of reading text while skipping irrelevant information if needed. The underlying model is a recurrent network that learns how far to jump after reading a few words of the input text. We employ a standard policy gradient method to train the model to make discrete jumping decisions. In our benchmarks on four different tasks, including number prediction, sentiment analysis, news article classification and automatic Q&A, our proposed model, a modified LSTM with jumping, is up to 6 times faster than the standard sequential LSTM, while maintaining the same or even better accuracy.

An Algebra for Feature Extraction
Vivek Srikumar

Though feature extraction is a necessary first step in statistical NLP, it is often seen as a mere preprocessing step. Yet, it can dominate computation time, both during training, and especially at deployment. In this paper, we formalize feature extraction from an algebraic perspective. Our formalization allows us to define a message passing algorithm that can restructure feature templates to be more computationally efficient. We show via experiments on text chunking and relation extraction that this restructuring does indeed speed up feature extraction in practice by reducing redundant computation.

Chunk-based Decoder for Neural Machine Translation
Shonosuke Ishiwatari, Jingtao Yao, Shujie Liu, Mu Li, Ming Zhou, Naoki Yoshinaga, Masaru Kitsuregawa, and Weijia Jia

Chunks (or phrases) once played a pivotal role in machine translation. By using a chunk rather than a word as the basic translation unit, local (intra-chunk) and global (inter-chunk) word orders and dependencies can be easily modeled. The chunk structure, despite its importance, has not been considered in the decoders used for neural machine translation (NMT). In this paper, we propose chunk-based decoders for (NMT), each of which consists of a chunk-level decoder and a word-level decoder. The chunk-level decoder models global dependencies while the word-level decoder decides the local word order in a chunk. To output a target sentence, the chunk-level decoder generates a chunk representation containing global information, which the word-level decoder then uses as a basis to predict the words inside the chunk. Experimental results show that our proposed decoders can significantly improve translation performance in a WAT '16 English-to-Japanese translation task.

Doubly-Attentive Decoder for Multi-modal Neural Machine Translation
Iacer Calixto, Qun Liu, and Nick Campbell

We introduce a Multi-modal Neural Machine Translation model in which a doubly-attentive decoder naturally incorporates spatial visual features obtained using pre-trained convolutional neural networks, bridging the gap between image description and translation. Our decoder learns to attend to source-language words and parts of an image independently by means of two separate attention mechanisms as it generates words in the target language. We find that our model can efficiently exploit not just back-translated in-domain multi-modal data but also large general-domain text-only MT corpora. We also report state-of-the-art results on the Multi30k data set.

A Teacher-Student Framework for Zero-Resource Neural Machine Translation
Yun Chen, Yang Liu, Yong Cheng, and Victor O.K. Li

While end-to-end neural machine translation (NMT) has made remarkable progress recently, it still suffers from the data scarcity problem for low-resource language pairs and domains. In this paper, we propose a method for zero-resource NMT by assuming that parallel sentences have close probabilities of generating a sentence in a third language. Based on the assumption, our method is able to train a source-to-target NMT model ("student") without parallel corpora available guided by an existing pivot-to-target NMT model ("teacher") on a source-pivot parallel corpus. Experimental results show that the proposed method significantly improves over a baseline pivot-based model by +3.0 BLEU points across various language pairs.

Improved Neural Machine Translation with a Syntax-Aware Encoder and Decoder
Huadong Chen, Shujian Huang, David Chiang, and Jiajun Chen

Most neural machine translation (NMT) models are based on the sequential encoder-decoder framework, which makes no use of syntactic information. In this paper, we improve this model by explicitly incorporating source-side syntactic trees. More specifically, we propose (1) a bidirectional tree encoder which learns both sequential and tree structured representations; (2) a tree-coverage model that lets the attention depend on the source-side syntax. Experiments on Chinese-English translation demonstrate that our proposed models outperform the sequential attentional model as well as a stronger baseline with a bottom-up tree encoder and word coverage.

Cross-lingual Name Tagging and Linking for 282 Languages
Xiaoman Pan, Boliang Zhang, Jonathan May, Joel Nothman, Kevin Knight, and Heng Ji

The ambitious goal of this work is to develop a cross-lingual name tagging and linking framework for 282 languages that exist in Wikipedia. Given a document in any of these languages, our framework is able to identify name mentions, assign a coarse-grained or fine-grained type to each mention, and link it to an English Knowledge Base (KB) if it is linkable. We achieve this goal by performing a series of new KB mining methods: generating "silver-standard" annotations by transferring annotations from English to other languages through cross-lingual links and KB properties, refining annotations through self-training and topic selection, deriving language-specific morphology features from anchor links, and mining word translation pairs from cross-lingual links. Both name tagging and linking results for 282 languages are promising on Wikipedia data and on-Wikipedia data.

Adversarial Training for Unsupervised Bilingual Lexicon Induction
Meng Zhang, Yang Liu, Huanbo Luan, and Maosong Sun

Word embeddings are well known to capture linguistic regularities of the language on which they are trained. Researchers also observe that these regularities can transfer across languages. However, previous endeavors to connect separate monolingual word embeddings typically require cross-lingual signals as supervision, either in the form of parallel corpus or seed lexicon. In this work, we show that such cross-lingual connection can actually be established without any form of supervision. We achieve this end by formulating the problem as a natural adversarial game, and investigating techniques that are crucial to successful training. We carry out evaluation on the unsupervised bilingual lexicon induction task. Even though this task appears intrinsically cross-lingual, we are able to demonstrate encouraging performance without any cross-lingual clues.

Estimating Code-Switching on Twitter with a Novel Generalized Word-Level Language Detection Technique
Shruti Rijhwani, Royal Sequiera, Monojit Choudhury, Kalika Bali, and Chandra Shekhar Maddila

Word-level language detection is necessary for analyzing code-switched text, where multiple languages could be mixed within a sentence. Existing models are restricted to code-switching between two specific languages and fail in real-world scenarios as text input rarely has a priori information on the languages used. We present a novel unsupervised word-level language detection technique for code-switched text for an arbitrarily large number of languages, which does not require any manually annotated training data. Our experiments with tweets in seven languages show a 74% relative error reduction in word-level labeling with respect to competitive baselines. We then use this system to conduct a large-scale quantitative analysis of code-switching patterns on Twitter, both global as well as region-specific, with 58M tweets.

Using Global Constraints and Reranking to Improve Cognates Detection
Michael Bloodgood and Benjamin Strauss

Global constraints and reranking have not been used in cognates detection research to date. We propose methods for using global constraints by performing rescoring of the score matrices produced by state of the art cognates detection systems. Using global constraints to perform rescoring is complementary to state of the art methods for performing cognates detection and results in significant performance improvements beyond current state of the art performance on publicly available datasets with different language pairs and various conditions such as different levels of baseline state of the art performance and different data size conditions, including with more realistic large data size conditions than have been evaluated with in the past.

One-Shot Neural Cross-Lingual Transfer for Paradigm Completion
Katharina Kann, Ryan Cotterell, and Hinrich Schütze

We present a novel cross-lingual transfer method for paradigm completion, the task of mapping a lemma to its inflected forms, using a neural encoder-decoder model, the state of the art for the monolingual task. We use labeled data from a high-resource language to increase performance on a low-resource language. In experiments on 21 language pairs from four different language families, we obtain up to 58% higher accuracy

than without transfer and show that even zero-shot and one-shot learning are possible. We further find that the degree of language relatedness strongly influences the ability to transfer morphological knowledge.

Morphological Inflection Generation with Hard Monotonic Attention
Roee Aharoni and Yoav Goldberg

We present a neural model for morphological inflection generation which employs a hard attention mechanism, inspired by the nearly-monotonic alignment commonly found between the characters in a word and the characters in its inflection. We evaluate the model on three previously studied morphological inflection generation datasets and show that it provides state of the art results in various setups compared to previous neural and non-neural approaches. Finally we present an analysis of the continuous representations learned by both the hard and soft (Bahdanau, 2014) attention models for the task, shedding some light on the features such models extract.

From Characters to Words to in Between: Do We Capture Morphology?
Clara Vania and Adam Lopez

Words can be represented by composing the representations of subword units such as word segments, characters, and/or character n-grams. While such representations are effective and may capture the morphological regularities of words, they have not been systematically compared, and it is not understood how they interact with different morphological typologies. On a language modeling task, we present experiments that systematically vary (1) the basic unit of representation, (2) the composition of these representations, and (3) the morphological typology of the language modeled. Our results extend previous findings that character representations are effective across typologies, and we find that a previously unstudied combination of character trigram representations composed with bi-LSTMs outperforms most others. But we also find room for improvement: none of the character-level models match the predictive accuracy of a model with access to true morphological analyses, even when learned from an order of magnitude more data.

Riemannian Optimization for Skip-Gram Negative Sampling
Alexander Fonarev, Oleksii Grinchuk, Gleb Gusev, Pavel Serdyukov, and Ivan Oseledets

Skip-Gram Negative Sampling (SGNS) word embedding model, well known by its implementation in "word2vec" software, is usually optimized by stochastic gradient descent. However, the optimization of SGNS objective can be viewed as a problem of searching for a good matrix with the low-rank constraint. The most standard way to solve this type of problems is to apply Riemannian optimization framework to optimize the SGNS objective over the manifold of required low-rank matrices. In this paper, we propose an algorithm that optimizes SGNS objective using Riemannian optimization and demonstrates its superiority over popular competitors, such as the original method to train SGNS and SVD over SPPMI matrix.

Deep Multitask Learning for Semantic Dependency Parsing
Hao Peng, Sam Thomson, and Noah A. Smith

We present a deep neural architecture that parses sentences into three semantic dependency graph formalisms. By using efficient, nearly arc-factored inference and a bidirectional-LSTM composed with a multi-layer perceptron, our base system is able to significantly improve the state of the art for semantic dependency parsing, without using hand-engineered features or syntax. We then explore two multitask learning approaches—one that shares parameters across formalisms, and one that uses higher-order structures to predict the graphs jointly. We find that both approaches improve performance across formalisms on average, achieving a new state of the art. Our code is open-source and available at https://github.com/Noahs-ARK/NeurboParser.

Improved Word Representation Learning with Sememes
Yilin Niu, Ruobing Xie, Zhiyuan Liu, and Maosong Sun

Sememes are minimum semantic units of word meanings, and the meaning of each word sense is typically composed by several sememes. Since sememes are not explicit for each word, people manually annotate word sememes and form linguistic common-sense knowledge bases. In this paper, we present that, word sememe information can improve word representation learning (WRL), which maps words into a low-dimensional semantic space and serves as a fundamental step for many NLP tasks. The key idea is to utilize word sememes to capture exact meanings of a word within specific contexts accurately. More specifically, we follow the framework of Skip-gram and present three sememe-encoded models to learn representations of sememes, senses and words, where we apply the attention scheme to detect word senses in various contexts. We conduct experiments on two tasks including word similarity and word analogy, and our models significantly outperform baselines. The results indicate that WRL can benefit from sememes via the attention scheme, and also confirm our models being capable of correctly modeling sememe information.

Learning Character-level Compositionality with Visual Features
Frederick Liu, Han Lu, Chieh Lo, and Graham Neubig

Previous work has modeled the compositionality of words by creating character-level models of meaning, reducing problems of sparsity for rare words. However, in many writing systems compositionality has an effect even on the character-level: the meaning of a character is derived by the sum of its parts. In this paper, we model this effect by creating embeddings for characters based on their visual characteristics, creating an image for the character and running it through a convolutional neural network to produce a visual character embedding. Experiments on a text classification task demonstrate that such model allows for better processing of instances with rare characters in languages such as Chinese, Japanese, and Korean. Additionally, qualitative analyses demonstrate that our proposed model learns to focus on the parts of characters that carry topical content which resulting in embeddings that are coherent in visual space.

A Progressive Learning Approach to Chinese SRL Using Heterogeneous Data
Qiaolin Xia, Lei Sha, Baobao Chang, and Zhifang Sui

Previous studies on Chinese semantic role labeling (SRL) have concentrated on a single semantically annotated corpus. But the training data of single corpus is often limited. Whereas the other existing semantically annotated corpora for Chinese SRL are scattered across different annotation frameworks. But still, Data sparsity remains a bottleneck. This situation calls for larger training datasets, or effective approaches which can take advantage of highly heterogeneous data. In this paper, we focus mainly on the latter, that is, to improve Chinese SRL by using heterogeneous corpora together. We propose a novel progressive learning model which augments the Progressive Neural Network with Gated Recurrent Adapters. The model can accommodate heterogeneous inputs and effectively transfer knowledge between them. We also release a new corpus, Chinese SemBank, for Chinese SRL. Experiments on CPB 1.0 show that our model outperforms state-of-the-art methods.

Revisiting Recurrent Networks for Paraphrastic Sentence Embeddings
John Wieting and Kevin Gimpel

We consider the problem of learning general-purpose, paraphrastic sentence embeddings, revisiting the setting of Wieting et al. (2016b). While they found LSTM recurrent networks to underperform word averaging, we present several developments that together produce the opposite conclusion. These include training on sentence pairs rather than phrase pairs, averaging states to represent sequences, and regularizing aggressively. These improve LSTMs in both transfer learning and supervised settings. We also introduce a new recurrent architecture, the Gated Recurrent Averaging Network, that is inspired by averaging and LSTMs while outperforming them both. We analyze our learned models, finding evidence of preferences for particular parts of speech and dependency relations.

Ontology-Aware Token Embeddings for Prepositional Phrase Attachment
Pradeep Dasigi, Waleed Ammar, Chris Dyer, and Eduard Hovy

Type-level word embeddings use the same set of parameters to represent all instances of a word regardless of its context, ignoring the inherent lexical ambiguity in language. Instead, we embed semantic concepts (or synsets) as defined in WordNet and represent a word token in a particular context by estimating a distribution over relevant semantic concepts. We use the new, context-sensitive embeddings in a model for predicting prepositional phrase (PP) attachments and jointly learn the concept embeddings and model parameters. We show that using context-sensitive embeddings improves the accuracy of the PP attachment model by 5.4% absolute points, which amounts to a 34.4% relative reduction in errors.

Identifying 1950s American Jazz Musicians: Fine-Grained IsA Extraction via Modifier Composition
Ellie Pavlick and Marius Pasca

We present a method for populating fine-grained classes (e.g., "1950s American jazz musicians") with instances (e.g., Charles Mingus). While state-of-the-art methods tend to treat class labels as single lexical units, the proposed method considers each of the individual modifiers in the class label relative to the head. An evaluation on the task of reconstructing Wikipedia category pages demonstrates a >10 point increase in AUC, over a strong baseline relying on widely-used Hearst patterns.

Parsing to 1-Endpoint-Crossing, Pagenumber-2 Graphs
Junjie Cao, Sheng Huang, Weiwei Sun, and Xiaojun Wan

We study the Maximum Subgraph problem in deep dependency parsing. We consider two restrictions to deep dependency graphs: (a) 1-endpoint-crossing and (b) pagenumber-2. Our main contribution is an exact algorithm that ob- tains maximum subgraphs satisfying both restrictions simultaneously in time O(n5). Moreover,

ignoring one linguistically-rare structure descreases the complexity to O(n4). We also extend our quartic-time algorithm into a practical parser with a discriminative disambiguation model and evaluate its performance on four linguistic data sets used in semantic dependency parsing.

Semi-supervised Multitask Learning for Sequence Labeling
Marek Rei
We propose a sequence labeling framework with a secondary training objective, learning to predict surrounding words for every word in the dataset. This language modeling objective incentivises the system to learn general-purpose patterns of semantic and syntactic composition, which are also useful for improving accuracy on different sequence labeling tasks. The architecture was evaluated on a range of datasets, covering the tasks of error detection in learner texts, named entity recognition, chunking and POS-tagging. The novel language modeling objective provided consistent performance improvements on every benchmark, without requiring any additional annotated or unannotated data.

Semantic Parsing of Pre-university Math Problems
Takuya Matsuzaki, Takumi Ito, Hidenao Iwane, Hirokazu Anai, and Noriko H. Arai
We have been developing an end-to-end math problem solving system that accepts natural language input. The current paper focuses on how we analyze the problem sentences to produce logical forms. We chose a hybrid approach combining a shallow syntactic analyzer and a manually-developed lexicalized grammar. A feature of the grammar is that it is extensively typed on the basis of a formal ontology for pre-university math. These types are helpful in semantic disambiguation inside and across sentences. Experimental results show that the hybrid system produces a well-formed logical form with 88% precision and 56% recall.

Poster Session P2 (Short Papers)

Time: *17:40–19:40* Location: Bayshore Grand Ballroom/Foyer

AliMe Chat: A Sequence to Sequence and Rerank based Chatbot Engine
Minghui Qiu, Feng-Lin Li, Siyu Wang, Xing Gao, Yan Chen, Weipeng Zhao, Haiqing Chen, Jun Huang, and Wei Chu
We propose AliMe Chat, an open-domain chatbot engine that integrates the joint results of Information Retrieval (IR) and Sequence to Sequence (Seq2Seq) based generation models. AliMe Chat uses an attentive Seq2Seq based rerank model to optimize the joint results. Extensive experiments show our engine outperforms both IR and generation based models. We launch AliMe Chat for a real-world industrial application and observe better results than another public chatbot.

A Conditional Variational Framework for Dialog Generation
Xiaoyu Shen, Hui Su, Yanran Li, Wenjie Li, Shuzi Niu, Yang Zhao, Akiko Aizawa, and Guoping Long
Deep latent variable models have been shown to facilitate the response generation for open-domain dialog systems. However, these latent variables are highly randomized, leading to uncontrollable generated responses. In this paper, we propose a framework allowing conditional response generation based on specific attributes. These attributes can be either manually assigned or automatically detected. Moreover, the dialog states for both speakers are modeled separately in order to reflect personal features. We validate this framework on two different scenarios, where the attribute refers to genericness and sentiment states respectively. The experiment result testified the potential of our model, where meaningful responses can be generated in accordance with the specified attributes.

Question Answering through Transfer Learning from Large Fine-grained Supervision Data
Sewon Min, Minjoon Seo, and Hannaneh Hajishirzi
We show that the task of question answering (QA) can significantly benefit from the transfer learning of models trained on a different large, fine-grained QA dataset. We achieve the state of the art in two well-studied QA datasets, WikiQA and SemEval-2016 (Task 3A), through a basic transfer learning technique from SQuAD. For WikiQA, our model outperforms the previous best model by more than 8%. We demonstrate that finer supervision provides better guidance for learning lexical and syntactic information than coarser supervision, through quantitative results and visual analysis. We also show that a similar transfer learning procedure achieves the state of the art on an entailment task.

Self-Crowdsourcing Training for Relation Extraction
Azad Abad, Moin Nabi, and Alessandro Moschitti
In this paper we introduce a self-training strategy for crowdsourcing. The training examples are automatically selected to train the crowd workers. Our experimental results show an impact of 5% Improvement in terms of F1 for relation extraction task, compared to the method based on distant supervision.

A Generative Attentional Neural Network Model for Dialogue Act Classification
Quan Hung Tran, Gholamreza Haffari, and Ingrid Zukerman
We propose a novel generative neural network architecture for Dialogue Act classification. Building upon the Recurrent Neural Network framework, our model incorporates a novel attentional technique and a label to label connection for sequence learning, akin to Hidden Markov Models. The experiments show that both of these innovations lead our model to outperform strong baselines for dialogue act classification on MapTask and Switchboard corpora. We further empirically analyse the effectiveness of each of the new innovations.

Salience Rank: Efficient Keyphrase Extraction with Topic Modeling
Nedelina Teneva and Weiwei Cheng
Topical PageRank (TPR) uses latent topic distribution inferred by Latent Dirichlet Allocation (LDA) to perform ranking of noun phrases extracted from documents. The ranking procedure consists of running PageRank K times, where K is the number of topics used in the LDA model. In this paper, we propose a modification of TPR, called Salience Rank. Salience Rank only needs to run PageRank once and extracts comparable or better keyphrases on benchmark datasets. In addition to quality and efficiency benefit, our method has the flexibility to extract keyphrases with varying tradeoffs between topic specificity and corpus specificity.

List-only Entity Linking
Ying Lin, Chin-Yew Lin, and Heng Ji
Traditional Entity Linking (EL) technologies rely on rich structures and properties in the target knowledge base (KB). However, in many applications, the KB may be as simple and sparse as lists of names of the same type (e.g., lists of products). We call it as List-only Entity Linking problem. Fortunately, some mentions may have more cues for linking, which can be used as seed mentions to bridge other mentions and the uninformative entities. In this work, we select most linkable mentions as seed mentions and disambiguate other mentions by comparing them with the seed mentions rather than directly with the entities. Our experiments on linking mentions to seven automatically mined lists show promising results and demonstrate the effectiveness of our approach.

Improving Native Language Identification by Using Spelling Errors
Lingzhen Chen, Carlo Strapparava, and Vivi Nastase
In this paper, we explore spelling errors as a source of information for detecting the native language of a writer, a previously under-explored area. We note that character n-grams from misspelled words are very indicative of the native language of the author. In combination with other lexical features, spelling error features lead to 1.2% improvement in accuracy on classifying texts in the TOEFL11 corpus by the author's native language, compared to systems participating in the NLI shared task.

Disfluency Detection using a Noisy Channel Model and a Deep Neural Language Model
Paria Jamshid Lou and Mark Johnson
This paper presents a model for disfluency detection in spontaneous speech transcripts called LSTM Noisy Channel Model. The model uses a Noisy Channel Model (NCM) to generate n-best candidate disfluency analyses and a Long Short-Term Memory (LSTM) language model to score the underlying fluent sentences of each analysis. The LSTM language model scores, along with other features, are used in a MaxEnt reranker to identify the most plausible analysis. We show that using an LSTM language model in the reranking process of noisy channel disfluency model improves the state-of-the-art in disfluency detection.

On the Equivalence of Holographic and Complex Embeddings for Link Prediction
Katsuhiko Hayashi and Masashi Shimbo
We show the equivalence of two state-of-the-art models for link prediction/knowledge graph completion: Nickel et al's holographic embeddings and Trouillon et al.'s complex embeddings. We first consider a spectral version of the holographic embeddings, exploiting the frequency domain in the Fourier transform for efficient computation. The analysis of the resulting model reveals that it can be viewed as an instance of the complex embeddings with a certain constraint imposed on the initial vectors upon training. Conversely, any set of complex embeddings can be converted to a set of equivalent holographic embeddings.

Sentence Embedding for Neural Machine Translation Domain Adaptation
Rui Wang, Andrew Finch, Masao Utiyama, and Eiichiro Sumita

Although new corpora are becoming increasingly available for machine translation, only those that belong to the same or similar domains are typically able to improve translation performance. Recently Neural Machine Translation (NMT) has become prominent in the field. However, most of the existing domain adaptation methods only focus on phrase-based machine translation. In this paper, we exploit the NMT's internal embedding of the source sentence and use the sentence embedding similarity to select the sentences which are close to in-domain data. The empirical adaptation results on the IWSLT English-French and NIST Chinese-English tasks show that the proposed methods can substantially improve NMT performance by 2.4-9.0 BLEU points, outperforming the existing state-of-the-art baseline by 2.3-4.5 BLEU points.

Data Augmentation for Low-Resource Neural Machine Translation
Marzieh Fadaee, Arianna Bisazza, and Christof Monz

The quality of a Neural Machine Translation system depends substantially on the availability of sizable parallel corpora. For low-resource language pairs this is not the case, resulting in poor translation quality. Inspired by work in computer vision, we propose a novel data augmentation approach that targets low-frequency words by generating new sentence pairs containing rare words in new, synthetically created contexts. Experimental results on simulated low-resource settings show that our method improves translation quality by up to 2.9 BLEU points over the baseline and up to 3.2 BLEU over back-translation.

Speeding Up Neural Machine Translation Decoding by Shrinking Run-time Vocabulary
Xing Shi and Kevin Knight

We speed up Neural Machine Translation (NMT) decoding by shrinking run-time target vocabulary. We experiment with two shrinking approaches: Locality Sensitive Hashing (LSH) and word alignments. Using the latter method, we get a 2x overall speed-up over a highly-optimized GPU implementation, without hurting BLEU. On certain low-resource language pairs, the same methods improve BLEU by 0.5 points. We also report a negative result for LSH on GPUs, due to relatively large overhead, though it was successful on CPUs. Compared with Locality Sensitive Hashing (LSH), decoding with word alignments is GPU-friendly, orthogonal to existing speedup methods and more robust across language pairs.

Chunk-Based Bi-Scale Decoder for Neural Machine Translation
Hao Zhou, Zhaopeng Tu, Shujian Huang, Xiaohua Liu, Hang Li, and Jiajun Chen

In typical neural machine translation~(NMT), the decoder generates a sentence word by word, packing all linguistic granularities in the same time-scale of RNN. In this paper, we propose a new type of decoder for NMT, which splits the decode state into two parts and updates them in two different time-scales. Specifically, we first predict a chunk time-scale state for phrasal modeling, on top of which multiple word time-scale states are generated. In this way, the target sentence is translated hierarchically from chunks to words, with information in different granularities being leveraged. Experiments show that our proposed model significantly improves the translation performance over the state-of-the-art NMT model.

Model Transfer for Tagging Low-resource Languages using a Bilingual Dictionary
Meng Fang and Trevor Cohn

Cross-lingual model transfer is a compelling and popular method for predicting annotations in a low-resource language, whereby parallel corpora provide a bridge to a high-resource language, and its associated annotated corpora. However, parallel data is not readily available for many languages, limiting the applicability of these approaches. We address these drawbacks in our framework which takes advantage of cross-lingual word embeddings trained solely on a high coverage dictionary. We propose a novel neural network model for joint training from both sources of data based on cross-lingual word embeddings, and show substantial empirical improvements over baseline techniques. We also propose several active learning heuristics, which result in improvements over competitive benchmark methods.

EuroSense: Automatic Harvesting of Multilingual Sense Annotations from Parallel Text
Claudio Delli Bovi, Jose Camacho-Collados, Alessandro Raganato, and Roberto Navigli

Parallel corpora are widely used in a variety of Natural Language Processing tasks, from Machine Translation to cross-lingual Word Sense Disambiguation, where parallel sentences can be exploited to automatically generate high-quality sense annotations on a large scale. In this paper we present EuroSense, a multilingual sense-annotated resource based on the joint disambiguation of the Europarl parallel corpus, with almost 123 million sense annotations for over 155 thousand distinct concepts and entities from a language-independent unified sense inventory. We evaluate the quality of our sense annotations intrinsically and extrinsically, showing their effectiveness as training data for Word Sense Disambiguation.

Challenging Language-Dependent Segmentation for Arabic: An Application to Machine Translation and Part-of-Speech Tagging
Hassan Sajjad, Fahim Dalvi, Nadir Durrani, Ahmed Abdelali, Yonatan Belinkov, and Stephan Vogel

Word segmentation plays a pivotal role in improving any Arabic NLP application. Therefore, a lot of research has been spent in improving its accuracy. Off-the-shelf tools, however, are: i) complicated to use and ii) domain/dialect dependent. We explore three language-independent alternatives to morphological segmentation us- ing: i) data-driven sub-word units, ii) characters as a unit of learning, and iii) word embeddings learned using a character CNN (Convolution Neural Network). On the tasks of Machine Translation and POS tagging, we found these methods to achieve close to, and occasionally surpass state-of-the-art performance. In our analysis, we show that a neural machine translation system is sensitive to the ratio of source and target tokens, and a ratio close to 1 or greater, gives optimal performance.

Fast and Accurate Neural Word Segmentation for Chinese
Deng Cai, Hai Zhao, Zhisong Zhang, Yuan Xin, Yongjian Wu, and Feiyue Huang

Neural models with minimal feature engineering have achieved competitive performance against traditional methods for the task of Chinese word segmentation. However, both training and working procedures of the current neural models are computationally inefficient. In this paper, we propose a greedy neural word segmenter with balanced word and character embedding inputs to alleviate the existing drawbacks. Our segmenter is truly end-to-end, capable of performing segmentation much faster and even more accurate than state-of-the-art neural models on Chinese benchmark datasets.

Pay Attention to the Ending:Strong Neural Baselines for the ROC Story Cloze Task
Zheng Cai, Lifu Tu, and Kevin Gimpel

We consider the ROC story cloze task (Mostafazadeh et al., 2016) and present several findings. We develop a model that uses hierarchical recurrent networks with attention to encode the sentences in the story and score candidate endings. By discarding the large training set and only training on the validation set, we achieve an accuracy of 74.7%. Even when we discard the story plots (sentences before the ending) and only train to choose the better of two endings, we can still reach 72.5%. We then analyze this "ending-only" task setting. We estimate human accuracy to be 78% and find several types of clues that lead to this high accuracy, including those related to sentiment, negation, and general ending likelihood regardless of the story context.

Neural Semantic Parsing over Multiple Knowledge-bases
Jonathan Herzig and Jonathan Berant

A fundamental challenge in developing semantic parsers is the paucity of strong supervision in the form of language utterances annotated with logical form. In this paper, we propose to exploit structural regularities in language in different domains, and train semantic parsers over multiple knowledge-bases (KBs), while sharing information across datasets. We find that we can substantially improve parsing accuracy by training a single sequence-to-sequence model over multiple KBs, when providing an encoding of the domain at decoding time. Our model achieves state-of-the-art performance on the Overnight dataset (containing eight domains), improves performance over a single KB baseline from 75.6% to 79.6%, while obtaining a 7x reduction in the number of model parameters.

Representing Sentences as Low-Rank Subspaces
Jiaqi Mu, Suma Bhat, and Pramod Viswanath

Sentences are important semantic units of natural language. A generic, distributional representation of sentences that can capture the latent semantics is beneficial to multiple downstream applications. We observe a simple geometry of sentences – the word representations of a given sentence (on average 10.23 words in all SemEval datasets with a standard deviation 4.84) roughly lie in a low-rank subspace (roughly, rank 4). Motivated by this observation, we represent a sentence by the low-rank subspace spanned by its word vectors. Such an unsupervised representation is empirically validated via semantic textual similarity tasks on 19 different datasets, where it outperforms the sophisticated neural network models, including skip-thought vectors, by 15% on average.

Improving Semantic Relevance for Sequence-to-Sequence Learning of Chinese Social Media Text Summarization
Shuming Ma, Xu Sun, Jingjing Xu, Houfeng Wang, Wenjie Li, and Qi Su

Current Chinese social media text summarization models are based on an encoder-decoder framework. Although its generated summaries are similar to source texts literally, they have low semantic relevance. In this work, our goal is to improve semantic relevance between source texts and summaries for Chinese social media

summarization. We introduce a Semantic Relevance Based neural model to encourage high semantic similarity between texts and summaries. In our model, the source text is represented by a gated attention encoder, while the summary representation is produced by a decoder. Besides, the similarity score between the representations is maximized during training. Our experiments show that the proposed model outperforms baseline systems on a social media corpus.

Determining Whether and When People Participate in the Events They Tweet About
Krishna Chaitanya Sanagavarapu, Alakananda Vempala, and Eduardo Blanco

This paper describes an approach to determine whether people participate in the events they tweet about. Specifically, we determine whether people are participants in events with respect to the tweet timestamp. We target all events expressed by verbs in tweets, including past, present and events that may occur in the future. We present new annotations using 1,096 event mentions, and experimental results showing that the task is challenging.

Separating Facts from Fiction: Linguistic Models to Classify Suspicious and Trusted News Posts on Twitter
Svitlana Volkova, Kyle Shaffer, Jin Yea Jang, and Nathan Hodas

Pew research polls report 62 percent of U.S. adults get news on social media (Gottfried and Shearer, 2016). In a December poll, 64 percent of U.S. adults said that "made-up news" has caused a "great deal of confusion" about the facts of current events (Barthel et al., 2016). Fabricated stories in social media, ranging from deliberate propaganda to hoaxes and satire, contributes to this confusion in addition to having serious effects on global stability. In this work we build predictive models to classify 130 thousand news posts as suspicious or verified, and predict four sub-types of suspicious news — satire, hoaxes, clickbait and propaganda. We show that neural network models trained on tweet content and social network interactions outperform lexical models. Unlike previous work on deception detection, we find that adding syntax and grammar features to our models does not improve performance. Incorporating linguistic features improves classification results, however, social interaction features are most in- formative for finer-grained separation be- tween four types of suspicious news posts.

Recognizing Counterfactual Thinking in Social Media Texts
Youngseo Son, Anneke Buffone, Joe Raso, Allegra Larche, Anthony Janocko, Kevin Zembroski, H. Andrew Schwartz, and Lyle Ungar

Counterfactual statements, describing events that did not occur and their consequents, have been studied in areas including problem-solving, affect management, and behavior regulation. People with more counterfactual thinking tend to perceive life events as more personally meaningful. Nevertheless, counterfactuals have not been studied in computational linguistics. We create a counterfactual tweet dataset and explore approaches for detecting counterfactuals using rule-based and supervised statistical approaches. A combined rule-based and statistical approach yielded the best results (F1 = 0.77) outperforming either approach used alone.

Temporal Orientation of Tweets for Predicting Income of Users
Mohammed Hasanuzzaman, Sabyasachi Kamila, Mandeep Kaur, Sriparna Saha, and Asif Ekbal

Automatically estimating a user's socio-economic profile from their language use in social media can significantly help social science research and various downstream applications ranging from business to politics. The current paper presents the first study where user cognitive structure is used to build a predictive model of income. In particular, we first develop a classifier using a weakly supervised learning framework to automatically time-tag tweets as past, present, or future. We quantify a user's overall temporal orientation based on their distribution of tweets, and use it to build a predictive model of income. Our analysis uncovers a correlation between future temporal orientation and income. Finally, we measure the predictive power of future temporal orientation on income by performing regression.

Character-Aware Neural Morphological Disambiguation
Alymzhan Toleu, Gulmira Tolegen, and Aibek Makazhanov

We develop a language-independent, deep learning-based approach to the task of morphological disambiguation. Guided by the intuition that the correct analysis should be "most similar" to the context, we propose dense representations for morphological analyses and surface context and a simple yet effective way of combining the two to perform disambiguation. Our approach improves on the language-dependent state of the art for two agglutinative languages (Turkish and Kazakh) and can be potentially applied to other morphologically complex languages.

Character Composition Model with Convolutional Neural Networks for Dependency Parsing on Morphologically Rich Languages
Xiang Yu and Ngoc Thang Vu

We present a transition-based dependency parser that uses a convolutional neural network to compose word representations from characters. The character composition model shows great improvement over the word-lookup model, especially for parsing agglutinative languages. These improvements are even better than using pre-trained word embeddings from extra data. On the SPMRL data sets, our system outperforms the previous best greedy parser (Ballesteros et. al, 2015) by a margin of 3% on average.

How (not) to train a dependency parser: The curious case of jackknifing part-of-speech taggers
Željko Agić and Natalie Schluter

In dependency parsing, jackknifing taggers is indiscriminately used as a simple adaptation strategy. Here, we empirically evaluate when and how (not) to use jackknifing in parsing. On 26 languages, we reveal a preference that conflicts with, and surpasses the ubiquitous ten-folding. We show no clear benefits of tagging the training data in cross-lingual parsing.

Poster Session P2 (Systems Demonstrations)

Time: *17:40–19:40* Location: Stanley Park Foyer/Cypress

Annotating tense, mood and voice for English, French and German
Anita Ramm, Sharid Loáiciga, Annemarie Friedrich, and Alexander Fraser

We present the first open-source tool for annotating morphosyntactic tense, mood and voice for English, French and German verbal complexes. The annotation is based on a set of language-specific rules, which are applied on dependency trees and leverage information about lemmas, morphological properties and POS-tags of the verbs. Our tool has an average accuracy of about 76%. The tense, mood and voice features are useful both as features in computational modeling and for corpus-linguistic research.

Automating Biomedical Evidence Synthesis: RobotReviewer
Iain Marshall, Joël Kuiper, Edward Banner, and Byron C. Wallace

We present RobotReviewer, an open-source web-based system that uses machine learning and NLP to semi-automate biomedical evidence synthesis, to aid the practice of Evidence-Based Medicine. RobotReviewer processes full-text journal articles (PDFs) describing randomized controlled trials (RCTs). It appraises the reliability of RCTs and extracts text describing key trial characteristics (e.g., descriptions of the population) using novel NLP methods. RobotReviewer then automatically generates a report synthesising this information. Our goal is for RobotReviewer to automatically extract and synthesise the full-range of structured data needed to inform evidence-based practice.

Benben: A Chinese Intelligent Conversational Robot
Wei-Nan Zhang, Ting Liu, Bing Qin, Yu Zhang, Wanxiang Che, Yanyan Zhao, and Xiao Ding

Recently, conversational robot is widely used in mobile terminals as the virtual assistant or companion. The goals of prevalent conversational robots mainly focus on four categories, namely chit-chat, task completion, question answering and recommendation. In this paper, we present a Chinese intelligent conversational robot, Benben, which is designed to achieve these goals in a unified architecture. Moreover, it also has some featured functions such as diet map, implicit feedback based conversation, interactive machine reading, news recommendation, etc. Since the release of Benben at June 6, 2016, there are 2,505 users (till Feb 22, 2017) and 11,107 complete human-robot conversations, which totally contain 198,998 single turn conversation pairs.

End-to-End Non-Factoid Question Answering with an Interactive Visualization of Neural Attention Weights
Andreas Rücklé and Iryna Gurevych

Advanced attention mechanisms are an important part of successful neural network approaches for non-factoid answer selection because they allow the models to focus on few important segments within rather long answer texts. Analyzing attention mechanisms is thus crucial for understanding strengths and weaknesses of particular models. We present an extensible, highly modular service architecture that enables to transform neural network models for non-factoid answer selection into fully featured end-to-end question answering systems. The primary objective of our system is to enable researchers a way to interactively explore and compare

attention-based neural networks for answer selection. Our interactive user interface helps researchers to better understand the capabilities of the different approaches and can aid qualitative analyses. The source-code of our system is publicly available.

ESTEEM: A Novel Framework for Qualitatively Evaluating and Visualizing Spatiotemporal Embeddings in Social Media
Dustin Arendt and Svitlana Volkova

Analyzing and visualizing large amounts of social media communications and contrasting short-term conversation changes over time and geolocations is extremely important for commercial and government applications. Earlier approaches for large-scale text stream summarization used dynamic topic models and trending words. Instead, we rely on text embeddings – low-dimensional word representations in a continuous vector space where similar words are embedded nearby each other. This paper presents ESTEEM, a novel tool for visualizing and evaluating spatiotemporal embeddings learned from streaming social media texts. Our tool allows users to monitor and analyze query words and their closest neighbors with an interactive interface. We used state-of-the-art techniques to learn embeddings and developed a visualization to represent dynamically changing relations between words in social media over time and other dimensions. This is the first interactive visualization of streaming text representations learned from social media texts that also allows users to contrast differences across multiple dimensions of the data.

Exploring Diachronic Lexical Semantics with JeSemE
Johannes Hellrich and Udo Hahn

Recent advances in distributional semantics combined with the availability of large-scale diachronic corpora offer new research avenues for the Digital Humanities. JeSemE, the Jena Semantic Explorer, renders assistance to a non-technical audience to investigate diachronic semantic topics. JeSemE runs as a website with query options and interactive visualizations of results, as well as a REST API for access to the underlying diachronic data sets.

Extended Named Entity Recognition API and Its Applications in Language Education
Tuan Duc Nguyen, Khai Mai, Thai-Hoang Pham, Minh Trung Nguyen, Truc-Vien T. Nguyen, Takashi Eguchi, Ryohei Sasano, and Satoshi Sekine

We present an Extended Named Entity Recognition API to recognize various types of entities and classify the entities into 200 different categories. Each entity is classified into a hierarchy of entity categories, in which the categories near the root are more general than the categories near the leaves of the hierarchy. This category information can be used in various applications such as language educational applications, online news services and recommendation engines. We show an application of the proposed API in a Japanese online news service for Japanese language studying users.

Hafez: an Interactive Poetry Generation System
Marjan Ghazvininejad, Xing Shi, Jay Priyadarshi, and Kevin Knight

Hafez is an automatic poetry generation system that integrates a Recurrent Neural Network (RNN) with a Finite State Acceptor (FSA). It generates sonnets given arbitrary topics. Furthermore, Hafez enables users to revise and polish generated poems by adjusting various style configurations. Experiments demonstrate that such "polish" mechanisms consider the user's intention and lead to a better poem. For evaluation, we build a web interface where users can rate the quality of each poem from 1 to 5 stars. We also speed up the whole system by a factor of 10, via vocabulary pruning and GPU computation, so that adequate feedback can be collected at a fast pace. Based on such feedback, the system learns to adjust its parameters to improve poetry quality. .

Interactive Visual Analysis of Transcribed Multi-Party Discourse
Mennatallah El-Assady, Annette Hautli-Janisz, Valentin Gold, Miriam Butt, Katharina Holzinger, and Daniel Keim

We present a first web-based Visual Analytics framework for the analysis of multiparty discourse data using verbatim text transcripts. Our framework supports a broad range of server-based processing steps, ranging from data mining and statistical analysis to deep linguistic parsing of English and German. On the client-side, browser-based Visual Analytics components enable multiple perspectives on the analyzed data. These interactive visualizations allow exploratory content analysis, argumentation pattern review, and speaker interaction modeling.

Life-iNet: A Structured Network-Based Knowledge Exploration and Analytics System for Life Sciences
Xiang Ren, Jiaming Shen, Meng Qu, Xuan Wang, Zeqiu Wu, Qi Zhu, Meng Jiang, Fangbo Tao, Saurabh Sinha, David Liem, Peipei Ping, Richard Weinshilboum, and Jiawei Han

Search engines running on scientific literature have been widely used by life scientists to find publications related to their research. However, existing search engines in life-science domain, such as PubMed, have limitations when applied to exploring and analyzing factual knowledge (e.g., disease-gene associations) in massive text corpora. These limitations are mainly due to the problems that factual information exists as an unstructured form in text, and also keyword and MeSH term-based queries cannot effectively imply semantic relations between entities. This demo paper presents the LifeNet system to address the limitations in existing search engines on facilitating life sciences research. LifeNet automatically constructs structured networks of factual knowledge from large amounts of background documents, to support efficient exploration of structured factual knowledge in the unstructured literature. It also provides functionalities for finding distinctive entities for given entity types, and generating hypothetical facts to assist literature-based knowledge discovery (e.g., drug target prediction).

Olelo: A Question Answering Application for Biomedicine
Mariana Neves, Hendrik Folkerts, Marcel Jankrift, Julian Niedermeier, Toni Stachewicz, Sören Tietböhl, Milena Kraus, and Matthias Uflacker

Despite the importance of the biomedical domain, there are few reliable applications to support researchers and physicians for retrieving particular facts that fit their needs. Users typically rely on search engines that only support keyword- and filter-based searches. We present Olelo, a question answering system for biomedicine. Olelo is built on top of an in-memory database, integrates domain resources, such as document collections and terminologies, and uses various natural language processing components. Olelo is fast, intuitive and easy to use. We evaluated the systems on two use cases: answering questions related to a particular gene and on the BioASQ benchmark. Olelo is available at: http://hpi.de/plattner/olelo.

OpenNMT: Open-Source Toolkit for Neural Machine Translation
Guillaume Klein, Yoon Kim, Yuntian Deng, Jean Senellart, and Alexander Rush

We describe an open-source toolkit for neural machine translation (NMT). The toolkit prioritizes efficiency, modularity, and extensibility with the goal of supporting NMT research into model architectures, feature representations, and source modalities, while maintaining competitive performance and reasonable training requirements. The toolkit consists of modeling and translation support, as well as detailed pedagogical documentation about the underlying techniques.

PyDial: A Multi-domain Statistical Dialogue System Toolkit
Stefan Ultes, Lina M. Rojas Barahona, Pei-Hao Su, David Vandyke, Dongho Kim, Iñigo Casanueva, Paweł Budzianowski, Nikola Mrkšić, Tsung-Hsien Wen, Milica Gasic, and Steve Young

Statistical Spoken Dialogue Systems have been around for many years. However, access to these systems has always been difficult as there is still no publicly available end-to-end system implementation. To alleviate this, we present CU-PyDial, an open-source end-to-end statistical spoken dialogue system toolkit which provides implementations of statistical approaches for all dialogue system modules. Moreover, it has been extended to provide multi-domain conversational functionality. It offers easy configuration, easy extensibility, and domain-independent implementations of the respective dialogue system modules. The toolkit is available for download under the Apache 2.0 license.

RelTextRank: An Open Source Framework for Building Relational Syntactic-Semantic Text Pair Representations
Kateryna Tymoshenko, Alessandro Moschitti, Massimo Nicosia, and Aliaksei Severyn

We present a highly-flexible UIMA-based pipeline for developing structural kernel-based systems for relational learning from text, i.e., for generating training and test data for ranking, classifying short text pairs or measuring similarity between pieces of text. For example, the proposed pipeline can represent an input question and answer sentence pairs as syntactic-semantic structures, enriching them with relational information, e.g., links between question class, focus and named entities, and serializes them as training and test files for the tree kernel-based reranking framework. The pipeline generates a number of dependency and shallow chunk-based representations shown to achieve competitive results in previous work. It also enables easy evaluation of the models thanks to cross-validation facilities.

Scattertext: a Browser-Based Tool for Visualizing how Corpora Differ
Jason Kessler

Scattertext is an open source tool for visualizing linguistic variation between document categories in a language-independent way. The tool presents a scatterplot, where each axis corresponds to the rank-frequency a term occurs in a category of documents. Through a tie-breaking strategy, the tool is able to display thousands of visible term-representing points and find space to legibly label hundreds of them. Scattertext also lends itself to a query-based visualization of how the use of terms with similar embeddings differs between document categories, as well as a visualization for comparing the importance scores of bag-of-words features to univariate metrics.

Semedico: A Comprehensive Semantic Search Engine for the Life Sciences
Erik Faessler and Udo Hahn

Semedico is a semantic search engine designed to support literature search in the life sciences by integrating the semantics of the domain at all stages of the search process - from query formulation via query processing up to the presentation of results. Semedico excels with an ad-hoc search approach which directly reflects relevance in terms of information density of entities and relations among them (events) and, a truly unique feature, ranks interaction events by certainty information reflecting the degree of factuality of the encountered event.

SuperAgent: A Customer Service Chatbot for E-commerce Websites
Lei Cui, Shaohan Huang, Furu Wei, Chuanqi Tan, Chaoqun Duan, and Ming Zhou

Conventional customer service chatbots are usually built upon human dialogs, yet confronted with the problems of data scale and privacy. In this paper, we present SuperAgent, which is a customer service chatbot leveraging large-scale and public available e-commerce data. Distinct from existing counterparts, SuperAgent takes advantage of data from in-page product descriptions as well as user-generated content from e-commerce websites, which is more practical and cost-effective to answer repetitive questions, thereby freeing up support staffs to answer much higher value questions. We demonstrate SuperAgent as an add-on extension to mainstream web browsers and show its usefulness to user's online shopping experience.

Swanson linking revisited: Accelerating literature-based discovery across domains using a conceptual influence graph
Gus Hahn-Powell, Marco A. Valenzuela-Escárcega, and Mihai Surdeanu

We introduce a modular approach for literature-based discovery consisting of a machine reading and knowledge assembly component that together produce a graph of influence relations (e.g., "A promotes B") from a collection of publications. A search engine is used to explore direct and indirect influence chains. Query results are substantiated with textual evidence, ranked according to their relevance, and presented in both a table-based view, as well as a network graph visualization. Our approach operates in both domain-specific settings, where there are knowledge bases and ontologies available to guide reading, and in multi-domain settings where such resources are absent. We demonstrate that this deep reading and search system reduces the effort needed to uncover "undiscovered public knowledge", and that with the aid of this tool a domain expert was able to drastically reduce her model building time from months to two days.

UCCAApp: Web-application for Syntactic and Semantic Phrase-based Annotation
Omri Abend, Shai Yerushalmi, and Ari Rappoport

We present UCCAApp, an open-source, flexible web-application for syntactic and semantic phrase-based annotation in general, and for UCCA annotation in particular. UCCAApp supports a variety of formal properties that have proven useful for syntactic and semantic representation, such as discontiguous phrases, multiple parents and empty elements, making it useful to a variety of other annotation schemes with similar formal properties. UCCAApp's user interface is intuitive and user friendly, so as to support annotation by users with no background in linguistics or formal representation. Indeed, a pilot version of the application has been successfully used in the compilation of the UCCA Wikipedia treebank by annotators with no previous linguistic training. The application and all accompanying resources are released as open source under the GNU public license, and are available online along with a live demo.

WebChild 2.0 : Fine-Grained Commonsense Knowledge Distillation
Niket Tandon, Gerard de Melo, and Gerhard Weikum

Despite important progress in the area of intelligent systems, most such systems still lack commonsense knowledge that appears crucial for enabling smarter, more human-like decisions. In this paper, we present a system based on a series of algorithms to distill fine-grained disambiguated commonsense knowledge from massive amounts of text. Our WebChild 2.0 knowledge base is one of the largest commonsense knowledge bases available, describing over 2 million disambiguated concepts and activities, connected by over 18 million assertions.

Zara Returns: Improved Personality Induction and Adaptation by an Empathetic Virtual Agent

Farhad Bin Siddique, Onno Kampman, Yang Yang, Anik Dey, and Pascale Fung

Virtual agents need to adapt their personality to the user in order to become more empathetic. To this end, we developed Zara the Supergirl, an interactive empathetic agent, using a modular approach. In this paper, we describe the enhanced personality module with improved recognition from speech and text using deep learning frameworks. From raw audio, an average F-score of 69.6 was obtained from real-time personality assessment using a Convolutional Neural Network (CNN) model. From text, we improved personality recognition results with a CNN model on top of pre-trained word embeddings and obtained an average F-score of 71.0. Results from our Human-Agent Interaction study confirmed our assumption that people have different agent personality preferences. We use insights from this study to adapt our agent to user personality.

Social Event

Tuesday, August 1, 2017, 18:00 – 21:00

Vancouver Aquarium (Stanley Park)
Vancouver Aquarium
845 Avison Way
Vancouver, BC
V6G 3E2
http://www.vanaqua.org/

The ACL 2017 Social and Networking Event will be held at the Vancouver Aquarium located in Stanley Park, on Tuesday, August 1, starting at 18:00. Here you will enjoy desserts, coffee and tea, and a cash bar. Bring your kids and enjoy a marine sanctuary in the heart of Stanley Park, home to thousands of incredible ocean species and amazing aquatic life. Since opening in 1956, the Vancouver Aquarium has connected more than 40 million people from around the world to our oceans and all the wonders within them. Enjoy networking with colleagues and have a relaxing evening!

5

Main Conference: Wednesday, August 2

Overview

7:30 – 9:00	**Breakfast**	*Bayshore Grand Foyer*
9:00 – 10:10	**Plenary Session** Invited Talk: Mirella Lapata – "Translating from Multiple Modalities to Text and Back"	*Bayshore Grand Ballroom*
10:10 – 10:40	**Coffee break**	*Bayshore Grand Foyer*

Session 7

10:40 – 12:30	Outstanding Papers 1 *Salons A/B/C*	Outstanding Papers 2 *Salons D/E/F*

12:00 – 13:30	**Lunch break**	
13:00 – 2:30	**ACL Business Meeting** *(Note: open to all; overlap with lunch break)*	*Salons D/E/F*
14:30 – 15:00	**Coffee break**	*Bayshore Grand Foyer*

Session 8

15:00 – 16:45	Outstanding Papers 3 *Salons A/B/C*	Outstanding Papers 4 *Salons D/E/F*

16:45 – 17:45	**Lifetime Achievement Award**	*Bayshore Grand Ballroom*
17:45 – 18:00	**Closing Session**	*Bayshore Grand Ballroom*

Keynote Address: Mirella Lapata

Translating from Multiple Modalities to Text and Back

Wednesday, August 2, 2017, 9:00–10:10

Bayshore Grand Ballroom

Abstract: Recent years have witnessed the development of a wide range of computational tools that process and generate natural language text. Many of these have become familiar to mainstream computer users in the from of web search, question answering, sentiment analysis, and notably machine translation. The accessibility of the web could be further enhanced with applications that not only translate between different languages (e.g., from English to French) but also within the same language, between different modalities, or different data formats. The web is rife with non-linguistic data (e.g., video, images, source code) that cannot be indexed or searched since most retrieval tools operate over textual data.

In this talk I will argue that in order to render electronic data more accessible to individuals and computers alike, new types of translation models need to be developed. I will focus on three examples, text simplification, source code generation, and movie summarization. I will illustrate how recent advances in deep learning can be extended in order to induce general representations for different modalities and learn how to translate between these and natural language.

Biography: Mirella Lapata is professor of natural language processing in the School of Informatics at the University of Edinburgh. Her research focuses on getting computers to understand, reason with, and generate. She is as an associate editor of the Journal of Artificial Intelligence Research and has served on the editorial boards of Transactions of the ACL and Computational Linguistics. She was the first recipient of the Karen Sparck Jones award of the British Computer Society, recognizing key contributions to NLP and information retrieval. She received two EMNLP best paper awards and currently holds a prestigious Consolidator Grant from the European Research Council.

Session 7 Overview – Wednesday, August 2, 2017

Track A	Track B
Outstanding Papers 1 Salons B/C	*Outstanding Papers 2* Salons E/F
Towards an Automatic Turing Test: Learning to Evaluate Dialogue Responses *R. Lowe, M. Noseworthy, I. V. Serban, N. Angelard-Gontier, Y. Bengio, and J. Pineau*	Visualizing and Understanding Neural Machine Translation *Y. Ding, Y. Liu, H. Luan, and M. Sun*
A Transition-Based Directed Acyclic Graph Parser for UCCA *D. Hershcovich, O. Abend, and A. Rappoport*	Detecting annotation noise in automatically labelled data *I. Rehbein and J. Ruppenhofer*
Abstract Syntax Networks for Code Generation and Semantic Parsing *M. Rabinovich, M. Stern, and D. Klein*	Attention Strategies for Multi-Source Sequence-to-Sequence Learning *J. Libovický and J. Helcl*
The Role of Prosody and Speech Register in Word Segmentation: A Computational Modelling Perspective *B. Ludusan, R. Mazuka, M. Bernard, A. Cristia, and E. Dupoux*	Understanding and Detecting Diverse Supporting Arguments on Controversial Issues *X. Hua and L. Wang*
A Two-Stage Parsing Method for Text-Level Discourse Analysis *Y. Wang, S. Li, and H. Wang*	A Neural Model for User Geolocation and Lexical Dialectology *A. Rahimi, T. Cohn, and T. Baldwin*
Error-repair Dependency Parsing for Ungrammatical Texts *K. Sakaguchi, M. Post, and B. Van Durme*	A Corpus of Natural Language for Visual Reasoning *A. Suhr, M. Lewis, J. Yeh, and Y. Artzi*

10:40 10:59 11:18 11:37 11:50 12:13

Parallel Session 7

Session Session 7A: Outstanding Papers 1

Towards an Automatic Turing Test: Learning to Evaluate Dialogue Responses
Ryan Lowe, Michael Noseworthy, Iulian Vlad Serban, Nicolas Angelard-Gontier, Yoshua Bengio, and Joelle Pineau 10:40–10:58

Automatically evaluating the quality of dialogue responses for unstructured domains is a challenging problem. Unfortunately, existing automatic evaluation metrics are biased and correlate very poorly with human judgements of response quality (Liu et al., 2016). Yet having an accurate automatic evaluation procedure is crucial for dialogue research, as it allows rapid prototyping and testing of new models with fewer expensive human evaluations. In response to this challenge, we formulate automatic dialogue evaluation as a learning problem.We present an evaluation model (ADEM)that learns to predict human-like scores to input responses, using a new dataset of human response scores. We show that the ADEM model's predictions correlate significantly, and at a level much higher than word-overlap metrics such as BLEU, with human judgements at both the utterance and system-level. We also show that ADEM can generalize to evaluating dialogue mod-els unseen during training, an important step for automatic dialogue evaluation.

A Transition-Based Directed Acyclic Graph Parser for UCCA
Daniel Hershcovich, Omri Abend, and Ari Rappoport 10:59–11:17

We present the first parser for UCCA, a cross-linguistically applicable framework for semantic representation, which builds on extensive typological work and supports rapid annotation. UCCA poses a challenge for existing parsing techniques, as it exhibits reentrancy (resulting in DAG structures), discontinuous structures and non-terminal nodes corresponding to complex semantic units. To our knowledge, the conjunction of these formal properties is not supported by any existing parser. Our transition-based parser, which uses a novel transition set and features based on bidirectional LSTMs, has value not just for UCCA parsing: its ability to handle more general graph structures can inform the development of parsers for other semantic DAG structures, and in languages that frequently use discontinuous structures.

Abstract Syntax Networks for Code Generation and Semantic Parsing
Maxim Rabinovich, Mitchell Stern, and Dan Klein 11:18–11:36

Tasks like code generation and semantic parsing require mapping unstructured (or partially structured) inputs to well-formed, executable outputs. We introduce abstract syntax networks, a modeling framework for these problems. The outputs are represented as abstract syntax trees (ASTs) and constructed by a decoder with a dynamically-determined modular structure paralleling the structure of the output tree. On the benchmark Hearthstone dataset for code generation, our model obtains 79.2 BLEU and 22.7% exact match accuracy, compared to previous state-of-the-art values of 67.1 and 6.1%. Furthermore, we perform competitively on the Atis, Jobs, and Geo semantic parsing datasets with no task-specific engineering.

The Role of Prosody and Speech Register in Word Segmentation: A Computational Modelling Perspective
Bogdan Ludusan, Reiko Mazuka, Mathieu Bernard, Alejandrina Cristia, and Emmanuel Dupoux 11:37–11:49

This study explores the role of speech register and prosody for the task of word segmentation. Since these two factors are thought to play an important role in early language acquisition, we aim to quantify their contribution for this task. We study a Japanese corpus containing both infant- and adult-directed speech and we apply four different word segmentation models, with and without knowledge of prosodic boundaries. The results showed that the difference between registers is smaller than previously reported and that prosodic boundary information helps more adult- than infant-directed speech.

A Two-Stage Parsing Method for Text-Level Discourse Analysis
Yizhong Wang, Sujian Li, and Houfeng Wang 11:50–12:12

Previous work introduced transition-based algorithms to form a unified architecture of parsing rhetorical structures (including span, nuclearity and relation), but did not achieve satisfactory performance. In this paper, we propose that transition-based model is more appropriate for parsing the naked discourse tree (i.e., identifying span and nuclearity) due to data sparsity. At the same time, we argue that relation labeling can benefit from naked tree structure and should be treated elaborately with consideration of three kinds of relations including within-sentence, across-sentence and across-paragraph relations. Thus, we design a pipelined two-stage parsing method for generating an RST tree from text. Experimental results show that our method achieves

state-of-the-art performance, especially on span and nuclearity identification.

Error-repair Dependency Parsing for Ungrammatical Texts
Keisuke Sakaguchi, Matt Post, and Benjamin Van Durme 12:13–12:25

We propose a new dependency parsing scheme which jointly parses a sentence and repairs grammatical errors by extending the non-directional transition-based formalism of Goldberg and Elhadad (2010) with three additional actions: SUBSTITUTE, DELETE, INSERT. Because these actions may cause an infinite loop in derivation, we also introduce simple constraints that ensure the parser termination. We evaluate our model with respect to dependency accuracy and grammaticality improvements for ungrammatical sentences, demonstrating the robustness and applicability of our scheme.

Session Session 7B: Outstanding Papers 2

Visualizing and Understanding Neural Machine Translation
Yanzhuo Ding, Yang Liu, Huanbo Luan, and Maosong Sun 10:40–10:58

While neural machine translation (NMT) has made remarkable progress in recent years, it is hard to interpret its internal workings due to the continuous representations and non-linearity of neural networks. In this work, we propose to use layer-wise relevance propagation (LRP) to compute the contribution of each contextual word to arbitrary hidden states in the attention-based encoder-decoder framework. We show that visualization with LRP helps to interpret the internal workings of NMT and analyze translation errors.

Detecting annotation noise in automatically labelled data
Ines Rehbein and Josef Ruppenhofer 10:59–11:17

We introduce a method for error detection in automatically annotated text, aimed at supporting the creation of high-quality language resources at affordable cost. Our method combines an unsupervised generative model with human supervision from active learning. We test our approach on in-domain and out-of-domain data in two languages, in AL simulations and in a real world setting. For all settings, the results show that our method is able to detect annotation errors with high precision and high recall.

Attention Strategies for Multi-Source Sequence-to-Sequence Learning
Jindřich Libovický and Jindřich Helcl 11:18–11:36

Modeling attention in neural multi-source sequence-to-sequence learning remains a relatively unexplored area, despite its usefulness in tasks that incorporate multiple source languages or modalities. We propose two novel approaches to combine the outputs of attention mechanisms over each source sequence, flat and hierarchical. We compare the proposed methods with existing techniques and present results of systematic evaluation of those methods on the WMT16 Multimodal Translation and Automatic Post-editing tasks. We show that the proposed methods achieve competitive results on both tasks.

Understanding and Detecting Diverse Supporting Arguments on Controversial Issues
Xinyu Hua and Lu Wang 11:37–11:49

We investigate the problem of sentence-level supporting argument detection from relevant documents for user-specified claims. A dataset containing claims and associated citation articles is collected from online debate website idebate.org. We then manually label sentence-level supporting arguments from the documents along with their types as study, factual, opinion, or reasoning. We further characterize arguments of different types, and explore whether leveraging type information can facilitate the supporting arguments detection task. Experimental results show that LambdaMART (Burges, 2010) ranker that uses features informed by argument types yields better performance than the same ranker trained without type information.

A Neural Model for User Geolocation and Lexical Dialectology
Afshin Rahimi, Trevor Cohn, and Timothy Baldwin 11:50–12:12

We propose a simple yet effective text-based user geolocation model based on a neural network with one hidden layer, which achieves state of the art performance over three Twitter benchmark geolocation datasets, in addition to producing word and phrase embeddings in the hidden layer that we show to be useful for detecting dialectal terms. As part of our analysis of dialectal terms, we release DAREDS, a dataset for evaluating dialect term detection methods.

A Corpus of Natural Language for Visual Reasoning
Alane Suhr, Mike Lewis, James Yeh, and Yoav Artzi 12:13–12:25

We present a new visual reasoning language dataset, containing 92,244 pairs of examples of natural statements grounded in synthetic images with 3,962 unique sentences. We describe a method of crowdsourcing linguistically-diverse data, and present an analysis of our data. The data demonstrates a broad set of linguistic phenomena, requiring visual and set-theoretic reasoning. We experiment with various models, and show the data presents a strong challenge for future research.

Session 8 Overview – Wednesday, August 2, 2017

Track A	Track B
Outstanding Papers 3 Salons B/C	*Outstanding Papers 4* Salons E/F
Abstractive Document Summarization with a Graph-Based Attentional Neural Model *J. Tan, X. Wan, and J. Xiao*	Joint Extraction of Entities and Relations Based on a Novel Tagging Scheme *S. Zheng, F. Wang, H. Bao, Y. Hao, P. Zhou, and B. Xu*
Probabilistic Typology: Deep Generative Models of Vowel Inventories *R. Cotterell and J. Eisner*	A Local Detection Approach for Named Entity Recognition and Mention Detection *M. Xu, H. Jiang, and S. Watcharawittayakul*
Adversarial Multi-Criteria Learning for Chinese Word Segmentation *X. Chen, Z. Shi, X. Qiu, and X. Huang*	Vancouver Welcomes You! Minimalist Location Metonymy Resolution *M. Gritta, M. T. Pilehvar, N. Limsopatham, and N. Collier*
Neural Joint Model for Transition-based Chinese Syntactic Analysis *S. Kurita, D. Kawahara, and S. Kurohashi*	Unifying Text, Metadata, and User Network Representations with a Neural Network for Geolocation Prediction *Y. Miura, M. Taniguchi, T. Taniguchi, and T. Ohkuma*
Robust Incremental Neural Semantic Graph Parsing *J. Buys and P. Blunsom*	Multi-Task Video Captioning with Video and Entailment Generation *R. Pasunuru and M. Bansal*

15:00 15:19 15:38 15:57 16:16

Parallel Session 8

Session Session 8A: Outstanding Papers 3

Abstractive Document Summarization with a Graph-Based Attentional Neural Model
Jiwei Tan, Xiaojun Wan, and Jianguo Xiao 15:00–15:18

Abstractive summarization is the ultimate goal of document summarization research, but previously it is less investigated due to the immaturity of text generation techniques. Recently impressive progress has been made to abstractive sentence summarization using neural models. Unfortunately, attempts on abstractive document summarization are still in a primitive stage, and the evaluation results are worse than extractive methods on benchmark datasets. In this paper, we review the difficulties of neural abstractive document summarization, and propose a novel graph-based attention mechanism in the sequence-to-sequence framework. The intuition is to address the saliency factor of summarization, which has been overlooked by prior works. Experimental results demonstrate our model is able to achieve considerable improvement over previous neural abstractive models. The data-driven neural abstractive method is also competitive with state-of-the-art extractive methods.

Probabilistic Typology: Deep Generative Models of Vowel Inventories
Ryan Cotterell and Jason Eisner 15:19–15:37

Linguistic typology studies the range of structures present in human language. The main goal of the field is to discover which sets of possible phenomena are universal, and which are merely frequent. For example, all languages have vowels, while most—but not all—languages have an /u/ sound. In this paper we present the first probabilistic treatment of a basic question in phonological typology: What makes a natural vowel inventory? We introduce a series of deep stochastic point processes, and contrast them with previous computational, simulation-based approaches. We provide a comprehensive suite of experiments on over 200 distinct languages.

Adversarial Multi-Criteria Learning for Chinese Word Segmentation
Xinchi Chen, Zhan Shi, Xipeng Qiu, and Xuanjing Huang 15:38–15:56

Different linguistic perspectives causes many diverse segmentation criteria for Chinese word segmentation (CWS). Most existing methods focus on improve the performance for each single criterion. However, it is interesting to exploit these different criteria and mining their common underlying knowledge. In this paper, we propose adversarial multi-criteria learning for CWS by integrating shared knowledge from multiple heterogeneous segmentation criteria. Experiments on eight corpora with heterogeneous segmentation criteria show that the performance of each corpus obtains a significant improvement, compared to single-criterion learning. Source codes of this paper are available on Github.

Neural Joint Model for Transition-based Chinese Syntactic Analysis
Shuhei Kurita, Daisuke Kawahara, and Sadao Kurohashi 15:57–16:15

We present neural network-based joint models for Chinese word segmentation, POS tagging and dependency parsing. Our models are the first neural approaches for fully joint Chinese analysis that is known to prevent the error propagation problem of pipeline models. Although word embeddings play a key role in dependency parsing, they cannot be applied directly to the joint task in the previous work. To address this problem, we propose embeddings of character strings, in addition to words. Experiments show that our models outperform existing systems in Chinese word segmentation and POS tagging, and perform preferable accuracies in dependency parsing. We also explore bi-LSTM models with fewer features.

Robust Incremental Neural Semantic Graph Parsing
Jan Buys and Phil Blunsom 16:16–16:34

Parsing sentences to linguistically-expressive semantic representations is a key goal of Natural Language Processing. Yet statistical parsing has focussed almost exclusively on bilexical dependencies or domain-specific logical forms. We propose a neural encoder-decoder transition-based parser which is the first full-coverage semantic graph parser for Minimal Recursion Semantics (MRS). The model architecture uses stack-based embedding features, predicting graphs jointly with unlexicalized predicates and their token alignments. Our parser is more accurate than attention-based baselines on MRS, and on an additional Abstract Meaning Representation (AMR) benchmark, and GPU batch processing makes it an order of magnitude faster than a high-precision grammar-based parser. Further, the 86.69% Smatch score of our MRS parser is higher than the upper-bound on AMR parsing, making MRS an attractive choice as a semantic representation.

Session Session 8B: Outstanding Papers 4

Joint Extraction of Entities and Relations Based on a Novel Tagging Scheme
Suncong Zheng, Feng Wang, Hongyun Bao, Yuexing Hao, Peng Zhou, and Bo Xu 15:00–15:18

Joint extraction of entities and relations is an important task in information extraction. To tackle this problem, we firstly propose a novel tagging scheme that can convert the joint extraction task to a tagging problem.. Then, based on our tagging scheme, we study different end-to-end models to extract entities and their relations directly, without identifying entities and relations separately. We conduct experiments on a public dataset produced by distant supervision method and the experimental results show that the tagging based methods are better than most of the existing pipelined and joint learning methods. What's more, the end-to-end model proposed in this paper, achieves the best results on the public dataset.

A Local Detection Approach for Named Entity Recognition and Mention Detection
Mingbin Xu, Hui Jiang, and Sedtawut Watcharawittayakul 15:19–15:37

In this paper, we study a novel approach for named entity recognition (NER) and mention detection (MD) in natural language processing. Instead of treating NER as a sequence labeling problem, we propose a new local detection approach, which relies on the recent fixed-size ordinally forgetting encoding (FOFE) method to fully encode each sentence fragment and its left/right contexts into a fixed-size representation. Subsequently, a simple feedforward neural network (FFNN) is learned to either reject or predict entity label for each individual text fragment. The proposed method has been evaluated in several popular NER and MD tasks, including CoNLL 2003 NER task and TAC-KBP2015 and TAC-KBP2016 Tri-lingual Entity Discovery and Linking (EDL) tasks. Our method has yielded pretty strong performance in all of these examined tasks. This local detection approach has shown many advantages over the traditional sequence labeling methods.

Vancouver Welcomes You! Minimalist Location Metonymy Resolution
Milan Gritta, Mohammad Taher Pilehvar, Nut Limsopatham, and Nigel Collier 15:38–15:56

Named entities are frequently used in a metonymic manner. They serve as references to related entities such as people and organisations. Accurate identification and interpretation of metonymy can be directly beneficial to various NLP applications, such as Named Entity Recognition and Geographical Parsing. Until now, metonymy resolution (MR) methods mainly relied on parsers, taggers, dictionaries, external word lists and other hand-crafted lexical resources. We show how a minimalist neural approach combined with a novel predicate window method can achieve competitive results on the SemEval 2007 task on Metonymy Resolution. Additionally, we contribute with a new Wikipedia-based MR dataset called RelocaR, which is tailored towards locations as well as improving previous deficiencies in annotation guidelines.

Unifying Text, Metadata, and User Network Representations with a Neural Network for Geolocation Prediction
Yasuhide Miura, Motoki Taniguchi, Tomoki Taniguchi, and Tomoko Ohkuma 15:57–16:15

We propose a novel geolocation prediction model using a complex neural network. Geolocation prediction in social media has attracted many researchers to use information of various types. Our model unifies text, metadata, and user network representations with an attention mechanism to overcome previous ensemble approaches. In an evaluation using two open datasets, the proposed model exhibited a maximum 3.8% increase in accuracy and a maximum of 6.6% increase in accuracy161 against previous models. We further analyzed several intermediate layers of our model, which revealed that their states capture some statistical characteristics of the datasets.

Multi-Task Video Captioning with Video and Entailment Generation
Ramakanth Pasunuru and Mohit Bansal 16:16–16:34

Video captioning, the task of describing the content of a video, has seen some promising improvements in recent years with sequence-to-sequence models, but accurately learning the temporal and logical dynamics involved in the task still remains a challenge, especially given the lack of sufficient annotated data. We improve video captioning by sharing knowledge with two related directed-generation tasks: a temporally-directed unsupervised video prediction task to learn richer context-aware video encoder representations, and a logically-directed language entailment generation task to learn better video-entailing caption decoder representations. For this, we present a many-to-many multi-task learning model that shares parameters across the encoders and decoders of the three tasks. We achieve significant improvements and the new state-of-the-art on several standard video captioning datasets using diverse automatic and human evaluations. We also show mutual multi-task improvements on the entailment generation task.

6

Workshops and Colocated Events

Sunday

Thursday–Friday

Thursday

Friday

WiNLP: Women & Underrepresented Minorities in Natural Language Processing

Organizers: *Libby Barak, Isabelle Augenstein, Chloé Braud, He He, and Margaret Mitchell*
Venue: Salon 2

Sunday, July 30, 2017

7:30–8:30 **ACL Registration**

8:30–10:35 **Session I: Opening, Invited Talk, Oral Presentations**

8:30–9:00 **Opening Remarks (Libby, Isabelle, Chloé, He, Margaret)**

9:00–9:35 **Invited Talk: Discourse and Computation: A life in tokens (Bonnie Webber)**

Oral Presentations

9:35–9:55 Grammatical gender associations outweigh topical gender bias in crosslinguistic word embeddings
Katherine McCurdy and Oğuz Serbetçi

9:55–10:15 Contextual and Structural Language Understanding in Dialogues
Yun-Nung Chen, Dilek Hakkani-Tur, Gokhan Tur, Jianfeng Gao, Asli Celikyilmaz, and Li Deng

10:15–10:35 Data-driven Natural Language Generation: Paving the Road to Success
Jekaterina Novikova, Ondřej Dušek, and Verena Rieser

10:35–11:00 **Coffee Break**

11:00–12:00 **Session II: Invited Talk, Mentoring**

11:00–11:35 **Invited Talk: Improving zero-shot learning for word-level translation (Ndapa Nakashole)**

11:35–12:00 **Mentoring Session (WiNLP Participants)**

12:00–13:00 **Lunch**

13:00–15:30 **Session III: Poster Presentations, Oral Presentations**

13:00–14:30 **Poster Session (see list below)**

14:30–15:30 **Oral Presentations**

14:30–14:50 Persuading People to Give: Modeling Persuasive Strategies via Semi Supervised Neural Nets
Diyi Yang

14:50–15:10 CANDiS: Coupled & Attention-Driven Neural Distant Supervision
Tushar Nagarajan, Sharmistha Jat, and Partha Talukdar

15:10–15:30 Semantic and Visual Cues for Humanitarian Computing of Natural Disaster Damage Images
Hadi Jomaa, Yara Rizk, and Mariette Awad

15:30–15:55 **Coffee Break**

15:55–18:00 **Session IV: Invited Talk, Oral Presentations, Closing, Mentor Meetup**

15:55–16:30 **Invited Talk: Empathetic Natural Language Processing (Pascale Fung)**

Oral Presentations

16:30–16:50 Detecting clinical conditions with Distributional Semantic Models
Felipe Paula, Rodrigo Wilkens, Marco Idiart, and Aline Villavicencio

16:50–17:10 Chasing Rhetorical Cues to Sarcasm in Online Debate
Shereen Oraby and Marilyn Walker

Closing

17:10–17:30 **Closing Remarks and Planning (Libby, Isabelle, Chloé, He, Margaret)**

17:30 **One-on-One Mentor Meetup**

18:00 **End of WiNLP**

18:00–21:30 **Session P1: Poster Session 1**

- Offensive Comments in the Brazilian Web: a dataset and baseline results
Viviane Moreira, Aline Villavicencio, and Rogers Prates de Pelle
- Planning in Language Grounded in Perception
Candace Ross, Andrei Barbu, and Boris Katz
- Empirical Evaluation of Named Entities on Arabic Sentiment Analysis of Levantine Dialect on Social Media
Hala Mulki
- Are you asking the right questions? Teaching Machines to Ask Clarification Questions
Sudha Rao and Hal Daumé III
- Sarcasm SIGN: Interpreting Sarcasm with Sentiment Based Monolingual Machine Translation
Lotem Peled and Roi Reichart
- Classify and Examine : Using Semi-supervised Learning with Ensemble Learning to Score Short-Answer Questions in Korean
Min-Ah Cheon, Chang-Hyun Kim, Jae-Hoon Kim, Eunhee Noh, Mi-Young Song, Jong-Im Park, Yuhyang Kim, and Do-Gil Lee
- Parameter Sharing for Morphological Generation Using Neural Sequence-to-Sequence Models
Katharina Kann, Ryan Cotterell, and Hinrich Schütze
- Convolutional Neural Networks for Churn Prediction in Microblogs
Mourad Gridach
- Using Distributional Semantics to Identify Idiomatic Language
Aline Villavicencio, Silvio Cordeiro, and Carlos Ramisch
- Neural Sentiment Classification with User and Product Attention
Huimin Chen, Maosong Sun, Cunchao Tu, Yankai Lin, and Zhiyuan Liu
- Adaptive Platform for Collaborative Translation
Kanika Kalra, Manasi Patwardhan, Rahul Kumar, Anand Sriraman, and Purushotam Radadia
- Name Variation in Community Question Answering Systems
Anietie Andy, Satoshi Sekine, Mugizi Rwebangira, and Mark Dredze
- Time Expression Resolution for Social Media Data
Jeniya Tabassum, Alan Ritter, and Wei Xu
- Neural Headline Generation with Sentence-wise Optimization
Yana A, Shiqi Shen, Zhiyuan Liu, and Maosong Sun
- Discovering Domain Specific Dialog Acts
Sabina Tomkins, Anbang Xu, Zhe Liu, and Yufan Guo
- Learning from Heterogeneous Data with PNN for Chinese SRL
Qiaolin Xia, Lei Sha, Baobao Chang, and Zhifang Sui

- Virtual Patient
  Yara Rizk, Khodor Kshoury, Mohamad Chehab, Petra Chidiac, Mariette Awad, and Joumana Antoun
- Using Fine-grained Phonetic Transcription to Improve the performance of Arabic Speech Recognition System
  Eiman Alsharhan
- Visual Verb Sense Disambiguation
  Spandana Gella
- From Characters to Words to in Between: Do We Capture Morphology?
  Clara Vania and Adam Lopez
- Regular Graph Grammars for NLP
  Sorcha Gilroy, Adam Lopez, and Sebastian Maneth
- Understanding Text Pre-processing for Latent Dirichlet Allocation
  Alexandra Schofield, Måns Magnusson, Laure Thompson, and David Mimno
- Combination of RNN and CRF for Extracting Adverse Drug Reactions from User Reviews
  Elena Tutubalina
- Who's to say what's funny? A computer using Language Models and Deep Learning, That's Who!
  Xinru Yan and Ted Pedersen
- Creation of an Annotated Corpus of Spanish Radiology Reports
  Viviana Cotik, Dario Filippo, Roland Roller, Feiyu Xu, and Hans Uszkoreit
- Argumentation Mining In Twitter: A Study Of Controversial Topics
  Aseel Addawood
- Affective Relationship Between Color and Text in Arabic Comic books
  Hadi Jomaa, Mohamad Kamereddine, Ammar Nayal, Yara Rizk, and Mariette Awad
- A Hybrid Approach to Recommender Systems for Technology Enhanced Learning: Using Probabilistic Topic Models, Collaborative Tagging and Domain Ontologies
  Lydia Odilinye and Fred Popowich
- Syntax Matters for Rhetorical Structure: The Case of Chiasmus
  Marie Dubremetz and Joakim Nivre
- Morph-Inflected Word Detection in Igbo via Bitext
  Enemouh Chioma, Mark Hepple, Ignatius Ezeani, and Ikechukwu Onyenwe
- Partial productivity of The Indefinite Noun Quantification Construction: implications for computational modeling in FrameNet Brasil
  Tatiane Tavares and Tiago Timponi Torrent
- Identifying frames at the sentence level in news articles
  Nona Naderi
- Unsupervised Word Discovery Using Attentional Encoder-Decoder Models
  Marcely Zanon Boito, Laurent Besacier, and Aline Villavicencio
- A Data-Driven Approach to Understanding Happiness
  Akari Asai, Vivian Li, Daniela Stepanov, and Wang-Chiew Tan
- Multi-task Domain Adaptation for Sequence Tagging
  Nanyun Peng and Mark Dredze
- Optimizing Statistical Machine Translation for Text Simplification
  Wei Xu, Courtney Napoles, Ellie Pavlick, Quanze Chen, and Chris Callison-Burch
- Copying Monolingual Data Improves Neural Machine Translation
  Anna Currey
- Insights on Sentiment Polarity Extraction from Bollywood Song Lyrics
  Drushti Apoorva and Radhika Mamidi

- Towards Efficient Human-Robot Dialogue Collection: Moving Fido into the Virtual World
  Cassidy Henry, Pooja Moolchandani, Kimberly A. Pollard, Claire Bonial, Ashley Foots, Ron Artstein, Cory Hayes, Clare R. Voss, David Traum, and Matthew Marge
- When does a Compliment become Sexist? Analysis and Classification of Ambivalent Sexism using Twitter Data
  Akshita Jha and Radhika Mamidi
- My Husband, the Dragonfly: The (Non-)Compositionality of Mansi Animal Names
  Veronika Vincze and Csilla Horváth
- Comment Relevance Classification in Facebook
  Chaya Liebeskind, Shmuel Liebeskind, and Yaakov HaCohen-Kerner
- Enterprise to Computer: Star Trek chatbot
  Grishma Jena, Mansi Vashisht, Abheek Basu, Joao Sedoc, and Lyle Ungar
- Neural Interactive Translation Prediction
  Rebecca Knowles and Philipp Koehn
- Psychological Stress Detection from Text Using Word Embeddings
  Genta Indra Winata, Onno Kampman, Yang Yang, and Pascale Fung
- WixNLP: Probabilistic Finite-State morphological analyzer for Wixarika language
  Jesús Manuel Mager Hois, Diónico Carrillo González, and Ivan Vladimir Meza Ruíz
- PubTermVariants: A Data-driven Approach to Finding Biomedical Term Variants
  Lana Yeganova, Won Kim, Sun Kim, Rezarta Islamaj Doğan, Wanli Liu, Donald Comeau, Zhiyong Lu, and W John Wilbur
- Safe Evaluation of Dialogue Management
  Layla El Asri and Adam Trischler
- What does an effective treatment recommendation look like: A corpus-based study on doctor-patient conversations
  Nan Wang and Yan Song
- Linguistic Theories and Resources for Natural Language Processing
  Miriam R L Petruck and Ellen K Dodge
- Argument Summarization in Social Media Dialogue
  Amita Misra and Marilyn Walker
- Functional Segment Identification and Emotion Tagging for Vietnamese Conversational Texts
  Ngo Thi Lan, Bui Van Hoang, Pham Khac Linh, Pham Bao Son, and Phan Xuan Hieu
- Learning Sentiment-Aware Word Representation in Compositional Matrix-Space Models
  Shima Asaadi and Sebastian Rudolph
- Exploring bilingual lexicon extraction for Spanish-Nahuatl
  Ximena Gutierrez-Vasques
- Sequential Approach to Rumour Stance Classification
  Elena Kochkina, Maria Liakata, and Isabelle Augenstein
- Comparison between Neural and Statistical translation after transliteration of Algerian Arabic Dialect
  Imane Guellil, Faical Azouaou, and Mourad Abbas
- How Biased Are We? Automated Detection of Gendered Language
  Ananya Ananya and Sameer Singh
- #MAGA or #TheResistance: Classifying Twitter users' political affiliation without looking at their words or friends
  Rachael Tatman

- Predicting Depression from User-generated Content in Social Media
  Hayda Almeida, Antoine Briand, and Marie-Jean Meurs
- User Trust When Machine Translation Goes Wrong
  Marianne Martindale and Marine Carpuat
- The role of domain in learning about referents
  Cassandra L. Jacobs
- Measuring Personality Adaptation in Task-Oriented Dialogs
  Zhichao Hu and Marilyn Walker
- Modeling of Political Discourse on Twitter
  Kristen Maria Johnson
- Collecting Artwork Descriptions Through Dialogue
  Lena Reed
- Bilingual Lexicon for Algerian Arabic Dialect Treatment in Social Media
  Imane Guellil and Faical Azouaou
- Debbie, the Debate Bot of the Future
  Geetanjali Rakshit, Kevin Bowden, Lena Reed, Amita Misra, and Marilyn Walker
- Generalized and Unsupervised Word-Level Language Detection
  Shruti Rijhwani, Royal Sequiera, Monojit Choudhury, and Kalika Bali
- Automatic Query Expansion for Patent Passage Retrieval using Paradigmatic and Syntagmatic Information
  Linda Andersson, Navid Rekabsaz, and Allan Hanbury

CoNLL: The SIGNLL Conference on Computational Natural Language Learning

Organizers: *Lucia Specia and Roger Levy*

Venue: Salon A

Thursday, August 3, 2017

8:45–9:00 **Opening Remarks**

Invited Talk by Chris Dyer

9:00–10:00 Should Neural Network Architecture Reflect Linguistic Structure?
Chris Dyer

Session 1

10:00–10:15 Exploring the Syntactic Abilities of RNNs with Multi-task Learning
Émile Enguehard, Yoav Goldberg, and Tal Linzen

Session 1L: Lightning Talks for Poster Session

10:15–10:17 The Effect of Different Writing Tasks on Linguistic Style: A Case Study of the ROC Story Cloze Task
Roy Schwartz, Maarten Sap, Ioannis Konstas, Leila Zilles, Yejin Choi, and Noah A. Smith

10:17–10:19 Parsing for Grammatical Relations via Graph Merging
Weiwei Sun, Yantao Du, and Xiaojun Wan

10:19–10:21 Leveraging Eventive Information for Better Metaphor Detection and Classification
I-Hsuan Chen, Yunfei Long, Qin Lu, and Chu-Ren Huang

10:21–10:23 Collaborative Partitioning for Coreference Resolution
Olga Uryupina and Alessandro Moschitti

10:23–10:25 Named Entity Disambiguation for Noisy Text
Yotam Eshel, Noam Cohen, Kira Radinsky, Shaul Markovitch, Ikuya Yamada, and Omer Levy

10:25–10:27 Tell Me Why: Using Question Answering as Distant Supervision for Answer Justification
Rebecca Sharp, Mihai Surdeanu, Peter Jansen, Marco A. Valenzuela-Escárcega, Peter Clark, and Michael Hammond

10:27–10:29 Learning What is Essential in Questions
Daniel Khashabi, Tushar Khot, Ashish Sabharwal, and Dan Roth

10:29–10:31 Top-Rank Enhanced Listwise Optimization for Statistical Machine Translation
Huadong Chen, Shujian Huang, David Chiang, Xin-Yu Dai, and Jiajun Chen

10:31–11:00 **Coffee Break**

Session ST1: CoNLL-SIGMORPHON Shared Task

11:00–12:30 Mans Hulden, Ryan Cotterell, Christo Kirov, and John Sylak-Glassman: Universal Morphological Reinflection in 52 Languages

12:30–14:00 **Lunch Break**

Session ST2: CoNLL Shared Task

14:00–15:30 Dan Zeman, Jan Hajič, et al.: Multilingual Parsing from Raw Text to Universal Dependencies

15:30–16:00 **Coffee Break**

Session 2

16:00–16:15 Embedding Words and Senses Together via Joint Knowledge-Enhanced Training
Massimiliano Mancini, Jose Camacho-Collados, Ignacio Iacobacci, and Roberto Navigli

16:15–16:30 Automatic Selection of Context Configurations for Improved Class-Specific Word Representations
Ivan Vulić, Roy Schwartz, Ari Rappoport, Roi Reichart, and Anna Korhonen

16:30–16:45 Modeling Context Words as Regions: An Ordinal Regression Approach to Word Embedding
Shoaib Jameel and Steven Schockaert

16:45–17:00 An Artificial Language Evaluation of Distributional Semantic Models
Fatemeh Torabi Asr and Michael Jones

17:00–17:15 Learning Word Representations with Regularization from Prior Knowledge
Yan Song, Chia-Jung Lee, and Fei Xia

Session 2L: Lightning Talks for Poster Session

17:15–17:17 Attention-based Recurrent Convolutional Neural Network for Automatic Essay Scoring
Fei Dong, Yue Zhang, and Jie Yang

17:17–17:19 Feature Selection as Causal Inference: Experiments with Text Classification
Michael J. Paul

17:19–17:21 A Joint Model for Semantic Sequences: Frames, Entities, Sentiments
Haoruo Peng, Snigdha Chaturvedi, and Dan Roth

17:21–17:23 Neural Sequence-to-sequence Learning of Internal Word Structure
Tatyana Ruzsics and Tanja Samardzic

17:23–17:25 A Supervised Approach to Extractive Summarisation of Scientific Papers
Ed Collins, Isabelle Augenstein, and Sebastian Riedel

17:25–17:27 An Automatic Approach for Document-level Topic Model Evaluation
Shraey Bhatia, Jey Han Lau, and Timothy Baldwin

17:27–17:29 Robust Coreference Resolution and Entity Linking on Dialogues: Character Identification on TV Show Transcripts
Henry Y. Chen, Ethan Zhou, and Jinho D. Choi

17:29–17:31 Cross-language Learning with Adversarial Neural Networks
Shafiq Joty, Preslav Nakov, Lluís Màrquez, and Israa Jaradat

17:31–18:31 **Business Meeting**

Friday, August 4, 2017

Invited talk by Naomi Feldman

8:45–9:45 Rational Distortions of Learners' Linguistic Input
Naomi Feldman

Session 3

9:45–10:00 Knowledge Tracing in Sequential Learning of Inflected Vocabulary
Adithya Renduchintala, Philipp Koehn, and Jason Eisner

10:00–10:15 A Probabilistic Generative Grammar for Semantic Parsing
Abulhair Saparov, Vijay Saraswat, and Tom M. Mitchell

Session 3L: Lightning Talks for Poster Session

10:15–10:17 Learning Contextual Embeddings for Structural Semantic Similarity using Categorical Information
Massimo Nicosia and Alessandro Moschitti

10:17–10:19 Making Neural QA as Simple as Possible but not Simpler
Dirk Weissenborn, Georg Wiese, and Laura Seiffe

10:19–10:21 Neural Domain Adaptation for Biomedical Question Answering
Georg Wiese, Dirk Weissenborn, and Mariana Neves

10:21–10:23 A phoneme clustering algorithm based on the obligatory contour principle
Mans Hulden

10:23–10:25 Learning Stock Market Sentiment Lexicon and Sentiment-Oriented Word Vector from StockTwits
Quanzhi Li and Sameena Shah

10:25–10:27 Learning local and global contexts using a convolutional recurrent network model for relation classification in biomedical text
Desh Raj, Sunil Sahu, and Ashish Anand

10:27–10:29 Idea density for predicting Alzheimer's disease from transcribed speech
Kairit Sirts, Olivier Piguet, and Mark Johnson

10:29–11:00 **Coffee Break**

11:00–14:00 **Poster Session & Lunch**

Session 4

14:00–14:15 Zero-Shot Relation Extraction via Reading Comprehension
Omer Levy, Minjoon Seo, Eunsol Choi, and Luke Zettlemoyer

14:15–14:30 The Covert Helps Parse the Overt
Xun Zhang, Weiwei Sun, and Xiaojun Wan

14:30–14:45 German in Flux: Detecting Metaphoric Change via Word Entropy
Dominik Schlechtweg, Stefanie Eckmann, Enrico Santus, Sabine Schulte im Walde, and Daniel Hole

14:45–15:00 Encoding of phonology in a recurrent neural model of grounded speech
Afra Alishahi, Marie Barking, and Grzegorz Chrupała

15:00–15:15 Multilingual Semantic Parsing And Code-Switching
Long Duong, Hadi Afshar, Dominique Estival, Glen Pink, Philip Cohen, and Mark Johnson

15:15–15:30 Optimizing Differentiable Relaxations of Coreference Evaluation Metrics
Phong Le and Ivan Titov

15:30–16:00 **Coffee Break**

Session 5

16:00–17:15 Neural Structural Correspondence Learning for Domain Adaptation
Yftah Ziser and Roi Reichart

16:15–16:30 A Simple and Accurate Syntax-Agnostic Neural Model for Dependency-based Semantic Role Labeling
Diego Marcheggiani, Anton Frolov, and Ivan Titov

16:30–17:45 Joint Prediction of Morphosyntactic Categories for Fine-Grained Arabic Part-of-Speech Tagging Exploiting Tag Dictionary Information
Go Inoue, Hiroyuki Shindo, and Yuji Matsumoto

16:45–17:00 Learning from Relatives: Unified Dialectal Arabic Segmentation
Younes Samih, Mohamed Eldesouki, Mohammed Attia, Kareem Darwish, Ahmed Abdelali, Hamdy Mubarak, and Laura Kallmeyer

17:00–17:15 Natural Language Generation for Spoken Dialogue System using RNN Encoder-Decoder Networks
Van-Khanh Tran and Le-Minh Nguyen

17:15–17:30 Graph-based Neural Multi-Document Summarization
Michihiro Yasunaga, Rui Zhang, Kshitijh Meelu, Ayush Pareek, Krishnan Srinivasan, and Dragomir Radev

17:30–17:35 **Best Paper Award**

17:35–17:45 **Closing Remarks**

*SEM: Sixth Joint Conference On Lexical And Computational Semantics

Organizers: *Nancy Ide, Aurélie Herbelot, and Lluís Màrquez*
Venue: Salon 1

Thursday, August 3, 2017

9:00–10:30 **Session S1: Invited Talk (Jointly with SemEval) and Best Paper Award**

9:00–9:15 **Opening Remarks**

9:15–10:15 **Invited Talk: From Naive Physics to Connotation: Modeling Commonsense in Frame Semantics (Yejin Choi)**

10:15–10:30 **Announcement of the Adam Kilgarriff Best Paper Award**

10:30–11:00 **Coffee Break**

11:00–12:30 **Session S2: Distributional Semantics**

11:00–11:30 What Analogies Reveal about Word Vectors and their Compositionality
Gregory Finley, Stephanie Farmer, and Serguei Pakhomov

11:30–12:00 Learning Antonyms with Paraphrases and a Morphology-Aware Neural Network
Sneha Rajana, Chris Callison-Burch, Marianna Apidianaki, and Vered Shwartz

12:00–12:30 Decoding Sentiment from Distributed Representations of Sentences
Edoardo Maria Ponti, Ivan Vulić, and Anna Korhonen

12:30–14:00 **Lunch Break**

14:00–15:30 **Session S3: Lexical Semantics and Lexical Resources**

14:00–14:30 Detecting Asymmetric Semantic Relations in Context: A Case-Study on Hypernymy Detection
Yogarshi Vyas and Marine Carpuat

14:30–15:00 Domain-Specific New Words Detection in Chinese
Ao Chen and Maosong Sun

15:00–15:30 Deep Learning Models For Multiword Expression Identification
Waseem Gharbieh, Virendrakumar Bhavsar, and Paul Cook

15:30–16:00 **Coffee Break**

16:00–16:30 **Session S4: Lexical Semantics and Lexical Resources (continued)**

16:00–16:30 Emotion Intensities in Tweets
Saif M. Mohammad and Felipe Bravo-Marquez

16:30–6:00 **Session S5: Poster Session**

- Deep Active Learning for Dialogue Generation
Nabiha Asghar, Pascal Poupart, Xin Jiang, and Hang Li
- Mapping the Paraphrase Database to WordNet
Anne Cocos, Marianna Apidianaki, and Chris Callison-Burch
- Semantic Frame Labeling with Target-based Neural Model
Yukun Feng, Dong Yu, Jian Xu, and Chunhua Liu

- Frame-Based Continuous Lexical Semantics through Exponential Family Tensor Factorization and Semantic Proto-Roles
  Francis Ferraro, Adam Poliak, Ryan Cotterell, and Benjamin Van Durme
- Distributed Prediction of Relations for Entities: The Easy, The Difficult, and The Impossible
  Abhijeet Gupta, Gemma Boleda, and Sebastian Padó
- Comparing Approaches for Automatic Question Identification
  Angel Maredia, Kara Schechtman, Sarah Ita Levitan, and Julia Hirschberg
- Does Free Word Order Hurt? Assessing the Practical Lexical Function Model for Croatian
  Zoran Medić, Jan Šnajder, and Sebastian Padó
- A Mixture Model for Learning Multi-Sense Word Embeddings
  Dai Quoc Nguyen, Dat Quoc Nguyen, Ashutosh Modi, Stefan Thater, and Manfred Pinkal
- Aligning Script Events with Narrative Texts
  Simon Ostermann, Michael Roth, Stefan Thater, and Manfred Pinkal
- The (too Many) Problems of Analogical Reasoning with Word Vectors
  Anna Rogers, Aleksandr Drozd, and Bofang Li
- Semantic Frames and Visual Scenes: Learning Semantic Role Inventories from Image and Video Descriptions
  Ekaterina Shutova, Andreas Wundsam, and Helen Yannakoudakis
- Acquiring Predicate Paraphrases from News Tweets
  Vered Shwartz, Gabriel Stanovsky, and Ido Dagan
- Evaluating Semantic Parsing against a Simple Web-based Question Answering Model
  Alon Talmor, Mor Geva, and Jonathan Berant

Friday, August 4, 2017

9:00–10:30 **Session S6: Invited Talk and Distributional Semantics**

9:00–10:00 **Invited Talk: What Do You Know About an Alligator When You Know the Company It Keeps? (Katrin Erk)**

10:00–10:30 Logical Metonymy in a Distributional Model of Sentence Comprehension
Emmanuele Chersoni, Alessandro Lenci, and Philippe Blache

10:30–11:00 **Coffee Break**

11:00–12:30 **Session S7: Linguistic and Formal Semantics**

11:00–11:30 Double Trouble: The Problem of Construal in Semantic Annotation of Adpositions
Jena D. Hwang, Archna Bhatia, Na-Rae Han, Tim O'Gorman, Vivek Srikumar, and Nathan Schneider

11:30–12:00 Issues of Mass and Count: Dealing with 'Dual-Life' Nouns
Tibor Kiss, Francis Jeffry Pelletier, Halima Husic, and Johanna Poppek

12:00–12:30 Parsing Graphs with Regular Graph Grammars
Sorcha Gilroy, Adam Lopez, and Sebastian Maneth

12:30–14:00 **Lunch Break**

14:00–15:30 **Session S8: Representations of Meaning**

14:00–14:30 Embedded Semantic Lexicon Induction with Joint Global and Local Optimization
Sujay Kumar Jauhar and Eduard Hovy

14:30–15:00 Generating Pattern-Based Entailment Graphs for Relation Extraction
Kathrin Eichler, Feiyu Xu, Hans Uszkoreit, and Sebastian Krause

15:00–15:30 Classifying Semantic Clause Types: Modeling Context and Genre Characteristics with Recurrent Neural Networks and Attention
Maria Becker, Michael Staniek, Vivi Nastase, Alexis Palmer, and Anette Frank

15:30–16:00 **Coffee Break**

16:00–17:30 **Session S9: Semantics for Applications**

16:00–16:30 Predictive Linguistic Features of Schizophrenia
Efsun Sarioglu Kayi, Mona Diab, Luca Pauselli, Michael Compton, and Glen Coppersmith

16:30–17:00 Learning to Solve Geometry Problems from Natural Language Demonstrations in Textbooks
Mrinmaya Sachan and Eric Xing

17:00–17:30 Ways of Asking and Replying in Duplicate Question Detection
João António Rodrigues, Chakaveh Saedi, Vladislav Maraev, João Silva, and António Branco

17:30–17:40 **Closing Remarks**

SemEval: 11th International Workshop on Semantic Evaluation

Organizers: *Steven Bethard, Marine Carpuat, Marianna Apidianaki, Saif M. Mohammad, Daniel Cer, and David Jurgens*

Venue: Salon B

Thursday, August 3, 2017

9:00–9:15 **Welcome / Opening Remarks**

9:15–10:30 **Invited Talk: From Naive Physics to Connotation: Modeling Commonsense in Frame Semantics (Yejin Choi)**

10:30–11:00 **Coffee**

11:00–12:30 **Task Descriptions**

11:00–11:15 SemEval-2017 Task 1: Semantic Textual Similarity Multilingual and Crosslingual Focused Evaluation
Daniel Cer, Mona Diab, Eneko Agirre, Inigo Lopez-Gazpio, and Lucia Specia

11:15–11:30 SemEval-2017 Task 2: Multilingual and Cross-lingual Semantic Word Similarity
Jose Camacho-Collados, Mohammad Taher Pilehvar, Nigel Collier, and Roberto Navigli

11:30–11:45 SemEval-2017 Task 3: Community Question Answering
Preslav Nakov, Doris Hoogeveen, Lluís Màrquez, Alessandro Moschitti, Hamdy Mubarak, Timothy Baldwin, and Karin Verspoor

11:45–12:00 SemEval-2017 Task 6: #HashtagWars: Learning a Sense of Humor
Peter Potash, Alexey Romanov, and Anna Rumshisky

12:00–12:15 SemEval-2017 Task 7: Detection and Interpretation of English Puns
Tristan Miller, Christian Hempelmann, and Iryna Gurevych

12:15–12:30 SemEval-2017 Task 8: RumourEval: Determining rumour veracity and support for rumours
Leon Derczynski, Kalina Bontcheva, Maria Liakata, Rob Procter, Geraldine Wong Sak Hoi, and Arkaitz Zubiaga

12:30–14:00 **Lunch**

14:00–15:30 **Best Of SemEval**

14:00–14:15 BIT at SemEval-2017 Task 1: Using Semantic Information Space to Evaluate Semantic Textual Similarity
Hao Wu, Heyan Huang, Ping Jian, Yuhang Guo, and Chao Su

14:15–14:30 ConceptNet at SemEval-2017 Task 2: Extending Word Embeddings with Multilingual Relational Knowledge
Robert Speer and Joanna Lowry-Duda

14:30–14:45 IIT-UHH at SemEval-2017 Task 3: Exploring Multiple Features for Community Question Answering and Implicit Dialogue Identification
Titas Nandi, Chris Biemann, Seid Muhie Yimam, Deepak Gupta, Sarah Kohail, Asif Ekbal, and Pushpak Bhattacharyya

14:45–15:00 HumorHawk at SemEval-2017 Task 6: Mixing Meaning and Sound for Humor Recognition
David Donahue, Alexey Romanov, and Anna Rumshisky

15:00–15:15 Idiom Savant at Semeval-2017 Task 7: Detection and Interpretation of English Puns
Samuel Doogan, Aniruddha Ghosh, Hanyang Chen, and Tony Veale

15:15–15:30 Turing at SemEval-2017 Task 8: Sequential Approach to Rumour Stance Classification with Branch-LSTM
Elena Kochkina, Maria Liakata, and Isabelle Augenstein

15:30–16:00 **Coffee**

16:00–16:30 **Discussion**

16:30–17:30 **Poster Session**

- CompiLIG at SemEval-2017 Task 1: Cross-Language Plagiarism Detection Methods for Semantic Textual Similarity
  Jérémy Ferrero, Laurent Besacier, Didier Schwab, and Frédéric Agnès
- UdL at SemEval-2017 Task 1: Semantic Textual Similarity Estimation of English Sentence Pairs Using Regression Model over Pairwise Features
  Hussein T. Al-Natsheh, Lucie Martinet, Fabrice Muhlenbach, and Djamel Abdelkader Zighed
- DT_Team at SemEval-2017 Task 1: Semantic Similarity Using Alignments, Sentence-Level Embeddings and Gaussian Mixture Model Output
  Nabin Maharjan, Rajendra Banjade, Dipesh Gautam, Lasang J. Tamang, and Vasile Rus
- FCICU at SemEval-2017 Task 1: Sense-Based Language Independent Semantic Textual Similarity Approach
  Basma Hassan, Samir AbdelRahman, Reem Bahgat, and Ibrahim Farag
- HCTI at SemEval-2017 Task 1: Use convolutional neural network to evaluate Semantic Textual Similarity
  Yang Shao
- LIM-LIG at SemEval-2017 Task1: Enhancing the Semantic Similarity for Arabic Sentences with Vectors Weighting
  El Moatez Billah Nagoudi, Jérémy Ferrero, and Didier Schwab
- OPI-JSA at SemEval-2017 Task 1: Application of Ensemble learning for computing semantic textual similarity
  Martyna Śpiewak, Piotr Sobecki, and Daniel Karaś
- Lump at SemEval-2017 Task 1: Towards an Interlingua Semantic Similarity
  Cristina España-Bonet and Alberto Barrón-Cedeño
- QLUT at SemEval-2017 Task 1: Semantic Textual Similarity Based on Word Embeddings
  Fanqing Meng, Wenpeng Lu, Yuteng Zhang, Jinyong Cheng, Yuehan Du, and Shuwang Han
- ResSim at SemEval-2017 Task 1: Multilingual Word Representations for Semantic Textual Similarity
  Johannes Bjerva and Robert Östling
- ITNLP-AiKF at SemEval-2017 Task 1: Rich Features Based SVR for Semantic Textual Similarity Computing
  Wenjie Liu, Chengjie Sun, Lei Lin, and Bingquan Liu
- Neobility at SemEval-2017 Task 1: An Attention-based Sentence Similarity Model
  WenLi Zhuang and Ernie Chang

- SEFUHH at SemEval-2017 Task 1: Unsupervised Knowledge-Free Semantic Textual Similarity via Paragraph Vector
  Mirela-Stefania Duma and Wolfgang Menzel
- STS-UHH at SemEval-2017 Task 1: Scoring Semantic Textual Similarity Using Supervised and Unsupervised Ensemble
  Sarah Kohail, Amr Rekaby Salama, and Chris Biemann
- UMDeep at SemEval-2017 Task 1: End-to-End Shared Weight LSTM Model for Semantic Textual Similarity
  Joe Barrow and Denis Peskov
- MITRE at SemEval-2017 Task 1: Simple Semantic Similarity
  John Henderson, Elizabeth Merkhofer, Laura Strickhart, and Guido Zarrella
- ECNU at SemEval-2017 Task 1: Leverage Kernel-based Traditional NLP features and Neural Networks to Build a Universal Model for Multilingual and Cross-lingual Semantic Textual Similarity
  Junfeng Tian, Zhiheng Zhou, Man Lan, and Yuanbin Wu
- PurdueNLP at SemEval-2017 Task 1: Predicting Semantic Textual Similarity with Paraphrase and Event Embeddings
  I-Ta Lee, Mahak Goindani, Chang Li, Di Jin, Kristen Marie Johnson, Xiao Zhang, Maria Leonor Pacheco, and Dan Goldwasser
- RTM at SemEval-2017 Task 1: Referential Translation Machines for Predicting Semantic Similarity
  Ergun Biçici
- LIPN-IIMAS at SemEval-2017 Task 1: Subword Embeddings, Attention Recurrent Neural Networks and Cross Word Alignment for Semantic Textual Similarity
  Ignacio Arroyo-Fernández and Ivan Vladimir Meza Ruiz
- L2F/INESC-ID at SemEval-2017 Tasks 1 and 2: Lexical and semantic features in word and textual similarity
  Pedro Fialho, Hugo Patinho Rodrigues, Luísa Coheur, and Paulo Quaresma
- HCCL at SemEval-2017 Task 2: Combining Multilingual Word Embeddings and Transliteration Model for Semantic Similarity
  Junqing He, Long Wu, Xuemin Zhao, and Yonghong Yan
- Citius at SemEval-2017 Task 2: Cross-Lingual Similarity from Comparable Corpora and Dependency-Based Contexts
  Pablo Gamallo
- Jmp8 at SemEval-2017 Task 2: A simple and general distributional approach to estimate word similarity
  Josué Melka and Gilles Bernard
- QLUT at SemEval-2017 Task 2: Word Similarity Based on Word Embedding and Knowledge Base
  Fanqing Meng, Wenpeng Lu, Yuteng Zhang, Ping Jian, Shumin Shi, and Heyan Huang
- RUFINO at SemEval-2017 Task 2: Cross-lingual lexical similarity by extending PMI and word embeddings systems with a Swadesh's-like list
  Sergio Jimenez, George Dueñas, Lorena Gaitan, and Jorge Segura
- MERALI at SemEval-2017 Task 2 Subtask 1: a Cognitively Inspired approach
  Enrico Mensa, Daniele P. Radicioni, and Antonio Lieto
- HHU at SemEval-2017 Task 2: Fast Hash-Based Embeddings for Semantic Word Similarity Assessment
  Behrang QasemiZadeh and Laura Kallmeyer

- Mahtab at SemEval-2017 Task 2: Combination of Corpus-based and Knowledge-based Methods to Measure Semantic Word Similarity
  Niloofar Ranjbar, Fatemeh Mashhadirajab, Mehrnoush Shamsfard, Rayeheh Hosseini pour, and Aryan Vahid pour
- Sew-Embed at SemEval-2017 Task 2: Language-Independent Concept Representations from a Semantically Enriched Wikipedia
  Claudio Delli Bovi and Alessandro Raganato
- Wild Devs' at SemEval-2017 Task 2: Using Neural Networks to Discover Word Similarity
  Răzvan-Gabriel Rotari, Ionut Hulub, Stefan Oprea, Mihaela Plamada-Onofrei, Alina Beatrice Lorent, Raluca Preisler, Adrian Iftene, and Diana Trandabat
- TrentoTeam at SemEval-2017 Task 3: An application of Grice Maxims in Ranking Community Question Answers
  Mohammed R. H. Qwaider, Abed Alhakim Freihat, and Fausto Giunchiglia
- UPC-USMBA at SemEval-2017 Task 3: Combining multiple approaches for CQA for Arabic
  Yassine El Adlouni, Imane Lahbari, Horacio Rodriguez, Mohammed Meknassi, Said Ouatik El Alaoui, and Noureddine Ennahnahi
- Beihang-MSRA at SemEval-2017 Task 3: A Ranking System with Neural Matching Features for Community Question Answering
  Wenzheng Feng, Yu Wu, Wei Wu, Zhoujun Li, and Ming Zhou
- MoRS at SemEval-2017 Task 3: Easy to use SVM in Ranking Tasks
  Miguel J. Rodrigues and Francisco M Couto
- EICA Team at SemEval-2017 Task 3: Semantic and Metadata-based Features for Community Question Answering
  Yufei Xie, Maoquan Wang, Jing Ma, Jian Jiang, and Zhao Lu
- FA3L at SemEval-2017 Task 3: A ThRee Embeddings Recurrent Neural Network for Question Answering
  Giuseppe Attardi, Antonio Carta, Federico Errica, Andrea Madotto, and Ludovica Pannitto
- SCIR-QA at SemEval-2017 Task 3: CNN Model Based on Similar and Dissimilar Information between Keywords for Question Similarity
  Le Qi, Yu Zhang, and Ting Liu
- LearningToQuestion at SemEval 2017 Task 3: Ranking Similar Questions by Learning to Rank Using Rich Features
  Naman Goyal
- SimBow at SemEval-2017 Task 3: Soft-Cosine Semantic Similarity between Questions for Community Question Answering
  Delphine Charlet and Geraldine Damnati
- FuRongWang at SemEval-2017 Task 3: Deep Neural Networks for Selecting Relevant Answers in Community Question Answering
  Sheng Zhang, Jiajun Cheng, Hui Wang, Xin Zhang, Pei Li, and Zhaoyun Ding
- KeLP at SemEval-2017 Task 3: Learning Pairwise Patterns in Community Question Answering
  Simone Filice, Giovanni Da San Martino, and Alessandro Moschitti
- SwissAlps at SemEval-2017 Task 3: Attention-based Convolutional Neural Network for Community Question Answering
  Jan Milan Deriu and Mark Cieliebak
- TakeLab-QA at SemEval-2017 Task 3: Classification Experiments for Answer Retrieval in Community QA
  Filip Šaina, Toni Kukurin, Lukrecija Puljić, Mladen Karan, and Jan Šnajder

- GW_QA at SemEval-2017 Task 3: Question Answer Re-ranking on Arabic Fora
  Nada Almarwani and Mona Diab
- NLM_NIH at SemEval-2017 Task 3: from Question Entailment to Question Similarity for Community Question Answering
  Asma Ben Abacha and Dina Demner-Fushman
- bunji at SemEval-2017 Task 3: Combination of Neural Similarity Features and Comment Plausibility Features
  Yuta Koreeda, Takuya Hashito, Yoshiki Niwa, Misa Sato, Toshihiko Yanase, Kenzo Kurotsuchi, and Kohsuke Yanai
- QU-BIGIR at SemEval 2017 Task 3: Using Similarity Features for Arabic Community Question Answering Forums
  Marwan Torki, Maram Hasanain, and Tamer Elsayed
- ECNU at SemEval-2017 Task 3: Using Traditional and Deep Learning Methods to Address Community Question Answering Task
  Guoshun Wu, Yixuan Sheng, Man Lan, and Yuanbin Wu
- UINSUSKA-TiTech at SemEval-2017 Task 3: Exploiting Word Importance Levels for Similarity Features for CQA
  Surya Agustian and Hiroya Takamura
- Talla at SemEval-2017 Task 3: Identifying Similar Questions Through Paraphrase Detection
  Byron Galbraith, Bhanu Pratap, and Daniel Shank
- QUB at SemEval-2017 Task 6: Cascaded Imbalanced Classification for Humor Analysis in Twitter
  Xiwu Han and Gregory Toner
- Duluth at SemEval-2017 Task 6: Language Models in Humor Detection
  Xinru Yan and Ted Pedersen
- DataStories at SemEval-2017 Task 6: Siamese LSTM with Attention for Humorous Text Comparison
  Christos Baziotis, Nikos Pelekis, and Christos Doulkeridis
- TakeLab at SemEval-2017 Task 6: #RankingHumorIn4Pages
  Marin Kukovačec, Juraj Malenica, Ivan Mršić, Antonio Šajatović, Domagoj Alagić, and Jan Šnajder
- SRHR at SemEval-2017 Task 6: Word Associations for Humour Recognition
  Andrew Cattle and Xiaojuan Ma
- #WarTeam at SemEval-2017 Task 6: Using Neural Networks for Discovering Humorous Tweets
  Iuliana Alexandra Fleşcan-Lovin-Arseni, Ramona Andreea Turcu, Cristina Sirbu, Larisa Alexa, Sandra Maria Amarandei, Nichita Herciu, Constantin Scutaru, Diana Trandabat, and Adrian Iftene
- SVNIT SemEval 2017 Task-6: Learning a Sense of Humor Using Supervised Approach
  Rutal Mahajan and Mukesh Zaveri
- Duluth at SemEval-2017 Task 7 : Puns Upon a Midnight Dreary, Lexical Semantics for the Weak and Weary
  Ted Pedersen
- UWaterloo at SemEval-2017 Task 7: Locating the Pun Using Syntactic Characteristics and Corpus-based Metrics
  Olga Vechtomova
- PunFields at SemEval-2017 Task 7: Employing Roget's Thesaurus in Automatic Pun Recognition and Interpretation
  Elena Mikhalkova and Yuri Karyakin

- JU CSE NLP SemEval 2017 Task 7: Employing Rules to Detect and Interpret English Puns
  Aniket Pramanick and Dipankar Das
- N-Hance at SemEval-2017 Task 7: A Computational Approach using Word Association for Puns
  Özge Sevgili, Nima Ghotbi, and Selma Tekir
- ELiRF-UPV at SemEval-2017 Task 7: Pun Detection and Interpretation
  Lluís-F. Hurtado, Encarna Segarra, Ferran Pla, Pascual Carrasco, and José-Ángel González
- BuzzSaw at SemEval-2017 Task 7: Global vs. Local Context for Interpreting and Locating Homographic English Puns with Sense Embeddings
  Dieke Oele and Kilian Evang
- UWAV at SemEval-2017 Task 7: Automated feature-based system for locating puns
  Ankit Vadehra
- ECNU at SemEval-2017 Task 7: Using Supervised and Unsupervised Methods to Detect and Locate English Puns
  Yuhuan Xiu, Man Lan, and Yuanbin Wu
- Fermi at SemEval-2017 Task 7: Detection and Interpretation of Homographic puns in English Language
  Vijayasaradhi Indurthi and Subba Reddy Oota
- UWaterloo at SemEval-2017 Task 8: Detecting Stance towards Rumours with Topic Independent Features
  Hareesh Bahuleyan and Olga Vechtomova
- IKM at SemEval-2017 Task 8: Convolutional Neural Networks for stance detection and rumor verification
  Yi-Chin Chen, Zhao-Yang Liu, and Hung-Yu Kao
- NileTMRG at SemEval-2017 Task 8: Determining Rumour and Veracity Support for Rumours on Twitter.
  Omar Enayet and Samhaa R. El-Beltagy
- Turing at SemEval-2017 Task 8: Sequential Approach to Rumour Stance Classification with Branch-LSTM
  Elena Kochkina, Maria Liakata, and Isabelle Augenstein
- Mama Edha at SemEval-2017 Task 8: Stance Classification with CNN and Rules
  Marianela García Lozano, Hanna Lilja, Edward Tjörnhammar, and Maja Karasalo
- DFKI-DKT at SemEval-2017 Task 8: Rumour Detection and Classification using Cascading Heuristics
  Ankit Srivastava, Georg Rehm, and Julian Moreno Schneider
- ECNU at SemEval-2017 Task 8: Rumour Evaluation Using Effective Features and Supervised Ensemble Models
  Feixiang Wang, Man Lan, and Yuanbin Wu
- IITP at SemEval-2017 Task 8 : A Supervised Approach for Rumour Evaluation
  Vikram Singh, Sunny Narayan, Md Shad Akhtar, Asif Ekbal, and Pushpak Bhattacharyya

Friday, August 4, 2017

9:00–9:30 **SemEval 2018 Tasks**

9:30–10:30 **State of SemEval Discussion**

10:30–11:00 **Coffee**

11:00–12:30 **Task Descriptions**

11:00–11:15 SemEval-2017 Task 4: Sentiment Analysis in Twitter
Sara Rosenthal, Noura Farra, and Preslav Nakov

11:15–11:30 SemEval-2017 Task 5: Fine-Grained Sentiment Analysis on Financial Microblogs and News
Keith Cortis, André Freitas, Tobias Daudert, Manuela Huerlimann, Manel Zarrouk, Siegfried Handschuh, and Brian Davis

11:30–11:45 SemEval-2017 Task 9: Abstract Meaning Representation Parsing and Generation
Jonathan May and Jay Priyadarshi

11:45–12:00 SemEval 2017 Task 10: ScienceIE - Extracting Keyphrases and Relations from Scientific Publications
Isabelle Augenstein, Mrinal Das, Sebastian Riedel, Lakshmi Vikraman, and Andrew McCallum

12:00–12:15 SemEval-2017 Task 11: End-User Development using Natural Language
Juliano Sales, Siegfried Handschuh, and André Freitas

12:15–12:30 SemEval-2017 Task 12: Clinical TempEval
Steven Bethard, Guergana Savova, Martha Palmer, and James Pustejovsky

12:30–14:00 **Lunch**

14:00–15:30 **Best Of SemEval**

14:00–14:15 BB_twtr at SemEval-2017 Task 4: Twitter Sentiment Analysis with CNNs and LSTMs
Mathieu Cliche

14:15–14:30 Lancaster A at SemEval-2017 Task 5: Evaluation metrics matter: predicting sentiment from financial news headlines
Andrew Moore and Paul Rayson

14:30–14:45 Sheffield at SemEval-2017 Task 9: Transition-based language generation from AMR.
Gerasimos Lampouras and Andreas Vlachos

14:45–15:00 The AI2 system at SemEval-2017 Task 10 (ScienceIE): semi-supervised end-to-end entity and relation extraction
Waleed Ammar, Matthew Peters, Chandra Bhagavatula, and Russell Power

15:00–15:15 LIMSI-COT at SemEval-2017 Task 12: Neural Architecture for Temporal Information Extraction from Clinical Narratives
Julien Tourille, Olivier Ferret, Xavier Tannier, and Aurélie Névéol

15:30–16:00 **Coffee**

16:00–16:30 **Discussion**

16:30–17:30 **Poster Session**

- OMAM at SemEval-2017 Task 4: Evaluation of English State-of-the-Art Sentiment Analysis Models for Arabic and a New Topic-based Model
Ramy Baly, Gilbert Badaro, Ali Hamdi, Rawan Moukalled, Rita Aoun, Georges El-Khoury, Ahmad Al Sallab, Hazem Hajj, Nizar Habash, Khaled Shaban, and Wassim El-Hajj

- NILC-USP at SemEval-2017 Task 4: A Multi-view Ensemble for Twitter Sentiment Analysis
  Edilson Anselmo Corrêa Júnior, Vanessa Queiroz Marinho, and Leandro Borges dos Santos
- deepSA at SemEval-2017 Task 4: Interpolated Deep Neural Networks for Sentiment Analysis in Twitter
  Tzu-Hsuan Yang, Tzu-Hsuan Tseng, and Chia-Ping Chen
- NNEMBs at SemEval-2017 Task 4: Neural Twitter Sentiment Classification: a Simple Ensemble Method with Different Embeddings
  Yichun Yin, Yangqiu Song, and Ming Zhang
- CrystalNest at SemEval-2017 Task 4: Using Sarcasm Detection for Enhancing Sentiment Classification and Quantification
  Raj Kumar Gupta and Yinping Yang
- SINAI at SemEval-2017 Task 4: User based classification
  Salud María Jiménez-Zafra, Arturo Montejo-Ráez, Maite Martin, and L. Alfonso Urena Lopez
- HLPUPenn at SemEval-2017 Task 4A: A simple, self-optimizing text classification system combining dense and sparse vectors
  Abeed Sarker and Graciela Gonzalez
- ej-sa-2017 at SemEval-2017 Task 4: Experiments for Target oriented Sentiment Analysis in Twitter
  Enkhzol Dovdon and José Saias
- SentiME++ at SemEval-2017 Task 4: Stacking State-of-the-Art Classifiers to Enhance Sentiment Classification
  Raphael Troncy, Enrico Palumbo, Efstratios Sygkounas, and Giuseppe Rizzo
- Amobee at SemEval-2017 Task 4: Deep Learning System for Sentiment Detection on Twitter
  Alon Rozental and Daniel Fleischer
- TWINA at SemEval-2017 Task 4: Twitter Sentiment Analysis with Ensemble Gradient Boost Tree Classifier
  Naveen Kumar Laskari and Suresh Kumar Sanampudi
- Tw-StAR at SemEval-2017 Task 4: Sentiment Classification of Arabic Tweets
  Hala Mulki, Hatem Haddad, Mourad Gridach, and Ismail Babaoğlu
- OMAM at SemEval-2017 Task 4: English Sentiment Analysis with Conditional Random Fields
  Chukwuyem Onyibe and Nizar Habash
- Tweester at SemEval-2017 Task 4: Fusion of Semantic-Affective and pairwise classification models for sentiment analysis in Twitter
  Athanasia Kolovou, Filippos Kokkinos, Aris Fergadis, Pinelopi Papalampidi, Elias Iosif, Nikolaos Malandrakis, Elisavet Palogiannidi, Haris Papageorgiou, Shrikanth Narayanan, and Alexandros Potamianos
- NRU-HSE at SemEval-2017 Task 4: Tweet Quantification Using Deep Learning Architecture
  Nikolay Karpov
- MI&T Lab at SemEval-2017 task 4: An Integrated Training Method of Word Vector for Sentiment Classification
  Jingjing Zhao, Yan Yang, and Bing Xu
- SiTAKA at SemEval-2017 Task 4: Sentiment Analysis in Twitter Based on a Rich Set of Features
  Mohammed Jabreel and Antonio Moreno

- Senti17 at SemEval-2017 Task 4: Ten Convolutional Neural Network Voters for Tweet Polarity Classification
  Hussam Hamdan
- DUTH at SemEval-2017 Task 4: A Voting Classification Approach for Twitter Sentiment Analysis
  Symeon Symeonidis, Dimitrios Effrosynidis, John Kordonis, and Avi Arampatzis
- SSN_MLRG1 at SemEval-2017 Task 4: Sentiment Analysis in Twitter Using Multi-Kernel Gaussian Process Classifier
  Angel Deborah S, S Milton Rajendram, and T T Mirnalinee
- YNUDLG at SemEval-2017 Task 4: A GRU-SVM Model for Sentiment Classification and Quantification in Twitter
  Ming Wang, Biao Chu, Qingxun Liu, and Xiaobing Zhou
- LSIS at SemEval-2017 Task 4: Using Adapted Sentiment Similarity Seed Words For English and Arabic Tweet Polarity Classification
  Amal Htait, Sébastien Fournier, and Patrice Bellot
- ELiRF-UPV at SemEval-2017 Task 4: Sentiment Analysis using Deep Learning
  José-Ángel González, Ferran Pla, and Lluís-F. Hurtado
- XJSA at SemEval-2017 Task 4: A Deep System for Sentiment Classification in Twitter
  Yazhou Hao, YangYang Lan, Yufei Li, and Chen Li
- Adullam at SemEval-2017 Task 4: Sentiment Analyzer Using Lexicon Integrated Convolutional Neural Networks with Attention
  Joosung Yoon, Kigon Lyu, and Hyeoncheol Kim
- EICA at SemEval-2017 Task 4: A Simple Convolutional Neural Network for Topic-based Sentiment Classification
  Maoquan Wang, Shiyun Chen, Yufei Xie, and Zhao Lu
- funSentiment at SemEval-2017 Task 4: Topic-Based Message Sentiment Classification by Exploiting Word Embeddings, Text Features and Target Contexts
  Quanzhi Li, Armineh Nourbakhsh, Xiaomo Liu, Rui Fang, and Sameena Shah
- DataStories at SemEval-2017 Task 4: Deep LSTM with Attention for Message-level and Topic-based Sentiment Analysis
  Christos Baziotis, Nikos Pelekis, and Christos Doulkeridis
- TwiSe at SemEval-2017 Task 4: Five-point Twitter Sentiment Classification and Quantification
  Georgios Balikas
- LIA at SemEval-2017 Task 4: An Ensemble of Neural Networks for Sentiment Classification
  Mickael Rouvier
- TopicThunder at SemEval-2017 Task 4: Sentiment Classification Using a Convolutional Neural Network with Distant Supervision
  Simon Müller, Tobias Huonder, Jan Milan Deriu, and Mark Cieliebak
- INGEOTEC at SemEval 2017 Task 4: A B4MSA Ensemble based on Genetic Programming for Twitter Sentiment Analysis
  Sabino Miranda-Jiménez, Mario Graff, Eric Sadit Tellez, and Daniela Moctezuma
- BUSEM at SemEval-2017 Task 4A Sentiment Analysis with Word Embedding and Long Short Term Memory RNN Approaches
  Deger Ayata, Murat Saraclar, and Arzucan Ozgur
- TakeLab at SemEval-2017 Task 4: Recent Deaths and the Power of Nostalgia in Sentiment Analysis in Twitter
  David Lozić, Doria Šarić, Ivan Tokić, Zoran Medić, and Jan Šnajder

- NileTMRG at SemEval-2017 Task 4: Arabic Sentiment Analysis
  Samhaa R. El-Beltagy, Mona El kalamawy, and Abu Bakr Soliman
- YNU-HPCC at SemEval 2017 Task 4: Using A Multi-Channel CNN-LSTM Model for Sentiment Classification
  Haowei Zhang, Jin Wang, Jixian Zhang, and Xuejie Zhang
- TSA-INF at SemEval-2017 Task 4: An Ensemble of Deep Learning Architectures Including Lexicon Features for Twitter Sentiment Analysis
  Amit Ajit Deshmane and Jasper Friedrichs
- UCSC-NLP at SemEval-2017 Task 4: Sense n-grams for Sentiment Analysis in Twitter
  José Abreu, Iván Castro, Claudia Martínez, Sebastián Oliva, and Yoan Gutiérrez
- ECNU at SemEval-2017 Task 4: Evaluating Effective Features on Machine Learning Methods for Twitter Message Polarity Classification
  Yunxiao Zhou, Man Lan, and Yuanbin Wu
- Fortia-FBK at SemEval-2017 Task 5: Bullish or Bearish? Inferring Sentiment towards Brands from Financial News Headlines
  Youness Mansar, Lorenzo Gatti, Sira Ferradans, Marco Guerini, and Jacopo Staiano
- SSN_MLRG1 at SemEval-2017 Task 5: Fine-Grained Sentiment Analysis Using Multiple Kernel Gaussian Process Regression Model
  Angel Deborah S, S Milton Rajendram, and T T Mirnalinee
- IBA-Sys at SemEval-2017 Task 5: Fine-Grained Sentiment Analysis on Financial Microblogs and News
  Zarmeen Nasim
- HHU at SemEval-2017 Task 5: Fine-Grained Sentiment Analysis on Financial Data using Machine Learning Methods
  Tobias Cabanski, Julia Romberg, and Stefan Conrad
- INF-UFRGS at SemEval-2017 Task 5: A Supervised Identification of Sentiment Score in Tweets and Headlines
  Tiago Zini, Karin Becker, and Marcelo Dias
- HCS at SemEval-2017 Task 5: Polarity detection in business news using convolutional neural networks
  Lidia Pivovarova, Llorenç Escoter, Arto Klami, and Roman Yangarber
- NLG301 at SemEval-2017 Task 5: Fine-Grained Sentiment Analysis on Financial Microblogs and News
  Chung-Chi Chen, Hen-Hsen Huang, and Hsin-Hsi Chen
- funSentiment at SemEval-2017 Task 5: Fine-Grained Sentiment Analysis on Financial Microblogs Using Word Vectors Built from StockTwits and Twitter
  Quanzhi Li, Sameena Shah, Armineh Nourbakhsh, Rui Fang, and Xiaomo Liu
- SentiHeros at SemEval-2017 Task 5: An application of Sentiment Analysis on Financial Tweets
  Narges Tabari, Armin Seyeditabari, and Wlodek Zadrozny
- DUTH at SemEval-2017 Task 5: Sentiment Predictability in Financial Microblogging and News Articles
  Symeon Symeonidis, John Kordonis, Dimitrios Effrosynidis, and Avi Arampatzis
- TakeLab at SemEval-2017 Task 5: Linear aggregation of word embeddings for fine-grained sentiment analysis of financial news
  Leon Rotim, Martin Tutek, and Jan Šnajder
- UW-FinSent at SemEval-2017 Task 5: Sentiment Analysis on Financial News Headlines using Training Dataset Augmentation
  Vineet John and Olga Vechtomova

- RiTUAL-UH at SemEval-2017 Task 5: Sentiment Analysis on Financial Data Using Neural Networks
  Sudipta Kar, Suraj Maharjan, and Thamar Solorio
- COMMIT at SemEval-2017 Task 5: Ontology-based Method for Sentiment Analysis of Financial Headlines
  Kim Schouten, Flavius Frasincar, and Franciska de Jong
- ECNU at SemEval-2017 Task 5: An Ensemble of Regression Algorithms with Effective Features for Fine-Grained Sentiment Analysis in Financial Domain
  Mengxiao Jiang, Man Lan, and Yuanbin Wu
- IITPB at SemEval-2017 Task 5: Sentiment Prediction in Financial Text
  Abhishek Kumar, Abhishek Sethi, Md Shad Akhtar, Asif Ekbal, Chris Biemann, and Pushpak Bhattacharyya
- IITP at SemEval-2017 Task 5: An Ensemble of Deep Learning and Feature Based Models for Financial Sentiment Analysis
  Deepanway Ghosal, Shobhit Bhatnagar, Md Shad Akhtar, Asif Ekbal, and Pushpak Bhattacharyya
- FEUP at SemEval-2017 Task 5: Predicting Sentiment Polarity and Intensity with Financial Word Embeddings
  Pedro Saleiro, Eduarda Mendes Rodrigues, Carlos Soares, and Eugénio Oliveira
- UIT-DANGNT-CLNLP at SemEval-2017 Task 9: Building Scientific Concept Fixing Patterns for Improving CAMR
  Khoa Nguyen and Dang Nguyen
- Oxford at SemEval-2017 Task 9: Neural AMR Parsing with Pointer-Augmented Attention
  Jan Buys and Phil Blunsom
- FORGe at SemEval-2017 Task 9: Deep sentence generation based on a sequence of graph transducers
  Simon Mille, Roberto Carlini, Alicia Burga, and Leo Wanner
- RIGOTRIO at SemEval-2017 Task 9: Combining Machine Learning and Grammar Engineering for AMR Parsing and Generation
  Normunds Gruzitis, Didzis Gosko, and Guntis Barzdins
- The Meaning Factory at SemEval-2017 Task 9: Producing AMRs with Neural Semantic Parsing
  Rik van Noord and Johan Bos
- PKU_ICL at SemEval-2017 Task 10: Keyphrase Extraction with Model Ensemble and External Knowledge
  Liang Wang and Sujian Li
- NTNU-1ScienceIE at SemEval-2017 Task 10: Identifying and Labelling Keyphrases with Conditional Random Fields
  Erwin Marsi, Utpal Kumar Sikdar, Cristina Marco, Biswanath Barik, and Rune Sætre
- EELECTION at SemEval-2017 Task 10: Ensemble of nEural Learners for kEyphrase ClassificaTION
  Steffen Eger, Erik-Lân Do Dinh, Ilia Kuznetsov, Masoud Kiaeeha, and Iryna Gurevych
- LABDA at SemEval-2017 Task 10: Extracting Keyphrases from Scientific Publications by combining the BANNER tool and the UMLS Semantic Network
  Isabel Segura-Bedmar, Cristóbal Colón-Ruiz, and Paloma Martínez
- The NTNU System at SemEval-2017 Task 10: Extracting Keyphrases and Relations from Scientific Publications Using Multiple Conditional Random Fields
  Lung-Hao Lee, Kuei-Ching Lee, and Yuen-Hsien Tseng

- MayoNLP at SemEval 2017 Task 10: Word Embedding Distance Pattern for Keyphrase Classification in Scientific Publications
  Sijia Liu, Feichen Shen, Vipin Chaudhary, and Hongfang Liu
- Know-Center at SemEval-2017 Task 10: Sequence Classification with the CODE Annotator
  Roman Kern, Stefan Falk, and Andi Rexha
- NTNU-2 at SemEval-2017 Task 10: Identifying Synonym and Hyponym Relations among Keyphrases in Scientific Documents
  Biswanath Barik and Erwin Marsi
- LABDA at SemEval-2017 Task 10: Relation Classification between keyphrases via Convolutional Neural Network
  Víctor Suárez-Paniagua, Isabel Segura-Bedmar, and Paloma Martínez
- WING-NUS at SemEval-2017 Task 10: Keyphrase Extraction and Classification as Joint Sequence Labeling
  Animesh Prasad and Min-Yen Kan
- MIT at SemEval-2017 Task 10: Relation Extraction with Convolutional Neural Networks
  Ji Young Lee, Franck Dernoncourt, and Peter Szolovits
- TTI-COIN at SemEval-2017 Task 10: Investigating Embeddings for End-to-End Relation Extraction from Scientific Papers
  Tomoki Tsujimura, Makoto Miwa, and Yutaka Sasaki
- SZTE-NLP at SemEval-2017 Task 10: A High Precision Sequence Model for Keyphrase Extraction Utilizing Sparse Coding for Feature Generation
  Gábor Berend
- LIPN at SemEval-2017 Task 10: Filtering Candidate Keyphrases from Scientific Publications with Part-of-Speech Tag Sequences to Train a Sequence Labeling Model
  Simon David Hernandez, Davide Buscaldi, and Thierry Charnois
- EUDAMU at SemEval-2017 Task 11: Action Ranking and Type Matching for End-User Development
  Marek Kubis, Paweł Skórzewski, and Tomasz Ziętkiewicz
- Hitachi at SemEval-2017 Task 12: System for temporal information extraction from clinical notes
  Sarath P R, Manikandan R, and Yoshiki Niwa
- NTU-1 at SemEval-2017 Task 12: Detection and classification of temporal events in clinical data with domain adaptation
  Po-Yu Huang, Hen-Hsen Huang, Yu-Wun Wang, Ching Huang, and Hsin-Hsi Chen
- XJNLP at SemEval-2017 Task 12: Clinical temporal information ex-traction with a Hybrid Model
  Yu Long, Zhijing Li, Xuan Wang, and Chen Li
- ULISBOA at SemEval-2017 Task 12: Extraction and classification of temporal expressions and events
  Andre Lamurias, Diana Sousa, Sofia Pereira, Luka Clarke, and Francisco M Couto
- GUIR at SemEval-2017 Task 12: A Framework for Cross-Domain Clinical Temporal Information Extraction
  Sean MacAvaney, Arman Cohan, and Nazli Goharian
- KULeuven-LIIR at SemEval-2017 Task 12: Cross-Domain Temporal Information Extraction from Clinical Records
  Artuur Leeuwenberg and Marie-Francine Moens

BUCC: 10th Workshop on Building and Using Comparable Corpora

Organizers: *Serge Sharoff, Pierre Zweigenbaum, and Reinhard Rapp*
Venue: Cypress 1

Thursday, August 3, 2017

9:00-9:05 **Opening**

9:05-10:00 **Invited presentation**

- Users and Data: The Two Neglected Children of Bilingual Natural Language Processing Research
  Philippe Langlais

10:00-10:30 **Session 1: Plagiarism detection**

- Deep Investigation of Cross-Language Plagiarism Detection Methods
  Jérémy Ferrero, Laurent Besacier, Didier Schwab, and Frédéric Agnès

10:30-11:00 **Coffee break**

11:00-12:00 **Session 2: Sentence alignment and lexicon acquisition**

- Sentence Alignment using Unfolding Recursive Autoencoders
  Jeenu Grover and Pabitra Mitra
- Acquisition of Translation Lexicons for Historically Unwritten Languages via Bridging Loanwords
  Michael Bloodgood and Benjamin Strauss

12:00-2:00 **Lunch**

2:00-3:30 **Session 3: Building comparable corpora**

- Toward a Comparable Corpus of Latvian, Russian and English Tweets
  Dmitrijs Milajevs
- Automatic Extraction of Parallel Speech Corpora from Dubbed Movies
  Alp Öktem, Mireia Farrús, and Leo Wanner
- A parallel collection of clinical trials in Portuguese and English
  Mariana Neves

3:30-4:00 **Coffee break**

4:00-5:40 **Session 4: Shared task session**

- Weighted Set-Theoretic Alignment of Comparable Sentences
  Andoni Azpeitia, Thierry Etchegoyhen, and Eva Martínez Garcia
- BUCC 2017 Shared Task: a First Attempt Toward a Deep Learning Framework for Identifying Parallel Sentences in Comparable Corpora
  Francis Grégoire and Philippe Langlais
- zNLP: Identifying Parallel Sentences in Chinese-English Comparable Corpora
  Zheng Zhang and Pierre Zweigenbaum
- BUCC2017: A Hybrid Approach for Identifying Parallel Sentences in Comparable Corpora
  Sainik Mahata, Dipankar Das, and Sivaji Bandyopadhyay

- Overview of the Second BUCC Shared Task: Spotting Parallel Sentences in Comparable Corpora
  Pierre Zweigenbaum, Serge Sharoff, and Reinhard Rapp

5:40-5:50 **Closing**

CLPsych: Computational Linguistics and Clinical Psychology – From Linguistic Signal to Clinical Reality

Organizers: *Kristy Hollingshead, Molly E. Ireland, and Kate Loveys*

Venue: Cypress 2

Thursday, August 3, 2017

9:00–9:20 **Opening Remarks (Kristy Hollingshead, Molly E. Ireland, and Kate Loveys)**

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

- A Cross-modal Review of Indicators for Depression Detection Systems
  Michelle Morales, Stefan Scherer, and Rivka Levitan
- In your wildest dreams: the language and psychological features of dreams
  Kate Niederhoffer, Jonathan Schler, Patrick Crutchley, Kate Loveys, and Glen Coppersmith

11:00–12:15 **Session: Poster Presentations**

- A Corpus Analysis of Social Connections and Social Isolation in Adolescents Suffering from Depressive Disorders
  Jia-Wen Guo, Danielle L Mowery, Djin Lai, Katherine Sward, and Mike Conway
- Monitoring Tweets for Depression to Detect At-risk Users
  Zunaira Jamil, Diana Inkpen, Prasadith Buddhitha, and Kenton White
- Examining Sentiment and Depression in Survivors of Intimate Partner Violence
  Joseph Costello, Catherine Kothari, Duncan Vos, Richard Brandt, and Angie Moe
- Ethical Challenges in Algorithmic Inference of Mental Illness with Large-Scale Social Data
  Stevie Chancellor, Vincent Silenzio, Eric Caine, and Munmun De Choudhury
- Validation of Twitter Self-Stated Models of Mental Health against Patient Medical Records
  Glen Coppersmith, Patrick Crutchley, Raina M. Merchant, and H. Andrew Schwartz
- Language Style Matching in Subclinically Depressed and Anxious Participants' Responses to Social Media-Style Posts
  Taleen Nalabandian and Molly E. Ireland

13:45–14:30 **Session: Mini-Oral Presentations**

- Investigating Patient Attitudes Towards the use of Social Media Data to Augment Depression Diagnosis and Treatment: a Qualitative Study
  Jude Mikal, Samantha Hurst, and Mike Conway
- Natural-language Interactive Narratives in Imaginal Exposure Therapy for Obsessive-Compulsive Disorder
  Melissa Roemmele, Paola Mardo, and Andrew Gordon
- Detecting Anxiety through Reddit
  Judy Hanwen Shen and Frank Rudzicz

- Detecting and Explaining Crisis
  Rohan Kshirsagar, Robert Morris, and Samuel Bowman

14:30–15:30 **Session: Oral Presentations 2**

- A Dictionary-Based Comparison of Autobiographies by People and Murderous Monsters
  Micah Iserman and Molly E. Ireland
- Small but Mighty: Affective Micropatterns for Quantifying Mental Health from Social Media Language
  Kate Loveys, Patrick Crutchley, Emily Wyatt, and Glen Coppersmith

16:00–17:00 **CLPsych2017 Shared Task: Results & Open Discussion (David Milne)**

17:00–17:30 **Closing Remarks**

NLP+CSS: Workshops on Natural Language Processing and Computational Social Science

Organizers: *David Bamman, A. Seza Doğruöz, Dirk Hovy, David Jurgens, Brendan O'Connor, Oren Tsur, and Svitlana Volkova*

Venue: Salon 3

Thursday, August 3, 2017

9:00–10:30 **Session 1**

9:00–9:15 **Welcome (Organizers)**

9:15–10:00 **Invited Talk 1**

10:00–10:30 **Spotlight Paper Session**

10:00–10:15 Language-independent Gender Prediction on Twitter
Nikola Ljubešić, Darja Fišer, and Tomaž Erjavec

10:15–10:30 When does a compliment become sexist? Analysis and classification of ambivalent sexism using twitter data
Akshita Jha and Radhika Mamidi

10:30–11:00 **Morning coffee break**

11:00–12:15 **Session 2**

11:00–11:45 **Invited Talk: Measuring Psychological Traits using Social Media (Lyle Ungar)**

11:45–12:15 **Spotlight Paper Session**

11:45–12:00 Personality Driven Differences in Paraphrase Preference
Daniel Preoţiuc-Pietro, Jordan Carpenter, and Lyle Ungar

12:00–12:15 **Never Tell Me the Odds: How Belief Dynamics Shape Audience Experience (non-archival) (Shengli Hu)**

12:15–14:00 **Lunch break**

14:00–15:30 **Session 3**

14:00–14:45 **Invited Talk: The War on Facts (Gideon Mann)**

14:45–3:30 **One-minute-madness paper presentation**

14:45–14:48 community2vec: Vector representations of online communities encode semantic relationships
Trevor Martin

14:48–14:51 Telling Apart Tweets Associated with Controversial versus Non-Controversial Topics
Aseel Addawood, Rezvaneh Rezapour, Omid Abdar, and Jana Diesner

14:51–14:54 Cross-Lingual Classification of Topics in Political Texts
Goran Glavaš, Federico Nanni, and Simone Paolo Ponzetto

14:54–14:57 **The Role of Network Structure for Gender Prediction (non-archival) (Kristen M. Altenburger and Johan Ugander)**

14:57–15:00 **The Role of Party and Incumbency in Identification of Argumentation Strategies in Political Debate (non-archival) (Justin Garten, Kenji Sagae, Zahra Kamel, Nitika Awasthi and Morteza Dehghani)**

15:00–15:03 Mining Social Science Publications for Survey Variables
Andrea Zielinski and Peter Mutschke

15:03–15:06 Linguistic Markers of Influence in Informal Interactions
Shrimai Prabhumoye, Samridhi Choudhary, Evangelia Spiliopoulou, Christopher Bogart, Carolyn Rose, and Alan W Black

15:06–15:09 Non-lexical Features Encode Political Affiliation on Twitter
Rachael Tatman, Leo Stewart, Amandalynne Paullada, and Emma Spiro

15:09–15:12 **Syntactic Alignment in Power Relations (non-archival) (Reihane Boghrati, Justin Garten and Morteza Dehghani)**

15:12–15:15 Modelling Participation in Small Group Social Sequences with Markov Rewards Analysis
Gabriel Murray

15:15–15:18 Code-Switching as a Social Act: The Case of Arabic Wikipedia Talk Pages
Michael Yoder, Shruti Rijhwani, Carolyn Rosé, and Lori Levin

15:18–15:21 How Does Twitter User Behavior Vary Across Demographic Groups?
Zach Wood-Doughty, Michael Smith, David Broniatowski, and Mark Dredze

15:21–15:24 Ideological Phrase Indicators for Classification of Political Discourse Framing on Twitter
Kristen Marie Johnson, I-Ta Lee, and Dan Goldwasser

15:24–15:27 **Market Evolution of Sharing Economy vs. Traditional Platforms: A Natural Language Processing Approach (Mohsen Mosleh and Babak Heydari)**

15:30–16:45 **Coffee break and posters**

16:45–17:45 **Session 4**

16:45–17:30 **Invited Talk (Brandon Stewart)**

17:30–17:45 **Closing remarks and wrap-up (Organizers)**

Repl4NLP: 2nd Workshop on Representation Learning for NLP

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

Venue: Salon C

Thursday, August 3, 2017

9:30–9:45 **Welcome and Opening Remarks**

9:45–10:30 **Keynote Session**

9:45–10:30 Learning Joint Embeddings of Vision and Language (Sanja Fidler)

10:30–11:00 **Coffee Break**

11:00–12:30 **Keynote Session**

11:00–11:45 Learning Representations of Social Meaning (Jacob Eisenstein)

11:45–12:30 Representations in the Brain (Alona Fyshe)

12:30–14:00 **Lunch**

14:00–14:45 **Keynote Session**

14:00–14:45 "A million ways to say I love you" or Learning to Paraphrase with Neural Machine Translation (Mirella Lapata)

14:45–15:00 **Best Paper Session**

15:00–16:30 **Poster Session, including Coffee Break**

- Sense Contextualization in a Dependency-Based Compositional Distributional Model
  Pablo Gamallo
- Context encoders as a simple but powerful extension of word2vec
  Franziska Horn
- Active Discriminative Text Representation Learning
  Ye Zhang, Matthew Lease, and Byron Wallace
- Using millions of emoji occurrences to pretrain any-domain models for detecting emotion, sentiment and sarcasm
  Bjarke Felbo, Alan Mislove, Anders Søgaard, Iyad Rahwan, and Sune Lehmann
- Evaluating Layers of Representation in Neural Machine Translation on Syntactic and Semantic Tagging
  Yonatan Belinkov, Lluís Màrquez, Hassan Sajjad, Nadir Durrani, Fahim Dalvi, and James Glass
- Machine Comprehension by Text-to-Text Neural Question Generation
  Xingdi Yuan, Tong Wang, Caglar Gulcehre, Alessandro Sordoni, Philip Bachman, Saizheng Zhang, Sandeep Subramanian, and Adam Trischler
- Emergent Predication Structure in Hidden State Vectors of Neural Readers
  Hai Wang, Takeshi Onishi, Kevin Gimpel, and David McAllester

- Towards Harnessing Memory Networks for Coreference Resolution
  Joe Cheri and Pushpak Bhattacharyya
- Combining Word-Level and Character-Level Representations for Relation Classification of Informal Text
  Dongyun Liang, Weiran Xu, and Yinge Zhao
- Regularized Topic Models for Sparse Interpretable Word Embeddings
  Anna Potapenko and Artem Popov
- Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings
  Tolga Bolukbasi, Kai-Wei Chang, James Zou, Venkatesh Saligrama, and Adam T. Kalai
- Transfer Learning for Neural Semantic Parsing
  Xing Fan, Emilio Monti, Lambert Mathias, and Markus Dreyer
- MUSE: Modularizing Unsupervised Sense Embeddings
  Guang-He Lee and Yun-Nung Chen
- Modeling Large-Scale Structured Relationships with Shared Memory for Knowledge Base Completion
  Yelong Shen, Po-Sen Huang, Ming-Wei Chang, and Jianfeng Gao
- Knowledge Base Completion: Baselines Strike Back
  Rudolf Kadlec, Ondrej Bajgar, and Jan Kleindienst
- Sequential Attention: A Context-Aware Alignment Function for Machine Reading
  Sebastian Brarda, Philip Yeres, and Samuel Bowman
- Semantic Vector Encoding and Similarity Search Using Fulltext Search Engines
  Jan Rygl, Jan Pomikálek, Radim Řehůřek, Michal Růžička, Vít Novotný, and Petr Sojka
- Multi-task Domain Adaptation for Sequence Tagging
  Nanyun Peng and Mark Dredze
- Beyond Bilingual: Multi-sense Word Embeddings using Multilingual Context
  Shyam Upadhyay, Kai-Wei Chang, Matt Taddy, Adam Kalai, and James Zou
- DocTag2Vec: An Embedding Based Multi-label Learning Approach for Document Tagging
  Sheng Chen, Akshay Soni, Aasish Pappu, and Yashar Mehdad
- Binary Paragraph Vectors
  Karol Grzegorczyk and Marcin Kurdziel
- Representing Compositionality based on Multiple Timescales Gated Recurrent Neural Networks with Adaptive Temporal Hierarchy for Character-Level Language Models
  Dennis Singh Moirangthem, Jegyung Son, and Minho Lee
- Learning Bilingual Projections of Embeddings for Vocabulary Expansion in Machine Translation
  Pranava Swaroop Madhyastha and Cristina España-Bonet
- Learning to Compose Words into Sentences with Reinforcement Learning
  Dani Yogatama, Phil Blunsom, Chris Dyer, Edward Grefenstette, and Wang Ling
- Prediction of Frame-to-Frame Relations in the FrameNet Hierarchy with Frame Embeddings
  Teresa Botschen, Hatem Mousselly Sergieh, and Iryna Gurevych
- Learning Joint Multilingual Sentence Representations with Neural Machine Translation
  Holger Schwenk and Matthijs Douze

- Transfer Learning for Speech Recognition on a Budget
  Julius Kunze, Louis Kirsch, Ilia Kurenkov, Andreas Krug, Jens Johannsmeier, and Sebastian Stober
- Gradual Learning of Matrix-Space Models of Language for Sentiment Analysis
  Shima Asaadi and Sebastian Rudolph
- Improving Language Modeling using Densely Connected Recurrent Neural Networks
  Fréderic Godin, Joni Dambre, and Wesley De Neve
- NewsQA: A Machine Comprehension Dataset
  Adam Trischler, Tong Wang, Xingdi Yuan, Justin Harris, Alessandro Sordoni, Philip Bachman, and Kaheer Suleman
- Intrinsic and Extrinsic Evaluation of Spatiotemporal Text Representations in Twitter Streams
  Lawrence Phillips, Kyle Shaffer, Dustin Arendt, Nathan Hodas, and Svitlana Volkova
- Rethinking Skip-thought: A Neighborhood based Approach
  Shuai Tang, Hailin Jin, Chen Fang, Zhaowen Wang, and Virginia de Sa
- A Frame Tracking Model for Memory-Enhanced Dialogue Systems
  Hannes Schulz, Jeremie Zumer, Layla El Asri, and Shikhar Sharma
- Plan, Attend, Generate: Character-Level Neural Machine Translation with Planning
  Caglar Gulcehre, Francis Dutil, Adam Trischler, and Yoshua Bengio
- Does the Geometry of Word Embeddings Help Document Classification? A Case Study on Persistent Homology-Based Representations
  Paul Michel, Abhilasha Ravichander, and Shruti Rijhwani
- Adversarial Generation of Natural Language
  Sandeep Subramanian, Sai Rajeswar, Francis Dutil, Chris Pal, and Aaron Courville
- Deep Active Learning for Named Entity Recognition
  Yanyao Shen, Hyokun Yun, Zachary Lipton, Yakov Kronrod, and Animashree Anandkumar
- The Coadaptation Problem when Learning How and What to Compose
  Andrew Drozdov and Samuel Bowman
- Learning when to skim and when to read
  Alexander Johansen and Richard Socher
- Learning to Embed Words in Context for Syntactic Tasks
  Lifu Tu, Kevin Gimpel, and Karen Livescu

16:30–17:30 **Panel Discussion**

17:30–17:40 **Closing Remarks**

RoboNLP: Language Grounding for Robotics

Organizers: *Mohit Bansal, Cynthia Matuszek, Jacob Andreas, Yoav Artzi, and Yonatan Bisk*
Venue: Salon 2

Thursday, August 3, 2017

9:00–9:15 **Welcome and Opening Remarks**

9:15–9:50 **Invited Talk 1 (Joyce Chai, MSU)**

9:50–10:25 **Invited Talk 2 (Ray Mooney, UT Austin)**

10:30–11:00 **Coffee break**

11:00–11:35 **Invited Talk 3 (Stefanie Tellax)**

11:35–12:10 **Invited Talk 4 (Karl Moritz Hermann, Google DeepMind)**

12:10–14:00 **Poster Session (Lunch from 12:30–14:00)**

- Grounding Language for Interactive Task Learning
  Peter Lindes, Aaron Mininger, James R. Kirk, and John E. Laird
- Guiding Interaction Behaviors for Multi-modal Grounded Language Learning
  Jesse Thomason, Jivko Sinapov, and Raymond Mooney
- Structured Learning for Context-aware Spoken Language Understanding of Robotic Commands
  Andrea Vanzo, Danilo Croce, Roberto Basili, and Daniele Nardi
- Communication with Robots using Multilayer Recurrent Networks
  Bedřich Pišl and David Mareček
- Grounding Symbols in Multi-Modal Instructions
  Yordan Hristov, Svetlin Penkov, Alex Lascarides, and Subramanian Ramamoorthy
- Exploring Variation of Natural Human Commands to a Robot in a Collaborative Navigation Task
  Matthew Marge, Claire Bonial, Ashley Foots, Cory Hayes, Cassidy Henry, Kimberly Pollard, Ron Artstein, Clare Voss, and David Traum
- Are Distributional Representations Ready for the Real World? Evaluating Word Vectors for Grounded Perceptual Meaning
  Li Lucy and Jon Gauthier
- Sympathy Begins with a Smile, Intelligence Begins with a Word: Use of Multimodal Features in Spoken Human-Robot Interaction
  Jekaterina Novikova, Christian Dondrup, Ioannis Papaioannou, and Oliver Lemon
- Towards Problem Solving Agents that Communicate and Learn
  Anjali Narayan-Chen, Colin Graber, Mayukh Das, Md Rakibul Islam, Soham Dan, Sriraam Natarajan, Janardhan Rao Doppa, Julia Hockenmaier, Martha Palmer, and Dan Roth

14:00–14:35 **Invited Talk 5 (Percy Liang, Stanford)**

14:45–15:10 **Invited Talk 6 (Jason Weston, Facebook AI Research)**

Selected Oral Submission 1

15:10–15:20 Learning how to Learn: An Adaptive Dialogue Agent for Incrementally Learning Visually Grounded Word Meanings
Yanchao Yu, Arash Eshghi, and Oliver Lemon

Selected Oral Submission 2

15:20–15:30 Natural Language Grounding and Grammar Induction for Robotic Manipulation Commands
Muhannad Alomari, Paul Duckworth, Majd Hawasly, David C. Hogg, and Anthony G. Cohn

15:30–16:00 **Coffee break**

Selected Oral Submission 3

16:00–16:10 A Tale of Two DRAGGNs: A Hybrid Approach for Interpreting Action-Oriented and Goal-Oriented Instructions
Siddharth Karamcheti, Edward Clem Williams, Dilip Arumugam, Mina Rhee, Nakul Gopalan, Lawson L.S. Wong, and Stefanie Tellex

16:10–16:45 **Invited Talk 7 (Nicholas Roy, MIT)**

16:45–17:45 **Panel**

TextGraphs-11: Graph-based Methods for Natural Language Processing

Organizers: *Martin Riedl, Swapna Somasundaran, Goran Glavaš, and Eduard Hovy*
Venue: Mackenzie

Thursday, August 3, 2017

9:00–9:10 **Opening remarks (Swapna Somasundaran and Goran Glavaš)**

9:10–10:10 **Invited talk (Apoorv Agarwal)**

10:10–10:30 On the "Calligraphy" of Books
Vanessa Queiroz Marinho, Henrique Ferraz de Arruda, Thales Sinelli, Luciano da Fontoura Costa, and Diego Raphael Amancio

10:30–11:00 **Coffee break**

11:00–11:20 Adapting predominant and novel sense discovery algorithms for identifying corpus-specific sense differences
Binny Mathew, Suman Kalyan Maity, Pratip Sarkar, Animesh Mukherjee, and Pawan Goyal

11:20–11:40 Merging knowledge bases in different languages
Jerónimo Hernández-González, Estevam R. Hruschka Jr., and Tom M. Mitchell

11:40–12:00 Parameter Free Hierarchical Graph-Based Clustering for Analyzing Continuous Word Embeddings
Thomas Alexander Trost and Dietrich Klakow

12:00–12:15 Spectral Graph-Based Method of Multimodal Word Embedding
Kazuki Fukui, Takamasa Oshikiri, and Hidetoshi Shimodaira

12:15–14:00 **Lunch**

14:00–15:00 **Invited talk (Michael Strube)**

15:00–15:15 Graph Methods for Multilingual FrameNets
Collin Baker and Michael Ellsworth

15:15–15:30 Extract with Order for Coherent Multi-Document Summarization
Mir Tafseer Nayeem and Yllias Chali

15:30–16:00 **Coffee break**

16:00–16:20 Work Hard, Play Hard: Email Classification on the Avocado and Enron Corpora
Sakhar Alkhereyf and Owen Rambow

16:20–16:40 A Graph Based Semi-Supervised Approach for Analysis of Derivational Nouns in Sanskrit
Amrith Krishna, Pavankumar Satuluri, Harshavardhan Ponnada, Muneeb Ahmed, Gulab Arora, Kaustubh Hiware, and Pawan Goyal

16:40–17:00 Evaluating text coherence based on semantic similarity graph
Jan Wira Gotama Putra and Takenobu Tokunaga

17:00–17:10 **Best paper award and closing remarks (Swapna Somasundaran and Goran Glavaš)**

ALW1: 1st Workshop on Abusive Language Online

Organizers: *Zeerak Waseem, Wendy Hui Kyong Chun, Dirk Hovy, and Joel Tetreault*
Venue: Mackenzie

Friday, August 4, 2017

8:45–9:05 **Opening Remarks**

9:05–9:50 **Invited Talk A: Carol Todd**

9:50–10:35 **Panel A: Sora Han, Liz Losh, Lucas Dixon**

10:35–11:00 **Break**

11:00–12:30 **Paper Presentations**

11:00–11:20 Dimensions of Abusive Language on Twitter
Isobelle Clarke and Dr. Jack Grieve

11:20–11:40 Constructive Language in News Comments
Varada Kolhatkar and Maite Taboada

11:40–12:00 Rephrasing Profanity in Chinese Text
Hui-Po Su, Zhen-Jie Huang, Hao-Tsung Chang, and Chuan-Jie Lin

12:00–12:20 Deep Learning for User Comment Moderation
John Pavlopoulos, Prodromos Malakasiotis, and Ion Androutsopoulos

12:20–14:00 **Lunch**

14:00–15:30 **Poster Session**

- Class-based Prediction Errors to Detect Hate Speech with Out-of-vocabulary Words
Joan Serrà, Ilias Leontiadis, Dimitris Spathis, Gianluca Stringhini, Jeremy Blackburn, and Athena Vakali
- One-step and Two-step Classification for Abusive Language Detection on Twitter
Ji Ho Park and Pascale Fung
- Legal Framework, Dataset and Annotation Schema for Socially Unacceptable Online Discourse Practices in Slovene
Darja Fišer, Tomaž Erjavec, and Nikola Ljubešić
- Abusive Language Detection on Arabic Social Media
Hamdy Mubarak, Kareem Darwish, and Walid Magdy
- Vectors for Counterspeech on Twitter
Lucas Wright, Derek Ruths, Kelly P Dillon, Haji Mohammad Saleem, and Susan Benesch
- Detecting Nastiness in Social Media
Niloofar Safi Samghabadi, Suraj Maharjan, Alan Sprague, Raquel Diaz-Sprague, and Thamar Solorio
- Technology Solutions to Combat Online Harassment
George Kennedy, Andrew McCollough, Edward Dixon, Alexei Bastidas, John Ryan, Chris Loo, and Saurav Sahay
- Understanding Abuse: A Typology of Abusive Language Detection Subtasks
Zeerak Waseem, Thomas Davidson, Dana Warmsley, and Ingmar Weber

- Using Convolutional Neural Networks to Classify Hate-Speech
  Björn Gambäck and Utpal Kumar Sikdar
- Illegal is not a Noun: Linguistic Form for Detection of Pejorative Nominalizations
  Alexis Palmer, Melissa Robinson, and Kristy K. Phillips

15:30–16:00 **Break**

16:00–16:45 **Invited Talk B: Brianna Wu**

16:45–17:30 **Panel B: Pascale Fung, Vinodkumar Prabhakaran, Jacqueline Wernimont, Margaret Mitchell**

17:30–17:40 **Closing Remarks**

BioNLP: Workshop on Biomedical Natural Language Processing

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

Venue: Salon 3

Friday, August 4, 2017

8:30–8:45 **Opening remarks**

8:45–10:30 **Session 1: Prediction and relation extraction**

8:45–9:00 Target word prediction and paraphasia classification in spoken discourse
Joel Adams, Steven Bedrick, Gerasimos Fergadiotis, Kyle Gorman, and Jan van Santen

9:00–9:15 Extracting Drug-Drug Interactions with Attention CNNs
Masaki Asada, Makoto Miwa, and Yutaka Sasaki

9:15–9:30 Insights into Analogy Completion from the Biomedical Domain
Denis Newman-Griffis, Albert Lai, and Eric Fosler-Lussier

9:30–9:45 Deep learning for extracting protein-protein interactions from biomedical literature
Yifan Peng and Zhiyong Lu

9:45–10:00 Stacking With Auxiliary Features for Entity Linking in the Medical Domain
Nazneen Fatema Rajani, Mihaela Bornea, and Ken Barker

10:00–10:30 **Invited Talk: "Results of the 5th edition of BioASQ Challenge" – Georgios Paliouras**

- Results of the fifth edition of the BioASQ Challenge
Anastasios Nentidis, Konstantinos Bougiatiotis, Anastasia Krithara, Georgios Paliouras, and Ioannis Kakadiaris

10:30–11:00 **Coffee Break**

11:00–12:30 **Session 2: BioASQ 2017 and more**

11:00–11:15 Tackling Biomedical Text Summarization: OAQA at BioASQ 5B
Khyathi Chandu, Aakanksha Naik, Aditya Chandrasekar, Zi Yang, Niloy Gupta, and Eric Nyberg

11:15–11:30 Macquarie University at BioASQ 5b – Query-based Summarisation Techniques for Selecting the Ideal Answers
Diego Molla

11:30–11:45 Neural Question Answering at BioASQ 5B
Georg Wiese, Dirk Weissenborn, and Mariana Neves

11:45–12:00 End-to-End System for Bacteria Habitat Extraction
Farrokh Mehryary, Kai Hakala, Suwisa Kaewphan, Jari Björne, Tapio Salakoski, and Filip Ginter

12:00–12:15 Creation and evaluation of a dictionary-based tagger for virus species and proteins
Helen Cook, Rudolfs Berzins, Cristina Leal Rodrıguez, Juan Miguel Cejuela, and Lars Juhl Jensen

12:15–12:30 Representation of complex terms in a vector space structured by an ontology for a normalization task
Arnaud Ferré, Pierre Zweigenbaum, and Claire Nédellec

12:30–14:00 **Lunch break**

14:00–15:30 **Session 3: From bio to clinical NLP**

14:00–14:15 Improving Correlation with Human Judgments by Integrating Semantic Similarity with Second–Order Vectors
Bridget McInnes and Ted Pedersen

14:15–14:30 Proactive Learning for Named Entity Recognition
Maolin Li, Nhung Nguyen, and Sophia Ananiadou

14:30–14:45 Biomedical Event Extraction using Abstract Meaning Representation
Sudha Rao, Daniel Marcu, Kevin Knight, and Hal Daumé III

14:45–15:00 Detecting Personal Medication Intake in Twitter: An Annotated Corpus and Baseline Classification System
Ari Klein, Abeed Sarker, Masoud Rouhizadeh, Karen O'Connor, and Graciela Gonzalez

15:00–15:15 Unsupervised Context-Sensitive Spelling Correction of Clinical Free-Text with Word and Character N-Gram Embeddings
Pieter Fivez, Simon Suster, and Walter Daelemans

15:15–15:30 Characterization of Divergence in Impaired Speech of ALS Patients
Archna Bhatia, Bonnie Dorr, Kristy Hollingshead, Samuel L. Phillips, and Barbara McKenzie

15:30–16:00 **Coffee Break**

16:00–16:30 **Session 4 More clinical NLP**

16:00–16:15 Deep Learning for Punctuation Restoration in Medical Reports
Wael Salloum, Greg Finley, Erik Edwards, Mark Miller, and David Suendermann-Oeft

16:15–16:30 Unsupervised Domain Adaptation for Clinical Negation Detection
Timothy Miller, Steven Bethard, Hadi Amiri, and Guergana Savova

16:30–18:00 **Poster Session**

- BioCreative VI Precision Medicine Track: creating a training corpus for mining protein-protein interactions affected by mutations
Rezarta Islamaj Dogan, Andrew Chatr-aryamontri, Sun Kim, Chih-Hsuan Wei, Yifan Peng, Donald Comeau, and Zhiyong Lu
- Painless Relation Extraction with Kindred
Jake Lever and Steven Jones
- Noise Reduction Methods for Distantly Supervised Biomedical Relation Extraction
Gang Li, Cathy Wu, and K. Vijay-Shanker
- Role-Preserving Redaction of Medical Records to Enable Ontology-Driven Processing
Seth Polsley, Atif Tahir, Muppala Raju, Akintayo Akinleye, and Duane Steward
- Annotation of pain and anesthesia events for surgery-related processes and outcomes extraction
Wen-wai Yim, Dario Tedesco, Catherine Curtin, and Tina Hernandez-Boussard
- Identifying Comparative Structures in Biomedical Text
Samir Gupta, A.S.M. Ashique Mahmood, Karen Ross, Cathy Wu, and K. Vijay-Shanker

- Tagging Funding Agencies and Grants in Scientific Articles using Sequential Learning Models
  Subhradeep Kayal, Zubair Afzal, George Tsatsaronis, Sophia Katrenko, Pascal Coupet, Marius Doornenbal, and Michelle Gregory
- Deep Learning for Biomedical Information Retrieval: Learning Textual Relevance from Click Logs
  Sunil Mohan, Nicolas Fiorini, Sun Kim, and Zhiyong Lu
- Detecting Dementia through Retrospective Analysis of Routine Blog Posts by Bloggers with Dementia
  Vaden Masrani, Gabriel Murray, Thalia Field, and Giuseppe Carenini
- Protein Word Detection using Text Segmentation Techniques
  Devi Ganesan, Ashish V. Tendulkar, and Sutanu Chakraborti
- External Evaluation of Event Extraction Classifiers for Automatic Pathway Curation: An extended study of the mTOR pathway
  Wojciech Kusa and Michael Spranger
- Toward Automated Early Sepsis Alerting: Identifying Infection Patients from Nursing Notes
  Emilia Apostolova and Tom Velez
- Enhancing Automatic ICD-9-CM Code Assignment for Medical Texts with PubMed
  Danchen Zhang, Daqing He, Sanqiang Zhao, and Lei Li
- Evaluating Feature Extraction Methods for Knowledge-based Biomedical Word Sense Disambiguation
  Sam Henry, Clint Cuffy, and Bridget McInnes
- Investigating the Documentation of Electronic Cigarette Use in the Veteran Affairs Electronic Health Record: A Pilot Study
  Danielle Mowery, Brett South, Olga Patterson, Shu-Hong Zhu, and Mike Conway
- Automated Preamble Detection in Dictated Medical Reports
  Wael Salloum, Greg Finley, Erik Edwards, Mark Miller, and David Suendermann-Oeft
- A Biomedical Question Answering System in BioASQ 2017
  Mourad Sarrouti and Said Ouatik El Alaoui
- Adapting Pre-trained Word Embeddings For Use In Medical Coding
  Kevin Patel, Divya Patel, Mansi Golakiya, Pushpak Bhattacharyya, and Nilesh Birari
- Initializing neural networks for hierarchical multi-label text classification
  Simon Baker and Anna Korhonen
- Biomedical Event Trigger Identification Using Bidirectional Recurrent Neural Network Based Models
  Rahul V S S Patchigolla, Sunil Sahu, and Ashish Anand
- Representations of Time Expressions for Temporal Relation Extraction with Convolutional Neural Networks
  Chen Lin, Timothy Miller, Dmitriy Dligach, Steven Bethard, and Guergana Savova
- Automatic Diagnosis Coding of Radiology Reports: A Comparison of Deep Learning and Conventional Classification Methods
  Sarvnaz Karimi, Xiang Dai, Hamedh Hassanzadeh, and Anthony Nguyen
- Automatic classification of doctor-patient questions for a virtual patient record query task
  Leonardo Campillos Llanos, Sophie Rosset, and Pierre Zweigenbaum
- Assessing the performance of Olelo, a real-time biomedical question answering application
  Mariana Neves, Fabian Eckert, Hendrik Folkerts, and Matthias Uflacker

- Clinical Event Detection with Hybrid Neural Architecture
  Adyasha Maharana and Meliha Yetisgen
- Extracting Personal Medical Events for User Timeline Construction using Minimal Supervision
  Aakanksha Naik, Christopher Bogart, and Carolyn Rose
- Detecting mentions of pain and acute confusion in Finnish clinical text
  Hans Moen, Kai Hakala, Farrokh Mehryary, Laura-Maria Peltonen, Tapio Salakoski, Filip Ginter, and Sanna Salanterä
- A Multi-strategy Query Processing Approach for Biomedical Question Answering: USTB_PRIR at BioASQ 2017 Task 5B
  Zan-Xia Jin, Bo-Wen Zhang, Fan Fang, Le-Le Zhang, and Xu-Cheng Yin

EventStory: Events and Stories in the News

Organizers: *Tommaso Caselli, Ben Miller, Marieke van Erp, Piek Vossen, Martha Palmer, Eduard Hovy, Teruko Mitamura, and David Caswell*

Venue: Cypress 1

Friday, August 4, 2017

9:00–10:30 **Session 1:**

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

9:05–10:05 **A theory of events unifying semantic parsing and reasoning (James F. Allen, University of Rocherster)**

10:05–10:30 newsLens: building and visualizing long-ranging news stories
Philippe Laban and Marti Hearst

10:30–11:00 **Coffee Break**

11:00–12:30 **Session 2:**

11:00–12:30 **Annotation Exercise**

12:30–14:00 **Lunch**

14:00–14:45 **Session 3:**

14:00–14:05 Detecting Changes in Twitter Streams using Temporal Clusters of Hashtags
Yunli Wang and Cyril Goutte

14:05–14:10 Event Detection Using Frame-Semantic Parser
Evangelia Spiliopoulou, Eduard Hovy, and Teruko Mitamura

14:10–14:15 Improving Shared Argument Identification in Japanese Event Knowledge Acquisition
Yin Jou Huang and Sadao Kurohashi

14:15–14:20 Tracing armed conflicts with diachronic word embedding models
Andrey Kutuzov, Erik Velldal, and Lilja Øvrelid

14:20–14:25 The Circumstantial Event Ontology (CEO)
Roxane Segers, Tommaso Caselli, and Piek Vossen

14:25–24:30 Event Detection and Semantic Storytelling: Generating a Travelogue from a large Collection of Personal Letters
Georg Rehm, Julian Moreno Schneider, Peter Bourgonje, Ankit Srivastava, Jan Nehring, Armin Berger, Luca König, Sören Räuchle, and Jens Gerth

14:30–14:35 Inference of Fine-Grained Event Causality from Blogs and Films
Zhichao Hu, Elahe Rahimtoroghi, and Marilyn Walker

14:35–14:40 On the Creation of a Security-Related Event Corpus
Martin Atkinson, Jakub Piskorski, Hristo Tanev, and Vanni Zavarella

14:40–14:45 Inducing Event Types and Roles in Reverse: Using Function to Discover Theme
Natalie Ahn

14:45–16:00 **Poster session**

15:30–16:00 **Coffee Break**

16:00–17:45 **Session 4:**

16:00–16:25 The Event StoryLine Corpus: A New Benchmark for Causal and Temporal Relation Extraction
Tommaso Caselli and Piek Vossen

16:25–16:50 The Rich Event Ontology
Susan Brown, Claire Bonial, Leo Obrst, and Martha Palmer

16:50–17:15 Integrating Decompositional Event Structures into Storylines
William Croft, Pavlina Peskova, and Michael Regan

17:15–17:45 **Discussion and Conclusion**

LaTeCH-CLfL: Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature

Organizers: *Beatrice Alex, Stefania Degaetano-Ortlieb, Anna Feldman, Anna Kazantseva, Nils Reiter, and Stan Szpakowicz*

Venue: Cypress 2

Friday, August 4, 2017

9:00–10:00 **Session 1**

9:00–9:30 Metaphor Detection in a Poetry Corpus
Vaibhav Kesarwani, Diana Inkpen, Stan Szpakowicz, and Chris Tanasescu

9:30–10:00 Machine Translation and Automated Analysis of the Sumerian Language
Émilie Pagé-Perron, Maria Sukhareva, Ilya Khait, and Christian Chiarcos

10:00–10:30 **Poster Teaser Talks**

11:00–12:30 **Session 2**

11:00–11:30 Investigating the Relationship between Literary Genres and Emotional Plot Development
Evgeny Kim, Sebastian Padó, and Roman Klinger

11:30–12:00 Enjambment Detection in a Large Diachronic Corpus of Spanish Sonnets
Pablo Ruiz, Clara Martínez Cantón, Thierry Poibeau, and Elena González-Blanco

12:00–12:30 Plotting Markson's "Mistress"
Conor Kelleher and Mark Keane

13:30–14:00 **SIGHUM Business Meeting**

14:00–15:00 **Invited Talk**

14:00–15:00 **Characterization (Andrew Piper)**

3:00–4:00 **Poster Session**

- An End-to-end Environment for Research Question-Driven Entity Extraction and Network Analysis
  Andre Blessing, Nora Echelmeyer, Markus John, and Nils Reiter
- An Ontology-Based Method for Extracting and Classifying Domain-Specific Compositional Nominal Compounds
  Maria Pia di Buono
- Modeling intra-textual variation with entropy and surprisal: topical vs. stylistic patterns
  Stefania Degaetano-Ortlieb and Elke Teich
- Finding a Character's Voice: Stylome Classification on Literary Characters
  Liviu P. Dinu and Ana Sabina Uban
- Phonological Soundscapes in Medieval Poetry
  Christopher Hench
- Annotation Challenges for Reconstructing the Structural Elaboration of Middle Low German
  Nina Seemann, Marie-Luis Merten, Michaela Geierhos, Doris Tophinke, and Eyke Hüllermeier

- Speeding up corpus development for linguistic research: language documentation and acquisition in Romansh Tuatschin
  Géraldine Walther and Benoît Sagot

16:00–17:30 **Session 4**

16:00–16:30 Distantly Supervised POS Tagging of Low-Resource Languages under Extreme Data Sparsity: The Case of Hittite
Maria Sukhareva, Francesco Fuscagni, Johannes Daxenberger, Susanne Görke, Doris Prechel, and Iryna Gurevych

16:30–17:00 A Dataset for Sanskrit Word Segmentation
Amrith Krishna, Pavankumar Satuluri, and Pawan Goyal

17:00–17:30 Lexical Correction of Polish Twitter Political Data
Maciej Ogrodniczuk and Mateusz Kopeć

NMT: 1st Workshop on Neural Machine Translation

Organizers: *Thang Luong, Graham Neubig, Alexandra Birch, and Andrew Finch*
Venue: Salon C

Friday, August 4, 2017

Session 1

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

9:40–10:30 **Keynote (Chris Dyer)**

10:30–11:00 **Coffee Break**

Session 2

11:00–11:50 **Keynote (Alexander Rush)**

11:50–12:20 **Best Paper Session**

12:20–13:40 **Lunch Break**

Session 3

13:40–14:30 **Keynote (Kevin Knight)**

14:30-15:20 **Keynote (Quoc Le)**

Session 4

15:20–15:30 **Poster Session**

- An Empirical Study of Adequate Vision Span for Attention-Based Neural Machine Translation
  Raphael Shu and Hideki Nakayama
- Analyzing Neural MT Search and Model Performance
  Jan Niehues, Eunah Cho, Thanh-Le Ha, and Alex Waibel
- Stronger Baselines for Trustable Results in Neural Machine Translation
  Michael Denkowski and Graham Neubig
- Six Challenges for Neural Machine Translation
  Philipp Koehn and Rebecca Knowles
- Cost Weighting for Neural Machine Translation Domain Adaptation
  Boxing Chen, Colin Cherry, George Foster, and Samuel Larkin
- Detecting Untranslated Content for Neural Machine Translation
  Isao Goto and Hideki Tanaka
- Beam Search Strategies for Neural Machine Translation
  Markus Freitag and Yaser Al-Onaizan
- Interactive Beam Search for Visualizing Neural Machine Translation (extended abstract)
  Jaesong Lee, JoongHwi Shin, and Jun-Seok Kim
- Graph Convolutional Encoders for Syntax-aware Neural Machine Translation
  Joost Bastings, Ivan Titov, Wilker Aziz, Diego Marcheggiani, and Khalil Sima'an

- An Empirical Study of Mini-Batch Creation Strategies for Neural Machine Translation
  Makoto Morishita, Yusuke Oda, Graham Neubig, Koichiro Yoshino, Katsuhito Sudoh, and Satoshi Nakamura
- Domain Aware Neural Dialogue System
  Sajal Choudhary, Prerna Srivastava, Joao Sedoc, and Lyle Ungar
- Detecting Cross-Lingual Semantic Divergence for Neural Machine Translation
  Marine Carpuat, Yogarshi Vyas, and Xing Niu

15:30–16:10 **Poster Session (continued) and Coffee Break**

Session 5

16:10–17:30 **Panel Discussion**

17:30–17:40 **Closing Remarks**

7 Anti-harassment policy

The open exchange of ideas, the freedom of thought and expression, and respectful scientific debate are central to the aims and goals of a ACL 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, ACL 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 ACL 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 ACL 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 ACL 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 ACL board members are listed at:

`https://www.aclweb.org/website/about`

The full policy and its implementation is defined at:

`https://www.aclweb.org/adminwiki/index.php?title=Anti-Harassment_Policy`

8 Local Guide

This guide was written by Anoop Sarkar, with additions by Christian Federmann.

Vancouver, BC has several events scheduled during the summer time and we would like to highlight some of the events that will be happening either during ACL 2017 or just before or just after the conference. This is also an important reminder to those planning to attend the conference that you should book your hotel room immediately. It is very likely that there will be a shortage of hotel rooms due to the various festivals in the city at around the same time as ACL 2017.

The main festivals happening in the city during ACL 2017 are:

Vancouver Pride Festival The Vancouver Pride Festivalthis year will held in downtown Vancouver on Sunday, August 6, 2017. There are more than 20 official events aimed at bringing together the LGBTQ community and their friends, allies and supporters. The main event is the festival parade which will be held in downtown Vancouver within walking distance of the conference hotel. There is more information at the Tourism Vancouver site.

- `http://www.vancouverpride.ca/`
- `https://www.tourismvancouver.com/events/festivals-and-events/vancouver-pride-festival/`

Celebration of Light The Celebration of Light is a competition in which three countries compete to put on the best fireworks display. The fireworks start at 10pm over English Bay in Downtown Vancouver. This year the three countries competing are: Japan on Saturday, July 29th; the UK on Wednesday, August 2nd; and Canada on Saturday, August 5, 2017. Tickets are on sale for venues close to the fireworks display. It is not a long walk to English Bay from the conference hotel and you might be able to catch the fireworks without purchasing a ticket. There is more information at the Tourism Vancouver site.

- `http://hondacelebrationoflight.com/`
- `https://www.tourismvancouver.com/events/festivals-and-events/celebration-of-light/`

Caribbean Days Caribbean Days is a multicultural festival celebrating Caribbean culture which will be held this year in North Vancouver, BC on July 29 and 30, 2017. North Vancouver is a short ferry ride or car/bus ride away from downtown Vancouver. There is more information at the Tourism Vancouver site.

- http://www.caribbeandays.ca/
- https://www.tourismvancouver.com/events/festivals-and-events/10-unique-vancouver-festivals/

Vancouver Folk Music Festival This event is a bit before the conference dates but is a large event and you might want to catch some of it before the conference begins. The Vancouver Folk Music Festival will be held at Jericho Beach in Vancouver, BC on July 14-16, 2017. There is more information at the Tourism Vancouver site.

- http://thefestival.bc.ca/
- https://www.tourismvancouver.com/events/festivals-and-events/vancouver-folk-music-festivals/

Bard on the Beach: Shakespeare Festival The Bard on the Beach stages Shakespeare in Vanier Park, Vancouver, BC during the late summer. Plays are held between June 1 to September 23, 2017. More information.

- https://bardonthebeach.org/
- https://www.tourismvancouver.com/events/festivals-and-events/bard-on-the-beach-shakespeare-festival/

Theatre under the Stars Theatre under the Stars stages Broadway musicals in Stanley Park, Vancouver, BC. Performances are held between July 7 and August 19, 2017. More information.

- https://www.tuts.ca/
- https://www.tourismvancouver.com/events/festivals-and-events/theatre-under-the-stars/

Transportation The Downtown Vancouver area is pretty well connected with buses and taxis. The city has recently introduced more bike lanes, so biking is also an option to get around. Note that there are no ride sharing solutions in Vancouver as they are illegal in BC.

- http://vancouver.ca/streets-transportation/visitors-guide-to-public-transit.aspx
- https://www.translink.ca/

Safety Vancouver is a pretty safe city. Parts of East Hastings Street are home to Vancouver's drug scene and can be unsettling at first sight. During day time, there should not be any problems though. In general, you should be aware of your surroundings in the evening, as you would be in any big city. Vancouver's West End and the False Creek areas are relaxed.

9 On-Site Childcare and Nursing Room

For the first time ever, ACL will be offering on-site childcare at the conference hotel by advanced reservation or walk-in. Walk-ins are subject to availability, but there should be plenty of space, so please bring your kids! The cost of the childcare is partially subsidized by the ACL. The cost for general registrants will be $10 USD per hour per child. For students registrants the cost will be $5 USD per hour per child.

The childcare rooms are Oak 1 and 2 Rooms, located across from The Stanley Park Ballroom (Salons 1-3). The care is available for children 6 months to 12 years old from 8 a.m. to 6 p.m. during the main conference, and 9 a.m. to 5 p.m. on the tutorials and workshop days. The carer givers are from a professional childcare company, KiddieCorp, all of whom are CPR/first aid certified and with child care experience. The childcare has activities and play materials such as arts and crafts as well as appropriate toys for babies as well as a quiet area for napping.

We also have the Arbutus Room available for nursing mothers with a fridge is available on request. Please stop by the registration desk to schedule time and for the key.

In addition to childcare, we invite you to bring your children, spouses and loved ones to the social event at the conference. You can purchase additional meals or additional tickets to the social event for your family members. The social event for this year is an evening at the Vancouver Aquarium Marine. It's a family friendly venue, and we hope to see you and your family there!

Below is information from KiddieCorp about their services.

Hello ACL Parents!

Thank you very much for your interest in the Association for Computational Linguistics children's program. Our goal is to provide your children with a program they want to attend, while providing you with that critical "peace of mind" feeling so you can attend the conference activities without worrying.

KiddieCorp is pleased to provide a children's program during ACL 2017. KiddieCorp has more than thirty years of experience providing high quality children's programs and youth services to conventions, trade shows and special events. We take caring for your children very seriously. KiddieCorp has enjoyed a long-time partnership with the American Academy of Pediatrics, which has helped to establish KiddieCorp as a premier provider of event children's program services.

Activities Activities include exciting themes, arts and crafts, group games, music and movement, board games, story time, dramatic play, etc. We provide activities appropriate for each age group, using safe, sturdy equipment that you can feel comfortable with. Children can make their own choices within KiddieCorp's program.

Commitment Our goal is to provide your children with a comfortable, safe and happy experience. Our staff to child ratios are high to ensure that every child feels special (1:2 for children ages 6 months through 11 months old; 1:3 for children ages 1 through 2 years old; 1:5 for children ages 3 through 5 years old; 1:7 for children ages 6 through 12 years old). KiddieCorp team members are selected according to their integrity, experience, education and enthusiasm. They must be wonderful with kids! In addition to our competitive and selective hiring process, KiddieCorp remains at the top of the industry by carrying ample liability insurance.

Where, When, For Whom The program is for children ages 6 months through 12 years old. The dates for the program are July 30 – August 4, 2017 and will be located in the Oak Room at the Westin Bayshore in Vancouver. Snacks and beverages will be provided and meals need to be supplied by parents when checking in your child each day.

Other Info

- Please label your child's belongings. We will maintain a lost and found, however, KiddieCorp does not accept responsibility for the loss or theft of any toy, book, or other personal items.

- For parents with infants, please bring diaper changing supplies, formula/baby food, and a change of clothes.

- Cancellation Policy: Cancellations must be made to KiddieCorp prior to June 30, 2017 for a full refund. Cancellations made after that date will be subject to a 50% cancellation fee. Once the program has begun, no refunds will be issued.

- KiddieCorp staff does not administer medication. To ensure a safe and fun-filled environment, any child who is ill will not be admitted to the children's program.

Need more information? KiddieCorp is always available to answer any questions. Feel free to stop by the Oak room, or to call or text KiddieCorp's on-site manager, Lisa. Lisa's number is +1-450-466-6897.

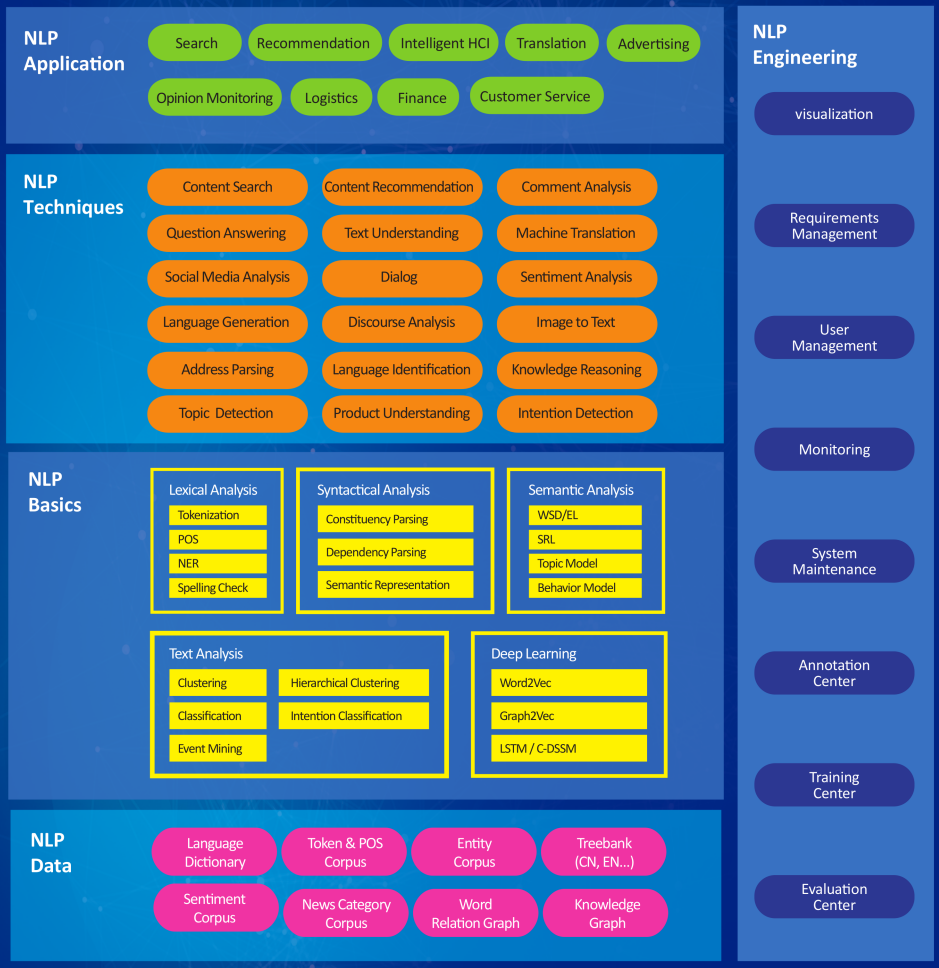
Example activities - A Pirate's Life For Me

Yo, Ho, Ho, Mateys! It's a pirate's life for us. We'll have a swashbuckling good time when we discover the KiddieCorp pirate life. Alrighty buccaneers, let's dress the part; grab an eye patch, pirate hat or handkerchief and we'll set sail with a serious game of Battleship. Before we head back to the mainland, we'll have an opportunity to create a pirate ship of our own out of Popsicle sticks at the arts and crafts station. We can also make pirate flags, treasure maps and our own beaded treasures. So let's raise the sails and get ready to have an adventure on the high seas, KiddieCorp pirate style.

Index

Alibaba Group
阿里巴巴集团

Data play an important role in the rich international eco-system of Alibaba. All kinds Natural Language Processing (NLP) for text data are very central and critical. The NLP techniques of Alibaba cover all kinds of multilingual NLP core data, multilingual foundational algorithms (lexical, syntactical, semantic, document analysis, and deep learning), and all kinds of NLP application techniques. They are widely applied in various Alibaba eco-system's business scenarios, such as search, recommendation, advertisement, finance, logistics, customer service, and cross-border e-commerce. Currently, NLP directly related technique platforms of Alibaba include large scale multilingual NLP platform AliNLP, machine translation platform AliTranX, and intelligent customer service platform AliMe, and so on.



We're a company of pioneers.

It's our job to make bold bets, and we get our energy from inventing on behalf of customers. Success is measured against the possible, not the probable. For today's pioneers, that's exactly why there's no place on Earth they'd rather build than Amazon.

Apply today.

We are hiring engineers and applied scientists around the world.

TEAM SPOTLIGHTS

Email your resume to
ACL2017@amazon.com
and learn more at
www.amazon.jobs/acl

Core Machine Learning
Amazon Music
Alexa
Personalization Sciences
AWS Deep Learning
Amazon Search/A9
Smart Home
Amazon Go
Advertising Technology
Alexa Shopping
Retail Systems Machine Learning
Customer Service Technology
Lab 126

MAKE THE COMPLICATED WORLD SIMPLER THROUGH TECHNOLOGY

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

Data Science at Bloomberg

Bloomberg's core product, the Terminal, is a must-have for the most influential people in finance. Bloomberg scientists and engineers develop finance-oriented NLP applications for NER & NED, question answering, sentiment analysis, market impact indicators, social media analysis, topic clustering and classification, recommendation systems and predictive models of market behavior. Our customers rely on this information to make swift financial decisions.

We are hiring in NLP, IR, ML and Data Science

We publish in leading academic venues and contribute to open source

We fund academic research through the Bloomberg Data Science Research Grants program

We welcome faculty visitors (summer and sabbatical)

We offer internships and postdocs

Find out more:

techatbloomberg.com/post-topic/data-science
github.com/bloomberg

Bloomberg

SAMSUNG

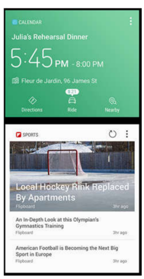
Positions in AI Research

We are looking for Natural Language Processing researchers

Research Areas & Related Products

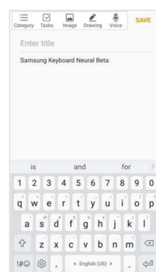
Core technology	Technology components
Language Understanding	Natural Language Understanding and Generation, Dialog System, Q&A, Machine Translation, Speech Recognition, Language Modeling
Inference / Prediction	Inference, Prediction, Planning, Recommendation
Knowledge Representation	Knowledge Creation, KB Management
Machine Learning	Deep Learning, Reinforcement Learning
Robotics	Autonomous Navigation, Visual-Motor Coordination
Visual Understanding	Object Recognition / Tracking, Visual Search, Human Recognition / Understanding, Scene Understanding

Dialog & QA



Recommender

Machine Translator



Input Prediction

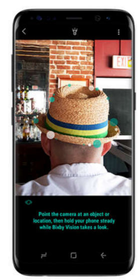


Image Recognition

Key Competencies

- M.S. or Ph.D. degree in Computer Science with specialization in deep learning / natural language processing / computer vision / robotics
- Experience with deep learning or related fields
- Experience in AI-related data challenges

HOW TO APPLY

- Please contact smin.seong@samsung.com
- Location : Seoul, South Korea

Artificial Intelligence Team

Software Center - Samsung Electronics.

Tencent AI Lab

Computer Vision

Automatic Speech Recognition

Machine Learning

Natural Language Processing

Make AI Everywhere

ai.tencent.com

SAP Leonardo Machine Learning

Explore the Intelligent Enterprise with SAP.
Visit us at our booth in the Grand Foyer.

SAP Run Simple

Emma is six.
The same age as Watson.

What does a world with Watson look like?

With Watson, the world is healthier, more productive, more secure, more creative, more engaging, more personal and more efficient.

Watson is already teaming with people in 45 countries and 20 industries—working on everything from moon shots to the everyday. Now it's your turn. What will you do with Watson?

Only with Watson.

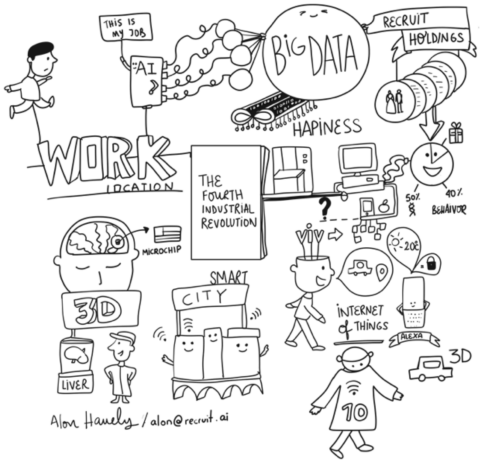
ibm.com/workwithwatson

IBM

Enriching lifestyles with Information Technology

Recruit Institute of Technology (RIT) is the technology hub and research lab for Recruit Holdings, a company that provides over 200 online services in the areas of human resources, travel, housing, restaurants and many other areas in which people make daily lifestyle decisions.

We conduct research in several areas, including natural language processing, machine learning, artificial intelligence, data management, and data integration. We collaborate with universities and publish in top-notch conferences.



Example Projects:

NLP for search engines: RIT is developing natural language processing techniques to improve the understanding of queries that users pose to vertical search engines and the textual content these engines store though conversational interfaces.

Chatbots: we are developing tools for easily building chatbots in vertical domains, including how to pose questions and extract answers and finding natural conversation flows.

Joybot: we are fielding our NLP technology in Joybot, a conversational agent that helps users set goals for increasing their happiness.

444 Castro St Suite 900, Mountain View, CA 94041, USA | www.recruit.ai

NAVER | LINE

ARTIFICIAL INTELLIGENCE

PRESENT AND FUTURE OF NAVER | LINE

Clova

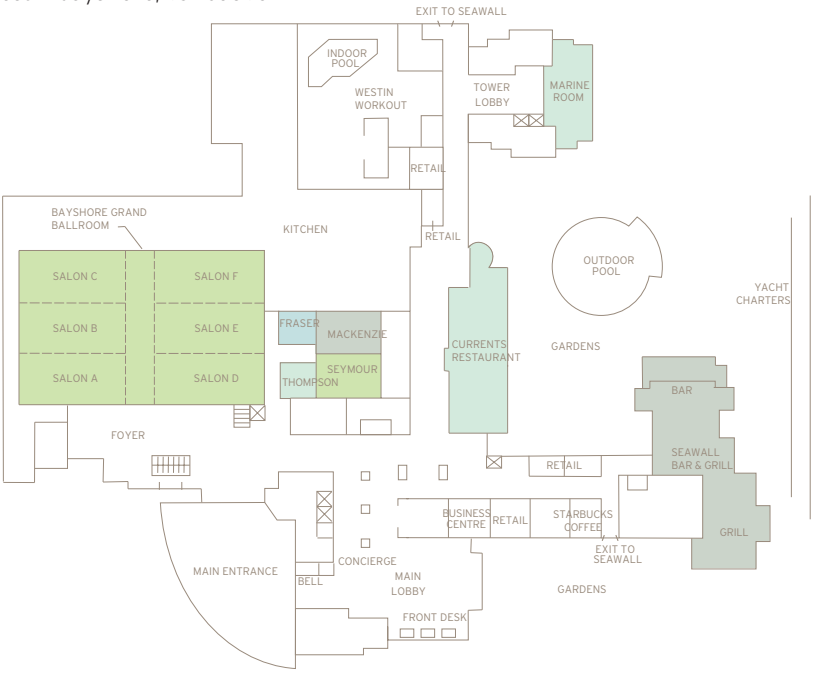
- Collaboration project between NAVER and LINE for AI-assistant platform aimed to artificial general intelligence
- Clova app: launched in Korea in May 2017
- WAVE and Champ: Summer and Winter in Korea / Japan in 2017

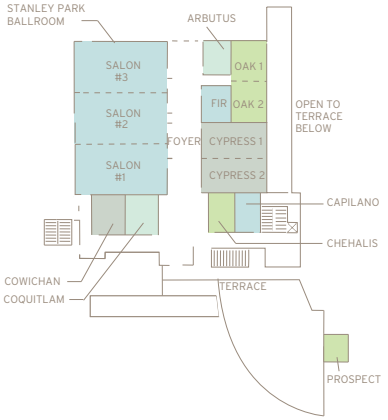
Core Technologies

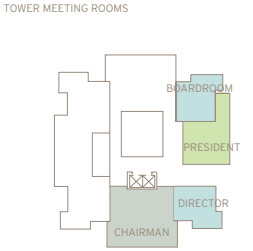
- Natural Language Processing
- Dialog management
- Recommendation
- Computer vision
- Pattern recognition

Based on machine learning including deep learning and reinforcement learning

floor plans
the westin bayshore, vancouver







THE WESTIN
BAYSHORE
Vancouver

We are grateful to our ACL 2017 Sponsors for their support!

Platinum

Alibaba Group 阿里巴巴集团 · amazon · Baidu 百度 百度一下 你就知道 · Bloomberg® · facebook · Google · SAMSUNG · Tencent 腾讯

Gold

ebay · ELSEVIER · KPMG · Maluuba A Microsoft company · SAP · IBM Research · RECRUIT Institute of Technology · NAVER · LINE

Silver

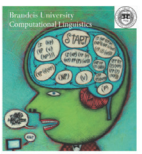
Orchestrating a brighter world · NEC · Microsoft

Bronze

Adobe · BOSCH Invented for life · CVTE 视源股份 · duolingo · HUAWEI · NUANCE · 搜狗搜索 · ORACLE

Supporter

grammarly · 头条 TOUTIAO · Yandex



Data Science Institute COLUMBIA UNIVERSITY · NEWSELA · NYU Center for Data Science · UNIVERSITY of WASHINGTON