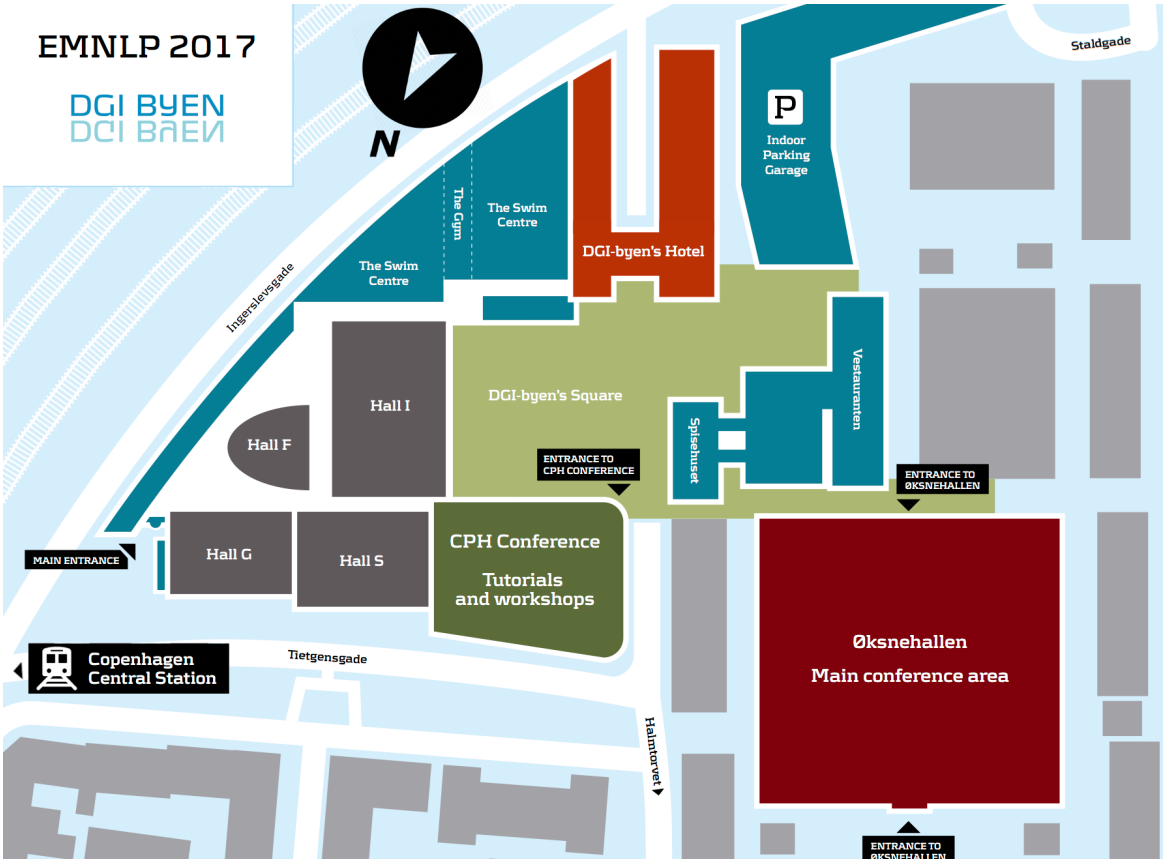
emnlp2017

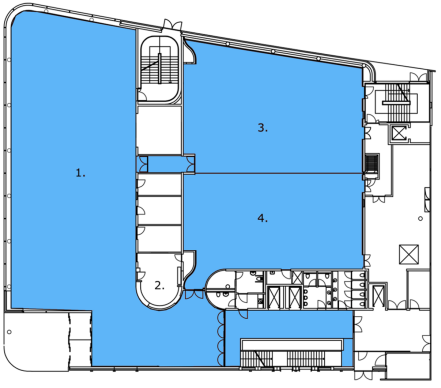
Conference Handbook

Copenhagen, Denmark
September 7-11, 2017

Cover design by Aurelia Bunescu
Cover photo by Matthew James
Deep gratitude to Matt Post for invaluable advice with creating this handbook
Handbook assembled by Joachim Bingel
Printing by Frederiksbergs Bogtrykkeri A/S

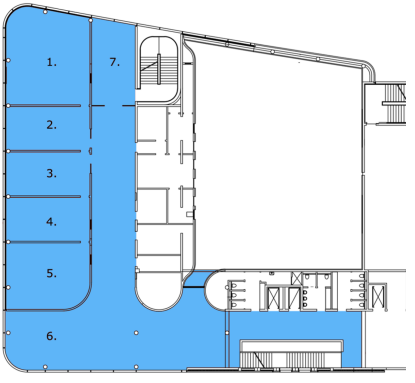

EMNLP 2017
DGI BYEN
N
Staldgade
P
Indoor Parking Garage
The Swim Centre
The Gym
The Swim Centre
DGI-byen's Hotel
Ingerslevsgade
Hall I
DGI-byen's Square
Vestauranten
Spisehuset
Hall F
ENTRANCE TO CPH CONFERENCE
ENTRANCE TO ØKSNEHALLEN
MAIN ENTRANCE
Hall G
Hall S
CPH Conference
Tutorials and workshops
Øksnehallen
Main conference area
Copenhagen Central Station
Tietgensgade
Halmtorvet
ENTRANCE TO ØKSNEHALLEN

CPH Conference Floorplan
(Most workshops, three tutorials)



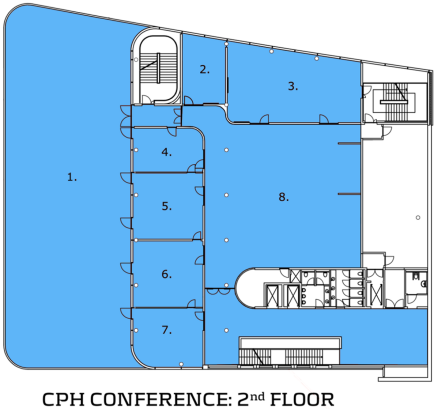
CPH CONFERENCE: GROUND FLOOR

1. Lobby
2. Reception
3. Sankt Hans Torv
4. Nørrebros Runddel



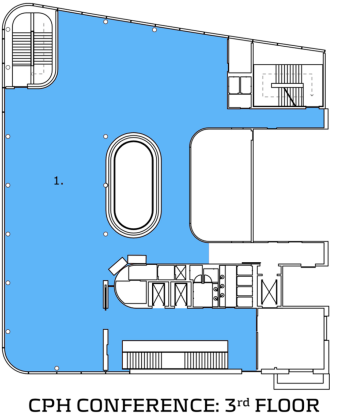
CPH CONFERENCE: 1st FLOOR

1. Kastrup Lufthavn
2. Christianshavn
3. Islands Brygge
4. Christiania
5. Amager Strandpark
6. Break/ Lounge area
7. Break/ Lounge area



CPH CONFERENCE: 2nd FLOOR

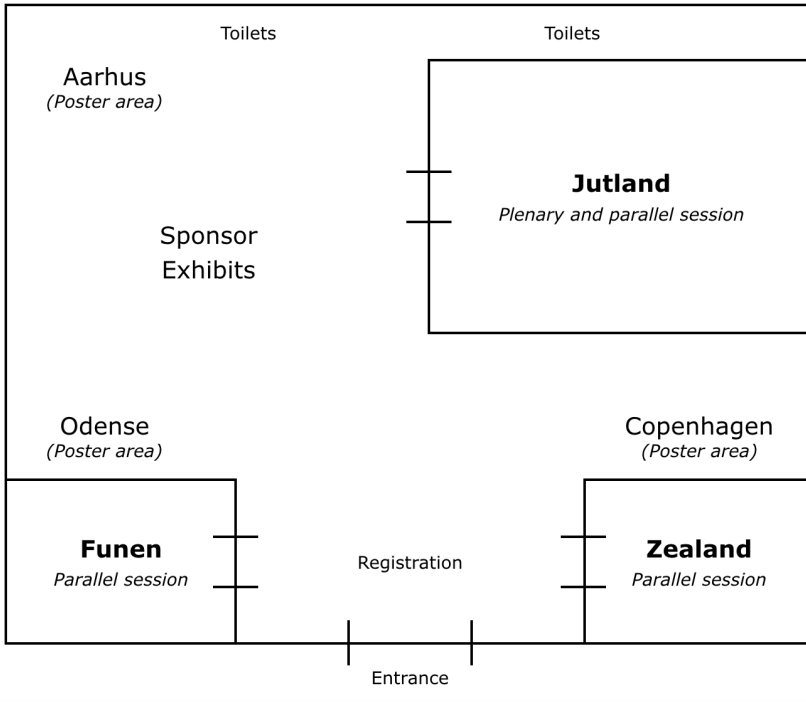
1. Roof terrace
2. Istedgade
3. Hovedbanen
4. Tivoli
5. Vesterbro Torv
6. Enghave Plads
7. Kødbyen
8. Break / Lounge area



CPH CONFERENCE: 3rd FLOOR

1. Østerbro

Øksnehallen Floorplan
(Main Conference, WMT, Tutorials 5 and 6)



Contents

1

Conference Information

Message from the General Chair

Thank you so much for joining us in Copenhagen. Welcome to a cosmopolitan city of fantastic restaurants, lovely seascapes, rich history, and lots and lots of cyclists!

We have an exciting program lined up for you, with three invited talks, fifteen workshops, seven tutorials, nine TACL presentations, 323 reviewed papers presented as both oral talks and posters, and twenty-one demos. I am especially grateful to our Program Chairs, Rebecca Hwa and Sebastian Riedel, who did a fantastic job managing a backbreaking 1,500 paper submissions (1466 reviewed papers). This involved 51 Area chairs and 980 reviewers. We tried some new things this year (never conducive to a smooth process) including a more careful handling of the COIs that result from Area Chair submissions, and the addition of a meta-review step to encourage more thoughtful reviewing. We are soliciting feedback on the meta-review process, from both reviewers and authors. Despite the additional time involvement, many of the Area Chairs embraced this new approach, and would like to repeat it. However, there are clearly a few dissenters, since Rebecca and Sebastian ended up writing around 200 meta-reviews themselves at the last minute. We are also trying to raise the visibility and status of the poster sessions by integrating them as parallel sessions alongside oral talks, with poster session chairs. This is in response to the survey results from EMNLP 2015 that indicated a decided preference for smaller, more frequent poster sessions during the day rather than evening mega-sessions. Finally, Rebecca and Sebastian are bringing you three outstanding invited speakers, Dan Jurafsky, Sharon Goldwater, and Nando de Freitas. No program chairs ever worked harder to bring you a superb set of presentations in an attendee friendly setting.

I am also very grateful to Victoria Fossum and Karl Moritz Hermann, our Workshop Chairs, who put together a terrific slate of fifteen workshops, and paid meticulous attention to ensuring that each workshop could hold exactly the poster sessions, invited talks and special events that it required. Our tutorial chairs, Alexandra Birch and Nathan Schneider, also outdid themselves, providing seven especially tempting tutorial offerings. Matt Post deserves to be singled out, for being an Advisor to our conscientious and successful Handbook Chair, Joachim Bingel, as well as becoming a welcome last minute addition to our excellent team of Demo Chairs, Lucia Specia and Michael Paul. Thanks are due to our Website Chair, Anders Johannsen, who responded promptly and deftly to all of our requests, and to Chloé Braud, who single-handedly took on the task of creating the conference app in Conference4me. Our Student Volunteer and Student Sponsorship Chairs, Zeljko Agic and Yonatan Bisk, worked hard to bring you the helpful and energetic volunteers who keep things running smoothly, for which we are also grateful.

Last but not least, many thanks to your hosts, our Local Arrangements Chairs, Dirk Hovy and Anders Søgaard and their team. Their concern has been increasing the enjoyment of your experience, and to that end they proposed a stunning venue, put together an amazing reception and Social Event, chose your conference bags, issued all the invitation letters for visas, helped create all the signs, etc., etc., etc. Dan Hardt, our Sponsorship chair, working with Anders and Dirk, raised an unusual amount of local sponsorships, all to defray the cost of the Social Event.

As always, we are extremely indebted to our generous sponsors. Our platinum sponsors are Amazon, Apple, Baidu, Bloomberg, Facebook, Google, and Siteimprove. Gold sponsors include Deloitte, ebay, IBM Research, Maluuba, Microsoft, SAP, Recruit Institute of Technology, textkernel, and Zalando. Silver sponsors are CVTE, Duolingo, Huawei, Nuance, Oracle, Snapchat, Sogou, Unsilo and Wizkids. Grammarly, NextAI and Yandex are our Bronze sponsors.

Finally, many, many thanks to our Area Chairs, our reviewers, and our authors, whose outstanding research is being showcased here for your delectation. *Nyd det mens det varer!*

Best Regards,
Martha Palmer
EMNLP 2017 General Chair

Message from the Program Committee Co-Chairs

Welcome to the 2017 Conference on Empirical Methods in Natural Language Processing! This is an exciting year; we have received a new record-high in the number of submissions: 1,509 papers. After discounting early withdraws, duplicates, and other invalid submissions, we sent out 1,418 submissions (836 long papers, 582 short papers) to be reviewed by the program committee. Ultimately, 216 long papers (25.8% acceptance rate) and 107 short papers (18.4% acceptance rate) have been accepted for presentation, making a total of 323 papers and an overall acceptance rate of 22.8%.

This year's technical program consists of three invited talks and 113 oral presentations and 219 poster presentations for the 323 long and short accepted papers as well as nine papers accepted to the Transactions of the Association for Computational Linguistics. To accommodate all the presentations in a compressed timeframe, we opted to have plenary sessions for the invited talks and the winners of the Best Paper Awards, while allotting three parallel oral sessions and thematically related poster sessions for all other presentations. We chose to have concurrent poster and oral sessions for several reasons. First, this is the preferred model of the majority (51.6%) of participants who filled out the EMNLP 2015 post-conference survey. Second, this allows us to spread out the poster presentations across three days in smaller thematically related clusters. Finally, this maximises the number of acceptances for the high quality submissions we received; by having more poster sessions, we are able to maintain the acceptance rates at the previous year's level despite an increase in submissions by 40%.

It would not have been possible to properly handle such a large number of submissions without the generous voluntary help from all the members of the program committee, which consists of 980 reviewers overseen by 51 area chairs. We continued last year's experiment of defining twelve relatively broad topic areas and assigning multiple area chairs to facilitate consistent ranking of larger sets of papers. Most technical program decisions, from the selection of papers to the modes of presentation to the choice of outstanding papers, are primarily made in a bottom-up fashion: reviewers assessed and scored papers, made recommendations for oral vs poster decisions, and marked papers suitable for best paper awards; area chairs ensured the quality of assessments, encouraged discussions and assembled opinions into their own recommendations; finally, we construct the technical program, considering the recommendations from the area chairs while taking into account venue constraints and balance across areas. A new experimental feature of this year's EMNLP reviewing process is the "meta review," in which the area chairs briefly summarize the major discussions between the reviewers to give authors a more transparent view of the process.

Per EMNLP tradition, awards are given to outstanding papers in three categories: Best Long Paper, Best Short Paper, and Best Resource Paper. The selection process is bottom-up: based on the reviewers and area chairs' recommendations, we nominated four papers for each category; we invited expert members to form a Best Papers committee for each category; each committee reviews the candidates and select the winners. The awarded papers will be presented at a special plenary session on the last day of the conference.

We are extremely grateful that three amazing speakers have agreed to give invited talks at EMNLP. Nando de Freitas (Google Deepmind) will discuss simulated physical environments, and whether language would benefit from the development of such environments, and could contribute toward improving such environments and agents within them. Sharon Goldwater (University of Edinburgh) will describe work on developing unsupervised speech technology for those of the world's 7,000 or so languages not spoken in large rich countries. Dan Jurafsky (Stanford University) will talk about processing the language of policing to automatically measure linguistic aspects of the interaction from discourse factors like conversational structure to social factors like respect.

The conference would not have been possible without the support of various people inside and outside of the committee. In particular, we would like to thank:

- Martha Palmer, whose encouragement and advice as the general chair has been invaluable every step of the way;
- Chris Callison-Burch, who has given us excellent advice and support in his capacity as the SIGDAT Secretary;
- Priscilla Rasmussen, who always has the right answers;
- Xavier Carreras and Kevin Duh, who generously shared their experiences as the chairs of EMNLP 2016;
- Anders Johannsen, who is lightning fast with website updates;
- Our 51 area chairs: David Bamman, Mohit Bansal, Roberto Basili, Chris Biemann, Jordan Boyd-Graber, Marine Carpuat, Joyce Chai, David Chiang, Jinho Choi, Jennifer Chu-Carroll, Trevor Cohn, Cristian Danescu-Niculescu-Mizil, Dipanjan Das, Hal Daume, Mona Diab, Mark Dredze, Jacob Eisenstein, Sanja Fidler, Alona Fyshe, Dan Gildea, Ed Grefenstette, Hannaneh Hajishirzi, Julia Hockenmaier, Kentaro Inui, Jing Jiang, Philipp Koehn, Mamoru Komachi, Anna Korhonen, Tom Kwiatkowski, Gina Levow, Bing Liu, Nitin Madnani, Mausam, Rada Mihalcea, Marie-Francine Moens, Saif M. Mohammad, Mari Ostendorf, Sameer Pradhan, Alexander Rush, Anoop Sarkar, William Schuler, Hinrich Schütze, Sameer Singh, Thamar Solorio, Vivek Srikumar, Amanda Stent, Tomek Strzalkowski, Mihai Surdeanu, Andreas Vlachos, Scott Wentau Yih, Zhang Yue;
- The best papers award committee members: Chris Brew, Mike Collins, Kevin Duh, Adam Lopez, Ani Nenkova, Bonnie Webber, Luke Zettlemoyer;
- Preethi Raghavan and Siddharth Patwardhan, the publications co-chairs and Joachim Bingel, the conference handbook chair;
- Dirk Hovy and Anders Søgaard, the local arrangements co-chairs;
- Rich Gerber and Paolo Gai at SoftConf.

Finally, we'd like to thank SIGDAT for the opportunity to serve as Program Co-Chairs of EMNLP 2017. It is an honor and a rewarding learning experience. We hope you will be as inspired by the technical program as we are.

EMNLP 2017 Program Co-Chairs
Rebecca Hwa, University of Pittsburgh
Sebastian Riedel, University College London

Organizing Committee

General Chair
Martha Palmer, University of Colorado

Local Arrangements Chair
Priscilla Rasmussen, ACL Business Manager

Program Committee Co-chairs
Rebecca Hwa, University of Pittsburgh
Sebastian Riedel, University College London

Local Arrangements Co-chairs
Dirk Hovy, University of Copenhagen
Anders Søgaard, University of Copenhagen

Local Sponsorship Chair
Daniel Hardt, Copenhagen Business School

Workshop Co-chairs
Victoria Fossum, Google
Karl Moritz Hermann, DeepMind

Tutorial Co-chairs
Alexandra Birch, University of Edinburgh
Nathan Schneider, Georgetown University

Demos Co-chairs
Lucia Specia, University of Sheffield
Matt Post, Johns Hopkins University
Michael Paul, University of Colorado

Publications Sr Chair
Siddharth Patwardhan, Apple

Publications Jr Chair
Preethi Raghavan, TJ Watson Lab, IBM

Publicity Chair
Isabelle Augenstein, University of Copenhagen

Web Chair
Anders Johannsen, Apple

Conference Handbook Chair
Joachim Bingel, University of Copenhagen

Conference Handbook Advisor
Matt Post, Johns Hopkins University

Handbook Proofreader
Pontus Stenetorp, University College London

Conference App Chair
Chloé Braud, University of Copenhagen

Student Scholarship Co-chair and Student Volunteer Coordinator
Željko Agić, IT University of Copenhagen
Yonatan Bisk, University of Southern California, ISI

SIGDAT Liason
Chris Callison-Birch, University of Pennsylvania

Program Committee

Program Committee Co-chairs
Rebecca Hwa, University of Pittsburgh
Sebastian Riedel, University College London

Area Chairs
Information Extraction, Information Retrieval, and Question Answering

Jing Jiang, Singapore Management University
Mausam, IIT Delhi
Hinrich Schütze, LMU Munich
Sameer Singh, UC Irvine
Mihai Surdeanu, University of Arizona
Tomek Strzalkowski, SUNY Albany
Scott Wen-tau Yih, MSR

Language and Vision
Sanja Fidler, University of Toronto
Hannaneh Hajishirzi, University of Washington

Linguistic Theories and Psycholinguistics
William Schuler, The Ohio State University

Machine Learning
Mohit Bansal, UNC Chapel Hill
Jordan Boyd-Graber, University of Colorado
Trevor Cohn, University of Melbourne
Hal Daumé, University of Maryland
Alona Fyshe, University of Victoria
Anoop Sarkar, Simon Fraser University

Machine Translation and Multilinguality
Marine Carpuat, University of Maryland
David Chiang, University of Notre Dame
Mona Diab, George Washington University
Dan Gildea, University of Rochester
Philipp Koehn, Johns Hopkins University

Segmentation, Tagging, and Parsing
Jinho Choi, Emory University
Julia Hockenmaier, University of Illinois at Urbana-Champaign
Alexander Rush, Harvard University
Zhang Yue, Singapore University of Technology and Design

Semantics
Roberto Basili, University of Roma, Tor Vergata
Chris Biemann, University of Hamburg
Ed Grefenstette, DeepMind
Tom Kwiatkowski, Google
Sameer Pradhan, cemantix.org and Boulder Learning, Inc
Vivek Srikumar, University of Utah

Sentiment Analysis and Opinion Mining
Bing Liu, University of Illinois at Chicago
Rada Mihalcea, University of Michigan
Saif M. Mohammad, National Research Council Canada

Social Media and Computational Social Science
David Bamman, University of California, Berkeley
Cristian Danescu-Niculescu-Mizil, Cornell University
Mark Dredze, Johns Hopkins University

Jacob Eisenstein, Georgia Tech

Spoken Language Processing

Mari Ostendorf, University of Washington

Summarization, Generation, Discourse, Dialogue

Joyce Chai, Michigan State University
Jennifer Chu-Carroll, Elemental Cognition
Kentaro Inui, Tohoku University
Gina Levow, University of Washington
Amanda Stent, Bloomberg LP

Text Mining and NLP Applications

Dipanjan Das, Google
Mamoru Komachi, Tokyo Metropolitan University
Anna Korhonen, University of Cambridge
Nitin Madnani, Educational Testing Service (ETS)
Marie-Francine Moens, KU Leuven
Thamar Solorio, University of Houston
Andreas Vlachos, University of Sheffield

Venue Info

Please note that EMNLP takes place in *four different buildings*. Almost all workshops are located in **CPH Conference**, while the WMT workshop, Tutorials 5 and 6 as well as the entire main conference take place in **Øksnehallen**.

Besides these, we will have Tutorials 2 and 4 located in **Spisehuset**, and the BEA workshop will take place in the Conference Hotel (**DGI Byen Hotel, Room 3**).

There are maps with room locations, as well as an area map, in the front matter.

Meal Info

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

- During the coffee breaks in the mornings and afternoons, **snacks** are provided.
- There will also be **snacks** at the welcome reception.
- A **full dinner** with various options is provided as part of the social event.

Welcome Reception

Friday, September 8, 2017, 19:00 – 22:00

CPH Conference, Room *Østerbro*

Catch up with your colleagues at the **Welcome Reception**! It will be held following the tutorials and workshops on Friday evening.

2

Tutorials: Thursday, September 7

Overview

7:30 – 18:00	**Registration**	*Lobby*
9:00 – 12:30	**Morning Tutorials**	
	Acquisition, Representation and Usage of Conceptual Hierarchies *Marius Pasca*	*Skt. Hans Torv*
	Computational Sarcasm *Pushpak Bhattacharyya and Aditya Joshi*	*Spisehuset*
10:30 – 11:00	**Coffee break**	*Multiple levels*
12:30 – 14:00	**Lunch break**	
14:00 – 17:30	**Afternoon Tutorials**	
	Graph-based Text Representations: Boosting Text Mining, NLP and Information Retrieval with Graphs *Fragkiskos D. Malliaros and Michalis Vazirgiannis*	*Skt. Hans Torv*
	Semantic Role Labeling *Diego Marcheggiani, Michael Roth, Ivan Titov, and Benjamin Van Durme*	*Spisehuset*
15:30 – 16:00	**Coffee break**	*Multiple levels*

Tutorial 1

Acquisition, Representation and Usage of Conceptual Hierarchies

Marius Pasca

Thursday, September 7, 2017, 9:00–12:30

Location: Skt. Hans Torv

Through subsumption and instantiation, individual instances ("artificial intelligence", "the spotted pig") otherwise spanning a wide range of domains can be brought together and organized under conceptual hierarchies. The hierarchies connect more specific concepts ("computer science subfields", "gastropubs") to more general concepts ("academic disciplines", "restaurants") through IsA relations. Explicit or implicit properties applicable to, and defining, more general concepts are inherited by their more specific concepts, down to the instances connected to the lower parts of the hierarchies. Subsumption represents a crisp, universally-applicable principle towards consistently representing IsA relations in any knowledge resource. Yet knowledge resources often exhibit significant differences in their scope, representation choices and intended usage, to cause significant differences in their expected usage and impact on various tasks.

This tutorial examines the theoretical foundations of subsumption, and its practical embodiment through IsA relations compiled manually or extracted automatically. It addresses IsA relations from their formal definition; through practical choices made in their representation within the larger and more widely-used of the available knowledge resources; to their automatic acquisition from document repositories, as opposed to their manual compilation by human contributors; to their impact in text analysis and information retrieval. As search engines move away from returning a set of links and closer to returning results that more directly answer queries, IsA relations play an increasingly important role towards a better understanding of documents and queries. The tutorial teaches the audience about definitions, assumptions and practical choices related to modeling and representing IsA relations in existing, human-compiled resources of instances, concepts and resulting conceptual hierarchies; methods for automatically extracting sets of instances within unlabeled or labeled concepts, where the concepts may be considered as a flat set or organized hierarchically; and applications of IsA relations in information retrieval.

Marius Pasca is a research scientist at Google. Current research interests include factual information extraction from unstructured text within documents and queries and its applications to Web search.

Tutorial 2

Computational Sarcasm

Pushpak Bhattacharyya and Aditya Joshi

Thursday, September 7, 2017, 9:00–12:30

Location: Spisehuset

Sarcasm is a form of verbal irony that is intended to express contempt or ridicule. Motivated by challenges posed by sarcastic text to sentiment analysis, computational approaches to sarcasm have witnessed a growing interest at NLP forums in the past decade. Computational sarcasm refers to automatic approaches pertaining to sarcasm. The tutorial will provide a bird's-eye view of the research in computational sarcasm for text, while focusing on significant milestones.

The tutorial begins with linguistic theories of sarcasm, with a focus on incongruity: a useful notion that underlies sarcasm and other forms of figurative language. Since the most significant work in computational sarcasm is sarcasm detection: predicting whether a given piece of text is sarcastic or not, sarcasm detection forms the focus hereafter. We begin our discussion on sarcasm detection with datasets, touching on strategies, challenges and nature of datasets. Then, we describe algorithms for sarcasm detection: rule-based (where a specific evidence of sarcasm is utilised as a rule), statistical classifier-based (where features are designed for a statistical classifier), a topic model-based technique, and deep learning-based algorithms for sarcasm detection. In case of each of these algorithms, we refer to our work on sarcasm detection and share our learnings. Since information beyond the text to be classified,

---

Prof. Pushpak Bhattacharyya is the current President of ACL (2016-17). He is the Director of IIT Patna and Vijay and Sita Vashee Chair Professor in IIT Bombay, Computer Science and Engineering Department. He was educated in IIT Kharagpur (B.Tech), IIT Kanpur (M.Tech) and IIT Bombay (PhD). He has been visiting scholar and faculty in MIT, Stanford, UT Houston and University Joseph Fouriere (France). Prof. Bhattacharyya's research areas are Natural Language Processing, Machine Learning and AI. He has guided more than 250 students (PhD, masters and Bachelors), has published more than 250 research papers and led government and industry projects of international and national importance. A significant contribution of his is Multilingual Lexical Knowledge Bases and Projection. Author of the text book 'Machine Translation' Prof. Bhattacharyya is loved by his students for his inspiring teaching and mentorship. He is a Fellow of National Academy of Engineering and recipient of Patwardhan Award of IIT Bombay and VNMM award of IIT Roorkey- both for technology development, and faculty grants of IBM, Microsoft, Yahoo and United Nations.

Aditya Joshi is a PhD student at IITB-Monash Research Academy, a joint PhD programme between Indian Institute of Technology Bombay, India and Monash University, Australia, since January 2013. His PhD advisors are Pushpak Bhattacharyya (IITB) and Mark Carman (Monash). His primary research focus is computational sarcasm where he has explored different ways in which incongruity can be captured in order to detect and generate sarcasm. In addition, he has worked on innovative applications of NLP such as sentiment analysis for Indian languages, drunk-texting prediction, news headline translation, political issue extraction, etc.

contextual information is useful for sarcasm detection, we then describe approaches that use such information through conversational context or author-specific context.

We then follow it by novel areas in computational sarcasm such as sarcasm generation, sarcasm v/s irony classification, etc. We then summarise the tutorial and describe future directions based on errors reported in past work. The tutorial will end with a demonstration of our work on sarcasm detection.

This tutorial will be of interest to researchers investigating computational sarcasm and related areas such as computational humour, figurative language understanding, emotion and sentiment sentiment analysis, etc. The tutorial is motivated by our continually evolving survey paper of sarcasm detection, that is available on arXiv at: Joshi, Aditya, Pushpak Bhattacharyya, and Mark James Carman. "Automatic Sarcasm Detection: A Survey." arXiv preprint arXiv:1602.03426 (2016).

Tutorial 3

Graph-based Text Representations: Boosting Text Mining, NLP and Information Retrieval with Graphs

Fragkiskos D. Malliaros and Michalis Vazirgiannis

Thursday, September 7, 2017, 9:00–12:30

Location: Skt. Hans Torv

Graphs or networks have been widely used as modeling tools in Natural Language Processing (NLP), Text Mining (TM) and Information Retrieval (IR). Traditionally, the unigram bag-of-words representation is applied; that way, a document is represented as a multi-set of its terms, disregarding dependencies between the terms. Although several variants and extensions of this modeling approach have been proposed (e.g., the n-gram model), the main weakness comes from the underlying term independence assumption. The order of the terms within a document is completely disregarded and any relationship between terms is not taken into account in the final task (e.g., text categorization). Nevertheless, as the heterogeneity of text collections is increasing (especially with respect to document length and vocabulary), the research community has started exploring different document representations aiming to capture more fine-grained contexts of co-occurrence between different terms, challenging the well-established unigram bag-of-words model. To this direction, graphs constitute a well-developed model that has been adopted for text representation. The goal of this tutorial is to offer a comprehensive presentation of recent methods that rely on graph-based text representations to deal with various tasks in NLP and IR. We will describe basic as well as novel graph theoretic concepts and we will examine how they can be applied in a wide

---

Fragkiskos D. Malliaros is currently a data science postdoctoral scholar in the Department of Computer Science and Engineering at UC San Diego and member of the Artificial Intelligence group. Right before that, he was a postdoctoral researcher in Ecole Polytechnique, France from where he also received his Ph.D. degree in 2015. He obtained his Diploma and his M.Sc. degree from the University of Patras, Greece in 2009 and 2011 respectively. He is the recipient of the 2012 Google European Doctoral Fellowship in Graph Mining and the 2015 Thesis Prize by Ecole Polytechnique. During the summer of 2014, he was a research intern at the Palo Alto Research Center (PARC), working on anomaly detection in social networks. His research interests span the broad area of data science, with focus on graph mining, social network analysis, applied machine learning and natural language processing.

Michalis Vazirgiannis is a Professor in Ecole Polytechnique, France and the leader of the Data Science and Mining (DaSciM) team. He holds a degree in Physics, a M.Sc. in Robotics, both from University of Athens, Greece, and a M.Sc. in Knowledge Based Systems from Heriot-Watt University (Edinburgh, UK). He acquired his Ph.D. degree from the Dept. of Informatics, University of Athens. He has worked as a researcher in different places: NTUA, GMD-IPSI (currently Frauhofer-IPSI), Germany Fern-Universitaet Hagen, in project VERSO (later GEMO) in INRIA/Paris, in IBM India Research Laboratory and in MPI fur Informatik (Saarbruecken, Germany). He held a Marie Curie Intra-European fellow in area of P2P Web Search, hosted by INRIA FUTURS, Paris. His current research interests are on graph mining, text mining and recommendation algorithms. He is chairing the "AXA Data Science" chair in Ecole Polytechnique and has collaborations with the industry including Google and Airbus.

range of text-related application domains. All the material associated to the tutorial will be available at: `http://fragkiskosm.github.io/projects/graph_text_tutorial`

Tutorial 4

Semantic Role Labeling

Diego Marcheggiani, Michael Roth, Ivan Titov, and Benjamin Van Durme

Thursday, September 7, 2017, 14:00–17:30

Location: Spisehuset

This tutorial describes semantic role labelling (SRL), the task of mapping text to shallow semantic representations of eventualities and their participants. The tutorial introduces the SRL task and discusses recent research directions related to the task. The audience of this tutorial will learn about the linguistic background and motivation for semantic roles, and also about a range of computational models for this task, from early approaches to the current state-of-the-art. We will further discuss recently proposed variations to the traditional SRL task, including topics such as semantic proto-role labeling.

We also cover techniques for reducing required annotation effort, such as methods exploiting unlabeled corpora (semi-supervised and unsupervised techniques), model adaptation across languages and domains, and methods for crowdsourcing semantic role annotation (e.g., question-answer driven SRL). Methods based on different machine learning paradigms, including neural networks, generative Bayesian models, graph-based algorithms and bootstrapping style techniques.

Beyond sentence-level SRL, we discuss work that involves semantic roles in discourse. In particular, we cover data sets and models related to the task of identifying implicit roles and linking them to discourse antecedents. We introduce different approaches to this task from

---

Diego Marcheggiani is a postdoctoral researcher at the University of Amsterdam. He graduated with a Ph.D. in Computer Science from the University of Venice and during this period he worked at the ISTI-CNR in Italy as a researcher. His research focus ranges from relation extraction to semantic role labeling and frame-semantic parsing. He is interested in supervised and unsupervised learning approaches in the scope of tensor factorization models and neural networks.

Michael Roth is a postdoctoral researcher and DFG research fellow at Saarland University and University of Illinois at Urbana-Champaign, respectively. He graduated with a Ph.D. in Computational Linguistics from Heidelberg University in 2013. His research focus lies on computational models of language that can facilitate automatic text understanding beyond the sentence level. Recent work includes neural-network based approaches to semantic role labeling and discourse-level frame-semantic parsing. His models are the current state-of-the-art on the CoNLL-2009 and FrameNet 1.5 data sets.

Ivan Titov is an Associate Professor at the University of Amsterdam. He is the recipient of an ERC Starting Grant, a personal Vidi Grant from the Dutch NSF (NWO) and a Google Focused Research Award. Ivan is an action editor for the Journal of Machine Learning Research (JMLR) and Transactions of ACL (TACL), as well as an editorial board member of the Journal of Artificial Intelligence Research (JAIR). His interests are in probabilistic modeling of language, primarily in semantics and syntax as well as in multilingual NLP and semi-supervised learning for NLP.

Benjamin Van Durme is an Assistant Professor at the Johns Hopkins University in Computer Science, with a courtesy appointment in Cognitive Science, and the lead of the Natural Language Understanding group at the Human Language Technology Center of Excellence. (HLTCOE). His research is broadly focused on discovering and extracting knowledge from language, exploring topics such as low resource, multilingual information extraction; scalable, streaming algorithms for processing large collections; and semantic analysis at various levels of complexity.

the literature, including models based on coreference resolution, centering, and selectional preferences. We also review how new insights gained through them can be useful for the traditional SRL task.

3

Tutorials: Friday, September 8

Overview

7:30 – 18:00	**Registration**	*Lobby*
14:00 – 17:30	**Afternoon Tutorials**	
	Memory Augmented Neural Networks for Natural Language Processing *Caglar Gulcehre and Sarath Chandar*	*Zealand*
	A Unified Framework for Structured Prediction: From Theory to Practice *Wei Lu*	*Jutland*
	Cross-Lingual Word Representations: Induction and Evaluation *Manaal Faruqui, Anders Søgaard, and Ivan Vulić*	*Nørrebros Runddel*
15:30 – 16:00	**Coffee break**	*Multiple levels*
19:00 – 22:00	**Welcome Reception**	*CPH Conference, Room Østerbro*

Tutorial 5

Memory Augmented Neural Networks for Natural Language Processing

Caglar Gulcehre and Sarath Chandar

Friday, September 8, 2017, 14:00–17:30

Location: Zealand

Designing of general-purpose learning algorithms is a long-standing goal of artificial intelligence. A general purpose AI agent should be able to have a memory that it can store and retrieve information from. Despite the success of deep learning in particular with the introduction of LSTMs and GRUs to this area, there are still a set of complex tasks that can be challenging for conventional neural networks. Those tasks often require a neural network to be equipped with an explicit, external memory in which a larger, potentially unbounded, set of facts need to be stored. They include but are not limited to, reasoning, planning, episodic question-answering and learning compact algorithms. Recently two promising approaches based on neural networks to this type of tasks have been proposed: Memory Networks and Neural Turing Machines.

In this tutorial, we will give an overview of this new paradigm of "neural networks with memory". We will present a unified architecture for Memory Augmented Neural Networks (MANN) and discuss the ways in which one can address the external memory and hence read/write from it. Then we will introduce Neural Turing Machines and Memory Networks as specific instantiations of this general architecture. In the second half of the tutorial, we will focus on recent advances in MANN which focus on the following questions: How can we read/write from an extremely large memory in a scalable way? How can we design

---

Caglar Gulcehre is currently a research scientist at Deepmind. He finished his PhD in University of Montreal under the supervision of Yoshua Bengio. His work mainly focuses on applications of neural networks, in particular recurrent architectures such as GRU and LSTMs on NLP and sequence to sequence learning tasks. His research also investigates different optimization approaches and architectures which are easier to optimize for neural networks. His recent research focuses on building neural network models that have external memory structures. He has done research internships at IBM Watson Research Center, Google Deep Mind. He was a PC Member at ECML and IJCAI 2016 Deep Reinforcement Learning Workshop. Prior to joining MILA as a PhD student, he finished his master degree in Middle East Technical University in Cognitive Science department.

Sarath Chandar is currently a PhD student in University of Montreal under the supervision of Yoshua Bengio and Hugo Larochelle. His work mainly focuses on Deep Learning for complex NLP tasks like question answering and dialog systems. He also investigates scalable training procedure and memory access mechanisms for memory network architectures. In the past, he has worked on multilingual representation learning and transfer learning across multiple languages. His research interests includes Machine Learning, Natural Language Processing, Deep Learning, and Reinforcement Learning. Before joining University of Montreal, he was a Research Scholar in IBM Research India for a year. He has previously given a tutorial on "Multilingual Multimodal Language Processing using Neural Networks" at NAACL 2016.

efficient non-linear addressing schemes? How can we do efficient reasoning using large scale memory and an episodic memory? The answer to any one of these questions introduces a variant of MANN. We will conclude the tutorial with several open challenges in MANN and its applications to NLP.

We will introduce several applications of MANN in NLP throughout the tutorial. Few examples include language modeling, question answering, visual question answering, and dialogue systems.

Tutorial 6

A Unified Framework for Structured Prediction: From Theory to Practice

Wei Lu

Friday, September 8, 2017, 14:00–17:30

Location: Jutland

Structured prediction is one of the most important topics in various fields, including machine learning, computer vision, natural language processing (NLP) and bioinformatics. In this tutorial, we present a novel framework that unifies various structured prediction models.

The hidden Markov model (HMM) and the probabilistic context-free grammars (PCFGs) are two classic generative models used for predicting outputs with linear-chain and tree structures, respectively. As HMM's discriminative counterpart, the linear-chain conditional random fields (CRFs) (Lafferty et al., 2001) model was later proposed. Such a model was shown to yield good performance on standard NLP tasks such as information extraction. Several extensions to such a model were then proposed afterward, including the semi-Markov CRFs (Sarawagi and Cohen, 2004), tree CRFs (Cohn and Blunsom, 2005), as well as discriminative parsing models and their latent variable variants (Petrov and Klein, 2007). On the other hand, utilizing a slightly different loss function, one could arrive at the structured support vector machines (Tsochantaridis et al., 2004) and its latent variable variant (Yu and Joachims, 2009) as well. Furthermore, new models that integrate neural networks and graphical models, such as neural CRFs (Do et al., 2010) were also proposed.

In this tutorial, we will be discussing how such a wide spectrum of existing structured prediction models can all be implemented under a unified framework that involves some basic building blocks. Based on such a framework, we show how some seemingly complicated structured prediction models such as a semantic parsing model (Lu et al., 2008; Lu, 2014) can be implemented conveniently and quickly. Furthermore, we also show that the framework can be used to solve certain structured prediction problems that otherwise cannot be easily handled by conventional structured prediction models. Specifically, we show how to use such a framework to construct models that are capable of predicting non-conventional structures, such as overlapping structures (Lu and Roth, 2015; Muis and Lu, 2016a). We will also discuss how to make use of the framework to build other related models such as topic mod-

---

Wei Lu is an Assistant Professor at the Singapore University of Technology and Design (SUTD), directing the StatNLP research group. He received his Ph.D. from the National University of Singapore (NUS) in 2009. He visited CSAIL, Massachusetts Institute of Technology (MIT) in 2007-2008, and worked as a postdoctoral research associate at the University of Illinois at Urbana-Champaign in 2011-2013. His research interests include developing mathematical models and machine learning algorithms for solving natural language processing problems. He is particularly interested in semantic processing (in a broad sense). His papers appeared at venues such as ACL, EMNLP, NAACL, AAAI, and CIKM. He served as a program committee member for conferences such as ACL, EMNLP, NAACL, EACL, AAAI, IJCAI and NIPS, and is currently a member of the standing reviewer team for TACL. He served as an area co-chair for ACL 2016 and received the best paper award at EMNLP 2011.

els and highlight its potential applications in some recent popular tasks (e.g., AMR parsing (Flanigan et al., 2014)).

The framework has been extensively used by our research group for developing various structured prediction models, including models for information extraction (Lu and Roth, 2015; Muis and Lu, 2016a; Jie et al., 2017), noun phrase chunking (Muis and Lu, 2016b), semantic parsing (Lu, 2015; Susanto and Lu, 2017), and sentiment analysis (Li and Lu, 2017). It is our hope that this tutorial will be helpful for many natural language processing researchers who are interested in designing their own structured prediction models rapidly. We also hope this tutorial allows researchers to strengthen their understandings on the connections between various structured prediction models, and that the open release of the framework will bring value to the NLP research community and enhance its overall productivity.

Tutorial 7

Cross-Lingual Word Representations: Induction and Evaluation

Manaal Faruqui, Anders Søgaard, and Ivan Vulić

Friday, September 8, 2017, 14:00–17:30

Location: Nørrebros Runddel

In recent past, NLP as a field has seen tremendous utility of distributional word vector representations as features in downstream tasks. The fact that these word vectors can be trained on unlabeled monolingual corpora of a language makes them an inexpensive resource in NLP. With the increasing use of monolingual word vectors, there is a need for word vectors that can be used as efficiently across multiple languages as monolingually. Therefore, learning bilingual and multilingual word embeddings/vectors is currently an important research topic. These vectors offer an elegant and language-pair independent way to represent content across different languages.

This tutorial aims to bring NLP researchers up to speed with the current techniques in cross-lingual word representation learning. We will first discuss how to induce cross-lingual word representations (covering both bilingual and multilingual ones) from various data types and resources (e.g., parallel data, comparable data, non-aligned monolingual data in different languages, dictionaries and theasuri, or, even, images, eye-tracking data). We will then discuss how to evaluate such representations, intrinsically and extrinsically. We will introduce researchers to state-of-the-art methods for constructing cross-lingual word representations and discuss their applicability in a broad range of downstream NLP applications.

We will deliver a detailed survey of the current methods, discuss best training and evaluation practices and use-cases, and provide links to publicly available implementations, datasets, and pre-trained models.

Manaal Faruqui is a research scientist at Google NYC currently working on industrial-scale NLP problems. Manaal received his PhD in the Language Technologies Institute at Carnegie Mellon University. He has worked on problems in the areas of representation learning, distributional semantics and multilingual learning. He has won one of the best paper awards at NAACL 2015. He organized the workshop on cross-lingual and multilingual models in NLP at NAACL 2016.

Anders Søgaard is a full professor of Computer Science (NLP and Machine Learning) at the University of Copenhagen. Anders is interested in transfer learning and has worked on semi-supervised learning, domain adaptation, and cross-language adaptation of NLP models. He is particularly interested in transferring models to very low-resource languages. He holds an ERC Starting Grant, as well as several grants from national research councils and private research foundations. He has won three best paper awards at major ACL conferences. He gave a tutorial on domain adaptation at COLING 2014.

Ivan Vulić is a research associate at the University of Cambridge. He received his PhD summa cum laude at KU Leuven in 2014. Ivan is interested in representation learning, distributional and multi-modal semantics in monolingual and multilingual contexts, and transfer learning for enabling cross-lingual NLP applications. His work has been published in top-tier ACL and IR conferences. He gave a tutorial on multilingual topic models at ECIR 2013 and WSDM 2014, and organized a Vision & Language workshop at EMNLP 2015.

4 Workshops

Note: almost all workshops are located in **CPH Conference** – please see the map with room locations in the front matter. Exceptions are Workshop 1 (WMT) and Workshop 10 (BEA), which take place in the main building and the conference hotel, respectively.

Thursday–Friday

Thursday

Friday

Workshop 1: Second Conference on Machine Translation (WMT)

Organizers: *Philipp Koehn and Barry Haddow*

Location: Funen

Thursday, September 7, 2017

8:45–9:00 **Opening Remarks**

9:00–10:30 **Session 1: Shared Tasks Overview Presentations I**

9:00–9:40 **Shared Task: News Translation**

- Findings of the 2017 Conference on Machine Translation (WMT17)
  Ondřej Bojar, Rajen Chatterjee, Christian Federmann, Yvette Graham, Barry Haddow, Matthias Huck, Philipp Koehn, Varvara Logacheva, Christof Monz, Matteo Negri, Matt Post, Raphael Rubino, Lucia Specia, and Marco Turchi

9:40–10:10 **Shared Task: Multimodal Translation**

- Findings of the Second Shared Task on Multimodal Machine Translation and Multilingual Image Description
  Desmond Elliott, Stella Frank, Loïc Barrault, Fethi Bougares, and Lucia Specia

10:10–10:30 **Shared Task: Biomedical Translation**

- Findings of the WMT 2017 Biomedical Translation Shared Task
  Antonio Jimeno Yepes, Aurelie Neveol, Mariana Neves, Karin Verspoor, Ondrej Bojar, Arthur Boyer, Cristian Grozea, Barry Haddow, Madeleine Kittner, Yvonne Lichtblau, Pavel Pecina, Roland Roller, Rudolf Rosa, Amy Siu, Philippe Thomas, and Saskia Trescher

10:30-11:00 **Coffee Break**

11:00–12:30 **Session 2: Shared Tasks Poster Session I**

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

- CUNI submission in WMT17: Chimera goes neural
  Roman Sudarikov, David Mareček, Tom Kocmi, Dusan Varis, and Ondřej Bojar
- LIMSIWMT'17
  Franck Burlot, Pooyan Safari, Matthieu Labeau, Alexandre Allauzen, and François Yvon
- SYSTRAN Purely Neural MT Engines for WMT2017
  Yongchao Deng, Jungi Kim, Guillaume Klein, Catherine Kobus, Natalia Segal, Christophe Servan, Bo Wang, Dakun Zhang, Josep Crego, and Jean Senellart
- FBK's Participation to the English-to-German News Translation Task of WMT 2017
  Mattia Antonino Di Gangi, Nicola Bertoldi, and Marcello Federico
- The JHU Machine Translation Systems for WMT 2017
  Shuoyang Ding, Huda Khayrallah, Philipp Koehn, Matt Post, Gaurav Kumar, and Kevin Duh

- The TALP-UPC Neural Machine Translation System for German/Finnish-English Using the Inverse Direction Model in Rescoring
  Carlos Escolano, Marta R. Costa-jussà, and José A. R. Fonollosa
- LIUM Machine Translation Systems for WMT17 News Translation Task
  Mercedes García-Martínez, Ozan Caglayan, Walid Aransa, Adrien Bardet, Fethi Bougares, and Loïc Barrault
- Extending hybrid word-character neural machine translation with multi-task learning of morphological analysis
  Stig-Arne Grönroos, Sami Virpioja, and Mikko Kurimo
- The AFRL-MITLL WMT17 Systems: Old, New, Borrowed, BLEU
  Jeremy Gwinnup, Timothy Anderson, Grant Erdmann, Katherine Young, Michaeel Kazi, Elizabeth Salesky, Brian Thompson, and Jonathan Taylor
- University of Rochester WMT 2017 NMT System Submission
  Chester Holtz, Chuyang Ke, and Daniel Gildea
- LMU Munich's Neural Machine Translation Systems for News Articles and Health Information Texts
  Matthias Huck, Fabienne Braune, and Alexander Fraser
- Rule-based Machine translation from English to Finnish
  Arvi Hurskainen and Jörg Tiedemann
- NRC Machine Translation System for WMT 2017
  Chi-kiu Lo, Boxing Chen, Colin Cherry, George Foster, Samuel Larkin, Darlene Stewart, and Roland Kuhn
- The Helsinki Neural Machine Translation System
  Robert Östling, Yves Scherrer, Jörg Tiedemann, Gongbo Tang, and Tommi Nieminen
- The QT21 Combined Machine Translation System for English to Latvian
  Jan-Thorsten Peter, Hermann Ney, Ondřej Bojar, Ngoc-Quan Pham, Jan Niehues, Alex Waibel, Franck Burlot, François Yvon, Mārcis Pinnis, Valters Sics, Joost Bastings, Miguel Rios, Wilker Aziz, Philip Williams, Frédéric Blain, and Lucia Specia
- The RWTH Aachen University English-German and German-English Machine Translation System for WMT 2017
  Jan-Thorsten Peter, Andreas Guta, Tamer Alkhouli, Parnia Bahar, Jan Rosendahl, Nick Rossenbach, Miguel Graça, and Hermann Ney
- The Karlsruhe Institute of Technology Systems for the News Translation Task in WMT 2017
  Ngoc-Quan Pham, Jan Niehues, Thanh-Le Ha, Eunah Cho, Matthias Sperber, and Alexander Waibel
- Tilde's Machine Translation Systems for WMT 2017
  Mārcis Pinnis, Rihards Krišlauks, Toms Miks, Daiga Deksne, and Valters Šics
- C-3MA: Tartu-Riga-Zurich Translation Systems for WMT17
  Matīss Rikters, Chantal Amrhein, Maksym Del, and Mark Fishel
- The University of Edinburgh's Neural MT Systems for WMT17
  Rico Sennrich, Alexandra Birch, Anna Currey, Ulrich Germann, Barry Haddow, Kenneth Heafield, Antonio Valerio Miceli Barone, and Philip Williams
- XMU Neural Machine Translation Systems for WMT 17
  Zhixing Tan, Boli Wang, Jinming Hu, and Yidong Chen
- The JAIST Machine Translation Systems for WMT 17
  Long Trieu, Trung-Tin Pham, and Le-Minh Nguyen
- Sogou Neural Machine Translation Systems for WMT17
  Yuguang Wang, Shanbo Cheng, Liyang Jiang, Jiajun Yang, Wei Chen, Muze Li, Lin Shi, Yanfeng Wang, and Hongtao Yang

- PJIIT's systems for WMT 2017 Conference
  Krzysztof Wolk and Krzysztof Marasek
- Hunter MT: A Course for Young Researchers in WMT17
  Jia Xu, Yi Zong Kuang, Shondell Baijoo, Jacob Hyun Lee, Uman Shahzad, Mir Ahmed, Meredith Lancaster, and Chris Carlan
- CASICT-DCU Neural Machine Translation Systems for WMT17
  Jinchao Zhang, Peerachet Porkaew, Jiawei Hu, Qiuye Zhao, and Qun Liu

11:00–12:30 **Shared Task: Multi-Modal Translation**

- LIUM-CVC Submissions for WMT17 Multimodal Translation Task
  Ozan Caglayan, Walid Aransa, Adrien Bardet, Mercedes García-Martínez, Fethi Bougares, Loïc Barrault, Marc Masana, Luis Herranz, and Joost van de Weijer
- DCU System Report on the WMT 2017 Multi-modal Machine Translation Task
  Iacer Calixto, Koel Dutta Chowdhury, and Qun Liu
- The AFRL-OSU WMT17 Multimodal Translation System: An Image Processing Approach
  Jeremy Gwinnup, John Duselis, Michael Hutt, James Davis, and Joshua Sandvick
- CUNI System for the WMT17 Multimodal Translation Task
  Jindřich Helcl and Jindřich Libovický
- Generating Image Descriptions using Multilingual Data
  Alan Jaffe
- OSU Multimodal Machine Translation System Report
  Mingbo Ma, Dapeng Li, Kai Zhao, and Liang Huang
- Sheffield MultiMT: Using Object Posterior Predictions for Multimodal Machine Translation
  Pranava Swaroop Madhyastha, Josiah Wang, and Lucia Specia
- NICT-NAIST System for WMT17 Multimodal Translation Task
  Jingyi Zhang, Masao Utiyama, Eiichro Sumita, Graham Neubig, and Satoshi Nakamura

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

- Automatic Threshold Detection for Data Selection in Machine Translation
  Mirela-Stefania Duma and Wolfgang Menzel

12:30–14:00 **Lunch**

14:00–15:30 **Session 3: Invited Talk**

14:00–15:30 **Holger Schwenk (Facebook): Multilingual Representions and Applications in NLP**

15:30–16:00 **Coffee Break**

16:00–17:30 **Session 4: Research Papers on Lexicon and Morphology**

16:00–16:15 Sense-Aware Statistical Machine Translation using Adaptive Context-Dependent Clustering
Xiao Pu, Nikolaos Pappas, and Andrei Popescu-Belis

16:15–16:30 Improving Word Sense Disambiguation in Neural Machine Translation with Sense Embeddings
Annette Rios Gonzales, Laura Mascarell, and Rico Sennrich

16:30–16:45 Word Representations in Factored Neural Machine Translation
Franck Burlot, Mercedes García-Martínez, Loïc Barrault, Fethi Bougares, and François Yvon

16:45–17:00 Modeling Target-Side Inflection in Neural Machine Translation
Aleš Tamchyna, Marion Weller-Di Marco, and Alexander Fraser

17:00–17:15 Evaluating the morphological competence of Machine Translation Systems
Franck Burlot and François Yvon

17:15–17:30 Target-side Word Segmentation Strategies for Neural Machine Translation
Matthias Huck, Simon Riess, and Alexander Fraser

Friday, September 8, 2017

9:00–10:30 **Session 5: Shared Tasks Overview Presentations II**

9:00–9:20 **Shared Task: Quality Estimation**

9:20–9:40 **Shared Task: Metrics**

- Results of the WMT17 Metrics Shared Task
Ondřej Bojar, Yvette Graham, and Amir Kamran

9:40–10:00 **Shared Task: Automatic Post-Editing**

10:00–10:15 **Shared Task: Bandit Learning**

- A Shared Task on Bandit Learning for Machine Translation
Artem Sokolov, Julia Kreutzer, Kellen Sunderland, Pavel Danchenko, Witold Szymaniak, Hagen Fürstenau, and Stefan Riezler

10:15–10:30 **Shared Task: Neural Training**

- Results of the WMT17 Neural MT Training Task
Ondřej Bojar, Jindřich Helcl, Tom Kocmi, Jindřich Libovický, and Tomáš Musil

10:30-11:00 **Coffee Break**

11:00–12:30 **Session 6: Shared Tasks Poster Session II**

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

- Sentence-level quality estimation by predicting HTER as a multi-component metric
Eleftherios Avramidis
- Predicting Translation Performance with Referential Translation Machines
Ergun Biçici
- Bilexical Embeddings for Quality Estimation
Frédéric Blain, Carolina Scarton, and Lucia Specia
- Improving Machine Translation Quality Estimation with Neural Network Features
Zhiming Chen, Yiming Tan, Chenlin Zhang, Qingyu Xiang, Lilin Zhang, Maoxi Li, and Mingwen Wang
- UHH Submission to the WMT17 Quality Estimation Shared Task
Melania Duma and Wolfgang Menzel

- Predictor-Estimator using Multilevel Task Learning with Stack Propagation for Neural Quality Estimation
  Hyun Kim, Jong-Hyeok Lee, and Seung-Hoon Na
- Unbabel's Participation in the WMT17 Translation Quality Estimation Shared Task
  André F. T. Martins, Fabio Kepler, and Jose Monteiro
- Feature-Enriched Character-Level Convolutions for Text Regression
  Gustavo Paetzold and Lucia Specia

11:00–12:30 **Shared Task: Metrics**

- UHH Submission to the WMT17 Metrics Shared Task
  Melania Duma and Wolfgang Menzel
- MEANT 2.0: Accurate semantic MT evaluation for any output language
  Chi-kiu Lo
- Blend: a Novel Combined MT Metric Based on Direct Assessment — CASICT-DCU submission to WMT17 Metrics Task
  Qingsong Ma, Yvette Graham, Shugen Wang, and Qun Liu
- CUNI Experiments for WMT17 Metrics Task
  David Mareček, Ondřej Bojar, Ondřej Hübsch, Rudolf Rosa, and Dusan Varis
- chrF++: words helping character n-grams
  Maja Popović
- bleu2vec: the Painfully Familiar Metric on Continuous Vector Space Steroids
  Andre Tättar and Mark Fishel

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

- LIG-CRIStAL Submission for the WMT 2017 Automatic Post-Editing Task
  Alexandre Berard, Laurent Besacier, and Olivier Pietquin
- Multi-source Neural Automatic Post-Editing: FBK's participation in the WMT 2017 APE shared task
  Rajen Chatterjee, M. Amin Farajian, Matteo Negri, Marco Turchi, Ankit Srivastava, and Santanu Pal
- The AMU-UEdin Submission to the WMT 2017 Shared Task on Automatic Post-Editing
  Marcin Junczys-Dowmunt and Marcin Junczys-Dowmunt
- Ensembling Factored Neural Machine Translation Models for Automatic Post-Editing and Quality Estimation
  Chris Hokamp
- Neural Post-Editing Based on Quality Estimation
  Yiming Tan, Zhiming Chen, Liu Huang, Lilin Zhang, Maoxi Li, and Mingwen Wang
- CUNI System for WMT17 Automatic Post-Editing Task
  Dusan Varis and Ondřej Bojar

11:00–12:30 **Shared Task: Bandit Learning**

- The UMD Neural Machine Translation Systems at WMT17 Bandit Learning Task
  Amr Sharaf, Shi Feng, Khanh Nguyen, Kiante Brantley, and Hal Daumé III
- LIMSI Submission for WMT'17 Shared Task on Bandit Learning
  Guillaume Wisniewski

11:00–12:30 **Shared Task: Neural Training**

- Variable Mini-Batch Sizing and Pre-Trained Embeddings
  Mostafa Abdou, Vladan Gloncak, and Ondřej Bojar

- The AFRL WMT17 Neural Machine Translation Training Task Submission
  Jeremy Gwinnup, Grant Erdmann, and Katherine Young

12:30–14:00 **Lunch**

14:00–15:15 **Session 7: Research Papers on Syntax and Deep Models**

14:00–14:15 Predicting Target Language CCG Supertags Improves Neural Machine Translation
Maria Nadejde, Siva Reddy, Rico Sennrich, Tomasz Dwojak, Marcin Junczys-Dowmunt, Philipp Koehn, and Alexandra Birch

14:15–14:30 Exploiting Linguistic Resources for Neural Machine Translation Using Multi-task Learning
Jan Niehues and Eunah Cho

14:30–14:45 Tree as a Pivot: Syntactic Matching Methods in Pivot Translation
Akiva Miura, Graham Neubig, Katsuhito Sudoh, and Satoshi Nakamura

14:45–15:00 Deep architectures for Neural Machine Translation
Antonio Valerio Miceli Barone, Jindřich Helcl, Rico Sennrich, Barry Haddow, and Alexandra Birch

15:00–15:15 Biasing Attention-Based Recurrent Neural Networks Using External Alignment Information
Tamer Alkhouli and Hermann Ney

15:15–16:00 **Coffee Break**

16:00–17:15 **Session 8: Research Papers on Domain Adaptation and External Data**

16:00–16:15 Effective Domain Mixing for Neural Machine Translation
Denny Britz, Quoc Le, and Reid Pryzant

16:15–16:30 Multi-Domain Neural Machine Translation through Unsupervised Adaptation
M. Amin Farajian, Marco Turchi, Matteo Negri, and Marcello Federico

16:30–16:45 Adapting Neural Machine Translation with Parallel Synthetic Data
Mara Chinea-Rios, Álvaro Peris, and Francisco Casacuberta

16:45–17:00 Copied Monolingual Data Improves Low-Resource Neural Machine Translation
Anna Currey, Antonio Valerio Miceli Barone, and Kenneth Heafield

17:00–17:15 Guiding Neural Machine Translation Decoding with External Knowledge
Rajen Chatterjee, Matteo Negri, Marco Turchi, Marcello Federico, Lucia Specia, and Frédéric Blain

Workshop 2: Subword and Character Level Models in NLP

Organizers: *Manaal Faruqui, Hinrich Schütze, Isabel Trancoso, and Yadollah Yaghoobzadeh*

Location: Nørrebros Runddel

Thursday, September 7, 2017

09:00–09:10 **Opening Remarks (Manaal Faruqui)**

09:10–09:50 **Invited Talk: Subword-level Information in NLP using Neural Networks (Tomas Mikolov)**

09:50–10:30 **Invited Talk: Chewing the Fat about Mincing Words (Noah Smith)**

10:30–11:00 **Coffee break**

11:00–11:40 **Invited Tutorial Talk: Neural WFSTs (Ryan Cotterell)**

11:40–12:10 **Best paper presentations**

12:10–14:00 **Poster session & Lunch break**

- Character and Subword-Based Word Representation for Neural Language Modeling Prediction
  Matthieu Labeau and Alexandre Allauzen
- Learning variable length units for SMT between related languages via Byte Pair Encoding
  Anoop Kunchukuttan and Pushpak Bhattacharyya
- Character Based Pattern Mining for Neology Detection
  Gaël Lejeune and Emmanuel Cartier
- Automated Word Stress Detection in Russian
  Maria Ponomareva, Kirill Milintsevich, Ekaterina Chernyak, and Anatoly Starostin
- A Syllable-based Technique for Word Embeddings of Korean Words
  Sanghyuk Choi, Taeuk Kim, Jinseok Seol, and Sang-goo Lee
- Supersense Tagging with a Combination of Character, Subword, and Word-level Representations
  Youhyun Shin and Sang-goo Lee
- Weakly supervised learning of allomorphy
  Miikka Silfverberg and Mans Hulden
- Character-based recurrent neural networks for morphological relational reasoning
  Olof Mogren and Richard Johansson
- Glyph-aware Embedding of Chinese Characters
  Falcon Dai and Zheng Cai
- Exploring Cross-Lingual Transfer of Morphological Knowledge In Sequence-to-Sequence Models
  Huiming Jin and Katharina Kann
- (EXTENDED ABSTRACT) Language Generation with Recurrent Generative Adversarial Networks without Pre-training
  Ofir Press, Amir Bar, Ben Bogin, Jonathan Berant, and Lior Wolf

- (EXTENDED ABSTRACT) Align and Copy: Hard Attention Models for Morphological Inflection Generation
  Tatyana Ruzsics, Peter Makarov, and Simon Clematide
- (EXTENDED ABSTRACT) Natural Language Generation through Character-Based RNNs with Finite-State Prior Knowledge
  Raghav Goyal, Marc Dymetman, and Eric Gaussier
- (EXTENDED ABSTRACT) Patterns versus Characters in Subword-aware Neural Language Modeling
  Zhenisbek Assylbekov and Rustem Takhanov

14:00–14:40 **Invited Talk: Fully Character Level Neural Machine Translation (Kyunghyun Cho)**

14:40–15:50 **Poster session & Coffee break**

- Unlabeled Data for Morphological Generation With Character-Based Sequence-to-Sequence Models
  Katharina Kann and Hinrich Schütze
- Vowel and Consonant Classification through Spectral Decomposition
  Patricia Thaine and Gerald Penn
- Syllable-level Neural Language Model for Agglutinative Language
  Seunghak Yu, Nilesh Kulkarni, Haejun Lee, and Jihie Kim
- Character-based Bidirectional LSTM-CRF with words and characters for Japanese Named Entity Recognition
  Shotaro Misawa, Motoki Taniguchi, Yasuhide Miura, and Tomoko Ohkuma
- Word Representation Models for Morphologically Rich Languages in Neural Machine Translation
  Ekaterina Vylomova, Trevor Cohn, Xuanli He, and Gholamreza Haffari
- Spell-Checking based on Syllabification and Character-level Graphs for a Peruvian Agglutinative Language
  Carlo Alva and Arturo Oncevay
- What do we need to know about an unknown word when parsing German
  Bich-Ngoc Do, Ines Rehbein, and Anette Frank
- A General-Purpose Tagger with Convolutional Neural Networks
  Xiang Yu, Agnieszka Falenska, and Ngoc Thang Vu
- Reconstruction of Word Embeddings from Sub-Word Parameters
  Karl Stratos
- Inflection Generation for Spanish Verbs using Supervised Learning
  Cristina Barros, Dimitra Gkatzia, and Elena Lloret
- Neural Paraphrase Identification of Questions with Noisy Pretraining
  Gaurav Singh Tomar, Thyago Duque, Oscar Täckström, Jakob Uszkoreit, and Dipanjan Das
- Sub-character Neural Language Modelling in Japanese
  Viet Nguyen, Julian Brooke, and Timothy Baldwin
- Byte-based Neural Machine Translation
  Marta R. Costa-jussà, Carlos Escolano, and José A. R. Fonollosa
- Improving Opinion-Target Extraction with Character-Level Word Embeddings
  Soufian Jebbara and Philipp Cimiano

15:50–16:30 **Invited Talk: Acoustic Word Embeddings (Karen Livescu)**

16:30–17:30 **Panel discussion**

17:30–17:45 Closing remarks (Hinrich Schütze)

Workshop 3: Natural Language Processing meets Journalism

Organizers: *Octavian Popescu and Carlo Strapparava*

Location: Amager Strandpark

Thursday, September 7, 2017

Morning

Oral Presentations

- Predicting News Values from Headline Text and Emotions
  Maria Pia di Buono, Jan Šnajder, Bojana Dalbelo Basic, Goran Glavaš, Martin Tutek, and Natasa Milic-Frayling
- Predicting User Views in Online News
  Daniel Hardt and Owen Rambow
- Tracking Bias in News Sources Using Social Media: the Russia-Ukraine Maidan Crisis of 2013–2014
  Peter Potash, Alexey Romanov, Mikhail Gronas, Anna Rumshisky, and Mikhail Gronas
- What to Write? A topic recommender for journalists
  Alessandro Cucchiarelli, Christian Morbidoni, Giovanni Stilo, and Paola Velardi
- Comparing Attitudes to Climate Change in the Media using sentiment analysis based on Latent Dirichlet Allocation
  Ye Jiang, Xingyi Song, Jackie Harrison, Shaun Quegan, and Diana Maynard
- Language-based Construction of Explorable News Graphs for Journalists
  Rémi Bois, Guillaume Gravier, Eric Jamet, Emmanuel Morin, Pascale Sébillot, and Maxime Robert
- Storyteller: Visual Analytics of Perspectives on Rich Text Interpretations
  Maarten van Meersbergen, Piek Vossen, Janneke van der Zwaan, Antske Fokkens, Willem van Hage, Inger Leemans, and Isa Maks
- Analyzing the Revision Logs of a Japanese Newspaper for Article Quality Assessment
  Hideaki Tamori, Yuta Hitomi, Naoaki Okazaki, and Kentaro Inui
- Improved Abusive Comment Moderation with User Embeddings
  John Pavlopoulos, Prodromos Malakasiotis, Juli Bakagianni, and Ion Androutsopoulos

Lunch

Poster Presentations

- Incongruent Headlines: Yet Another Way to Mislead Your Readers
  Sophie Chesney, Maria Liakata, Massimo Poesio, and Matthew Purver
- Unsupervised Event Clustering and Aggregation from Newswire and Web Articles
  Swen Ribeiro, Olivier Ferret, and Xavier Tannier
- Semantic Storytelling, Cross-lingual Event Detection and other Semantic Services for a Newsroom Content Curation Dashboard
  Julian Moreno-Schneider, Ankit Srivastava, Peter Bourgonje, David Wabnitz, and Georg Rehm

- Deception Detection in News Reports in the Russian Language: Lexics and Discourse
  Dina Pisarevskaya
- Fake news stance detection using stacked ensemble of classifiers
  James Thorne, Mingjie Chen, Giorgos Myrianthous, Jiashu Pu, Xiaoxuan Wang, and Andreas Vlachos
- From Clickbait to Fake News Detection: An Approach based on Detecting the Stance of Headlines to Articles
  Peter Bourgonje, Julian Moreno Schneider, and Georg Rehm
- 'Fighting' or 'Conflict'? An Approach to Revealing Concepts of Terms in Political Discourse
  Linyuan Tang and Kyo Kageura
- A News Chain Evaluation Methodology along with a Lattice-based Approach for News Chain Construction
  Mustafa Toprak, Özer Özkahraman, and Selma Tekir
- Using New York Times Picks to Identify Constructive Comments
  Varada Kolhatkar and Maite Taboada
- An NLP Analysis of Exaggerated Claims in Science News
  Yingya Li, Jieke Zhang, and Bei Yu

Best paper announcement and Conclusions

Please note: This workshop's schedule was not finalized by the time of printing. Please see `http://nlpj2017.fbk.eu/program` for the full program.

Workshop 4: 3rd Workshop on Noisy User-generated Text

Organizers: *Leon Derczynski, Wei Xu, Alan Ritter, and Timothy Baldwin*

Location: Tivoli & Vesterbro Torv

Thursday, September 7, 2017

9:00–9:05 **Opening**

9:05–9:50 **Invited Talk: Common Sense Knowledge as an Emergent Property of Neural Conversational Models (Bill Dolan)**

9:50–10:35 **Oral Session I**

9:50–10:05 Boundary-based MWE segmentation with text partitioning
Jake Williams

10:05–10:20 Towards the Understanding of Gaming Audiences by Modeling Twitch Emotes
Francesco Barbieri, Luis Espinosa Anke, Miguel Ballesteros, Juan Soler, and Horacio Saggion

10:20–10:35 Churn Identification in Microblogs using Convolutional Neural Networks with Structured Logical Knowledge
Mourad Gridach, Hatem Haddad, and Hala Mulki

10:35–11:00 **Coffee Break**

11:00–12:30 **Oral Session II**

11:00–11:15 To normalize, or not to normalize: The impact of normalization on Part-of-Speech tagging
Rob van der Goot, Barbara Plank, and Malvina Nissim

11:15–11:30 Constructing an Alias List for Named Entities during an Event
Anietie Andy, Mark Dredze, Mugizi Rwebangira, and Chris Callison-Burch

11:30–11:45 Incorporating Metadata into Content-Based User Embeddings
Linzi Xing and Michael J. Paul

11:45–12:00 Simple Queries as Distant Labels for Predicting Gender on Twitter
Chris Emmery, Grzegorz Chrupała, and Walter Daelemans

12:00–12:15 A Dataset and Classifier for Recognizing Social Media English
Su Lin Blodgett, Johnny Wei, and Brendan O'Connor

12:15–12:30 Evaluating hypotheses in geolocation on a very large sample of Twitter
Bahar Salehi and Anders Søgaard

12:30–14:00 **Lunch**

14:00–14:45 **Invited Talk: Tweets in Finance (Miles Osborne)**

14:45–14:55 **Lightning Talks**

- The Effect of Error Rate in Artificially Generated Data for Automatic Preposition and Determiner Correction
Fraser Bowen, Jon Dehdari, and Josef Van Genabith
- An Entity Resolution Approach to Isolate Instances of Human Trafficking Online
Chirag Nagpal, Kyle Miller, Benedikt Boecking, and Artur Dubrawski

- Noisy Uyghur Text Normalization
  Osman Tursun and Ruket Cakici
- Crowdsourcing Multiple Choice Science Questions
  Johannes Welbl, Nelson F. Liu, and Matt Gardner
- A Text Normalisation System for Non-Standard English Words
  Emma Flint, Elliot Ford, Olivia Thomas, Andrew Caines, and Paula Buttery
- Huntsville, hospitals, and hockey teams: Names can reveal your location
  Bahar Salehi, Dirk Hovy, Eduard Hovy, and Anders Søgaard
- Improving Document Clustering by Removing Unnatural Language
  Myungha Jang, Jinho D. Choi, and James Allan
- Lithium NLP: A System for Rich Information Extraction from Noisy User Generated Text on Social Media
  Preeti Bhargava, Nemanja Spasojevic, and Guoning Hu

14:55–15:30 **Shared Task Session**

14:55–15:10 Results of the WNUT2017 Shared Task on Novel and Emerging Entity Recognition
Leon Derczynski, Eric Nichols, Marieke van Erp, and Nut Limsopatham

15:10–15:20 A Multi-task Approach for Named Entity Recognition in Social Media Data
Gustavo Aguilar, Suraj Maharjan, Adrian Pastor López Monroy, and Thamar Solorio

15:20–15:30 Distributed Representation, LDA Topic Modelling and Deep Learning for Emerging Named Entity Recognition from Social Media
Patrick Jansson and Shuhua Liu

- Multi-channel BiLSTM-CRF Model for Emerging Named Entity Recognition in Social Media
  Bill Y. Lin, Frank Xu, Zhiyi Luo, and Kenny Zhu
- Transfer Learning and Sentence Level Features for Named Entity Recognition on Tweets
  Pius von Däniken and Mark Cieliebak
- Context-Sensitive Recognition for Emerging and Rare Entities
  Jake Williams and Giovanni Santia
- A Feature-based Ensemble Approach to Recognition of Emerging and Rare Named Entities
  Utpal Kumar Sikdar and Björn Gambäck

15:30–16:30 **Poster Session**

16:30–17:15 **Invited Talk: Modeling Language as a Social Construct (Dirk Hovy)**

17:15–17:30 **Closing and Best Paper Awards**

Workshop 5: 2nd Workshop on Structured Prediction for Natural Language Processing

Organizers: *Kai-Wei Chang, Ming-Wei Chang, Vivek Srikumar, and Alexander M. Rush*

Location: Hovedbanen

Thursday, September 7, 2017

9:00–10:30 **Section 1**

9:00–9:15 **Welcome (Organizers)**

9:15–10:00 **Invited Talk**

10:00–10:30 Dependency Parsing with Dilated Iterated Graph CNNs
Emma Strubell and Andrew McCallum

10:30–11:00 **Coffee Break**

11:00–12:15 **Section 2**

11:00–11:45 **Invited Talk**

11:45–12:15 **Poster Madness**

12:15–14:00 **Lunch**

14:00–15:30 **Section 3**

14:00–14:45 **Poster Session**

- Entity Identification as Multitasking
Karl Stratos
- Towards Neural Machine Translation with Latent Tree Attention
James Bradbury and Richard Socher
- Structured Prediction via Learning to Search under Bandit Feedback
Amr Sharaf and Hal Daumé III
- Syntax Aware LSTM model for Semantic Role Labeling
Feng Qian, Lei Sha, Baobao Chang, LuChen Liu, and Ming Zhang
- Spatial Language Understanding with Multimodal Graphs using Declarative Learning based Programming
Parisa Kordjamshidi, Taher Rahgooy, and Umar Manzoor
- Boosting Information Extraction Systems with Character-level Neural Networks and Free Noisy Supervision
Philipp Meerkamp and Zhengyi Zhou

14:45–15:30 **Invited Talk**

15:30–16:00 **Coffee Break**

16:00–17:30 **Section 4**

16:00–16:45 **Invited Talk**

16:45–17:15 Piecewise Latent Variables for Neural Variational Text Processing
Iulian Vlad Serban, Alexander Ororbia II, Joelle Pineau, and Aaron Courville

17:15–17:30 **Closing**

Workshop 6: New Frontiers in Summarization

Organizers: *Lu Wang, Jackie Chi Kit Cheung, Giuseppe Carenini, and Fei Liu*

Location: Enghave Plads & Kødbyen

Thursday, September 7, 2017

08:45–10:30 **Morning Session 1**

08:45–08:50 **Opening Remarks**

08:50–09:50 **Invited Talk (Andreas Kerren)**

09:50–10:10 Video Highlights Detection and Summarization with Lag-Calibration based on Concept-Emotion Mapping of Crowdsourced Time-Sync Comments
Qing Ping and Chaomei Chen

10:10–10:30 Multimedia Summary Generation from Online Conversations: Current Approaches and Future Directions
Enamul Hoque and Giuseppe Carenini

10:30–11:00 **Break**

11:00–12:30 **Morning Session 2**

11:00–12:00 **Invited Talk (Katja Filippova)**

12:00–12:15 Low-Resource Neural Headline Generation
Ottokar Tilk and Tanel Alumäe

12:15–12:30 Towards Improving Abstractive Summarization via Entailment Generation
Ramakanth Pasunuru, Han Guo, and Mohit Bansal

12:30–14:00 **Lunch**

14:00–15:30 **Poster Session**

- Coarse-to-Fine Attention Models for Document Summarization
Jeffrey Ling and Alexander Rush
- Automatic Community Creation for Abstractive Spoken Conversations Summarization
Karan Singla, Evgeny Stepanov, Ali Orkan Bayer, Giuseppe Carenini, and Giuseppe Riccardi
- Combining Graph Degeneracy and Submodularity for Unsupervised Extractive Summarization
Antoine Tixier, Polykarpos Meladianos, and Michalis Vazirgiannis
- TL;DR: Mining Reddit to Learn Automatic Summarization
Michael Völske, Martin Potthast, Shahbaz Syed, and Benno Stein
- Topic Model Stability for Hierarchical Summarization
John Miller and Kathleen McCoy
- Learning to Score System Summaries for Better Content Selection Evaluation.
Maxime Peyrard, Teresa Botschen, and Iryna Gurevych

- Revisiting the Centroid-based Method: A Strong Baseline for Multi-Document Summarization
  Demian Gholipour Ghalandari
- Reader-Aware Multi-Document Summarization: An Enhanced Model and The First Dataset
  Piji Li, Lidong Bing, and Wai Lam
- A Pilot Study of Domain Adaptation Effect for Neural Abstractive Summarization
  Xinyu Hua and Lu Wang

15:30–17:15 **Afternoon Session**

15:30–16:30 **Invited Talk (Ani Nenkova)**

17:10–17:15 **Closing Remarks**

Workshop 7: Workshop on Speech-Centric Natural Language Processing

Organizers: *Nicholas Ruiz and Srinivas Bangalore*

Location: Kastrup Airport

Thursday, September 7, 2017

8:50–9:00 **Opening Remarks (Nicholas Ruiz and Srinivas Bangalore)**

9:00–10:00 **Invited Talk. Modelling turn-taking in spoken interaction (Gabriel Skantze)**

10:00–10:30 **Session I**

- Functions of Silences towards Information Flow in Spoken Conversation
  Shammur Absar Chowdhury, Evgeny Stepanov, Morena Danieli, and Giuseppe Riccardi

10:30–11:00 **Coffee Break**

11:00–12:30 **Session II**

- Encoding Word Confusion Networks with Recurrent Neural Networks for Dialog State Tracking
  Glorianna Jagfeld and Ngoc Thang Vu
- Analyzing Human and Machine Performance In Resolving Ambiguous Spoken Sentences
  Hussein Ghaly and Michael Mandel
- Parsing transcripts of speech
  Andrew Caines, Michael McCarthy, and Paula Buttery
- Enriching ASR Lattices with POS Tags for Dependency Parsing
  Moritz Stiefel and Ngoc Thang Vu

12:30–14:00 **Lunch**

14:00–15:30 **Session III**

- End-to-End Information Extraction without Token-Level Supervision
  Rasmus Berg Palm, Dirk Hovy, Florian Laws, and Ole Winther
- Spoken Term Discovery for Language Documentation using Translations
  Antonios Anastasopoulos, Sameer Bansal, David Chiang, Sharon Goldwater, and Adam Lopez
- Amharic-English Speech Translation in Tourism Domain
  Michael Melese, Laurent Besacier, and Million Meshesha
- Speech- and Text-driven Features for Automated Scoring of English Speaking Tasks
  Anastassia Loukina, Nitin Madnani, and Aoife Cahill

15:30–16:00 **Coffee Break / Poster Discussion**

16:00–16:25 **Session IV**

- Improving coreference resolution with automatically predicted prosodic information
  Ina Roesiger, Sabrina Stehwien, Arndt Riester, and Ngoc Thang Vu

16:25–17:50 **Round-table: Issues in Speech-centric NLP**

17:50–18:00 **Closing**

Workshop 8: 3rd Workshop on Discourse in Machine Translation

Organizers: *Bonnie Webber, Andrei Popescu-Belis, and Jörg Tiedemann*

Location: Tivoli & Vesterbro Torv

Friday, September 8, 2017

09:00–10:30 **Session 1**

09:00–09:10 **Introduction**

09:10–09:40 Findings of the 2017 DiscoMT Shared Task on Cross-lingual Pronoun Prediction
Sharid Loáiciga, Sara Stymne, Preslav Nakov, Christian Hardmeier, Jörg Tiedemann, Mauro Cettolo, and Yannick Versley

09:40–10:10 Validation of an Automatic Metric for the Accuracy of Pronoun Translation (APT)
Lesly Miculicich Werlen and Andrei Popescu-Belis

10:10–10:30 **Poster Boaster**

10:30–11:00 **Coffee Break**

11:00–12:30 **Session 2a: Regular Track Posters**

- Using a Graph-based Coherence Model in Document-Level Machine Translation
  Leo Born, Mohsen Mesgar, and Michael Strube
- Treatment of Markup in Statistical Machine Translation
  Mathias Müller

11:00–12:30 **Session 2b: Shared Task Posters**

- A BiLSTM-based System for Cross-lingual Pronoun Prediction
  Sara Stymne, Sharid Loáiciga, and Fabienne Cap
- Neural Machine Translation for Cross-Lingual Pronoun Prediction
  Sébastien Jean, Stanislas Lauly, Orhan Firat, and Kyunghyun Cho
- Predicting Pronouns with a Convolutional Network and an N-gram Model
  Christian Hardmeier
- Cross-Lingual Pronoun Prediction with Deep Recurrent Neural Networks v2.0
  Juhani Luotolahti, Jenna Kanerva, and Filip Ginter

11:00–12:30 **Session 2c: Posters Related to Oral Presentations**

- Combining the output of two coreference resolution systems for two source languages to improve annotation projection
  Yulia Grishina
- Discovery of Discourse-Related Language Contrasts through Alignment Discrepancies in English-German Translation
  Ekaterina Lapshinova-Koltunski and Christian Hardmeier
- Neural Machine Translation with Extended Context
  Jörg Tiedemann and Yves Scherrer

- Translating Implicit Discourse Connectives Based on Cross-lingual Annotation and Alignment
  Hongzheng Li, Philippe Langlais, and Yaohong Jin

12:30–14:00 **Lunch Break**

14:00–15:30 **Session 3**

14:00–14:30 Neural Machine Translation with Extended Context
Jörg Tiedemann and Yves Scherrer

14:30–14:50 Discovery of Discourse-Related Language Contrasts through Alignment Discrepancies in English-German Translation
Ekaterina Lapshinova-Koltunski and Christian Hardmeier

14:50–15:10 Translating Implicit Discourse Connectives Based on Cross-lingual Annotation and Alignment
Hongzheng Li, Philippe Langlais, and Yaohong Jin

15:10–15:30 Combining the output of two coreference resolution systems for two source languages to improve annotation projection
Yulia Grishina

15:30–16:00 **Coffee Break**

16:00–17:30 **Session 4**

16:00–16:30 Lexical Chains meet Word Embeddings in Document-level Statistical Machine Translation
Laura Mascarell

16:30–16:50 On Integrating Discourse in Machine Translation
Karin Sim Smith

16:50–17:30 **Final Discussion and Conclusion**

Workshop 9: Workshop on Stylistic Variation

Organizers: *Julian Brooke, Thamar Solorio, and Moshe Koppel*

Location: Amager Strandpark

Friday, September 8, 2017

9:00–9:10 **Opening remarks (Julian Brooke, Thamar Solorio, and Moshe Koppel)**

9:10–10:00 **Invited Talk: Style Analysis for Practical Semantic Interpretation of Text (Ani Nenkova)**

10:00–10:30 From Shakespeare to Twitter: What are Language Styles all about?
Wei Xu

10:30–11:00 **Coffee Break**

11:00–12:30 **Technical Papers I**

11:00–11:30 Shakespearizing Modern Language Using Copy-Enriched Sequence to Sequence Models
Harsh Jhamtani, Varun Gangal, Eduard Hovy, and Eric Nyberg

11:30–12:00 Discovering Stylistic Variations in Distributional Vector Space Models via Lexical Paraphrases
Xing Niu and Marine Carpuat

12:00–12:30 Harvesting Creative Templates for Generating Stylistically Varied Restaurant Reviews
Shereen Oraby, Sheideh Homayon, and Marilyn Walker

12:30–14:00 **Lunch**

14:00–14:50 **Invited Talk: Problems in Personality Profiling (Walter Daelemans)**

14:50–15:30 **Poster Session**

- Is writing style predictive of scientific fraud?
Chloé Braud and Anders Søgaard
- "Deep" Learning : Detecting Metaphoricity in Adjective-Noun Pairs
Yuri Bizzoni, Stergios Chatzikyriakidis, and Mehdi Ghanimifard
- Authorship Attribution with Convolutional Neural Networks and POS-Eliding
Julian Hitschler, Esther van den Berg, and Ines Rehbein
- Topic and audience effects on distinctively Scottish vocabulary usage in Twitter data
Philippa Shoemark, James Kirby, and Sharon Goldwater
- Differences in type-token ratio and part-of-speech frequencies in male and female Russian written texts
Tatiana Litvinova, Pavel Seredin, Olga Litvinova, and Olga Zagorovskaya
- Modeling Communicative Purpose with Functional Style: Corpus and Features for German Genre and Register Analysis
Thomas Haider and Alexis Palmer
- Stylistic Variation in Television Dialogue for Natural Language Generation
Grace Lin and Marilyn Walker

15:30–16:00 **Coffee Break**

16:00–17:30 **Technical Papers II**

16:00–16:30 Controlling Linguistic Style Aspects in Neural Language Generation
Jessica Ficler and Yoav Goldberg

16:30–17:00 Approximating Style by N-gram-based Annotation
Melanie Andresen and Heike Zinsmeister

17:00–17:30 Assessing the Stylistic Properties of Neurally Generated Text in Authorship Attribution
Enrique Manjavacas, Jeroen De Gussem, Walter Daelemans, and Mike Kestemont

17:30–17:35 **Closing Remarks (Julian Brooke, Thamar Solorio, and Moshe Koppel)**

Workshop 10: 12th Workshop on Innovative Use of NLP for Building Educational Applications

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

Location: DGI Byen Hotel, Room 3

Friday, September 8, 2017

08:45–09:00 **Load Oral Presentations**

09:00–10:30 **Session 1**

09:00–09:15 **Opening Remarks**

09:15–09:40 Question Difficulty – How to Estimate Without Norming, How to Use for Automated Grading
Ulrike Pado

09:40–10:05 Combining CNNs and Pattern Matching for Question Interpretation in a Virtual Patient Dialogue System
Lifeng Jin, Michael White, Evan Jaffe, Laura Zimmerman, and Douglas Danforth

10:05–10:30 Continuous fluency tracking and the challenges of varying text complexity
Beata Beigman Klebanov, Anastassia Loukina, John Sabatini, and Tenaha O'Reilly

10:30–11:00 **Break**

11:00–12:35 **Session 2**

11:00–11:25 Auxiliary Objectives for Neural Error Detection Models
Marek Rei and Helen Yannakoudakis

11:25–11:50 Linked Data for Language-Learning Applications
Robyn Loughnane, Kate McCurdy, Peter Kolb, and Stefan Selent

11:50–12:10 Predicting Specificity in Classroom Discussion
Luca Lugini and Diane Litman

12:10–12:35 A Report on the 2017 Native Language Identification Shared Task
Shervin Malmasi, Keelan Evanini, Aoife Cahill, Joel Tetreault, Robert Pugh, Christopher Hamill, Diane Napolitano, and Yao Qian

12:35–14:00 **Lunch**

14:00–15:30 **Poster Session**

14:00–14:45 **Poster Session A**

- Evaluation of Automatically Generated Pronoun Reference Questions
Arief Yudha Satria and Takenobu Tokunaga
- Predicting Audience's Laughter During Presentations Using Convolutional Neural Network
Lei Chen and Chong Min Lee
- Collecting fluency corrections for spoken learner English
Andrew Caines, Emma Flint, and Paula Buttery

- Exploring Relationships Between Writing & Broader Outcomes With Automated Writing Evaluation
  Jill Burstein, Dan McCaffrey, Beata Beigman Klebanov, and Guangming Ling
- An Investigation into the Pedagogical Features of Documents
  Emily Sheng, Prem Natarajan, Jonathan Gordon, and Gully Burns
- Combining Multiple Corpora for Readability Assessment for People with Cognitive Disabilities
  Victoria Yaneva, Constantin Orasan, Richard Evans, and Omid Rohanian
- Automatic Extraction of High-Quality Example Sentences for Word Learning Using a Determinantal Point Process
  Arseny Tolmachev and Sadao Kurohashi
- Distractor Generation for Chinese Fill-in-the-blank Items
  Shu Jiang and John Lee
- An Error-Oriented Approach to Word Embedding Pre-Training
  Youmna Farag, Marek Rei, and Ted Briscoe
- Investigating neural architectures for short answer scoring
  Brian Riordan, Andrea Horbach, Aoife Cahill, Torsten Zesch, and Chong Min Lee
- Human and Automated CEFR-based Grading of Short Answers
  Anaïs Tack, Thomas François, Sophie Roekhaut, and Cédrick Fairon
- GEC into the future: Where are we going and how do we get there?
  Keisuke Sakaguchi, Courtney Napoles, and Joel Tetreault
- Detecting Off-topic Responses to Visual Prompts
  Marek Rei
- Combining Textual and Speech Features in the NLI Task Using State-of-the-Art Machine Learning Techniques
  Pavel Ircing, Jan Svec, Zbynek Zajic, Barbora Hladka, and Martin Holub
- Native Language Identification Using a Mixture of Character and Word N-grams
  Elham Mohammadi, Hadi Veisi, and Hessam Amini
- Ensemble Methods for Native Language Identification
  Sophia Chan, Maryam Honari Jahromi, Benjamin Benetti, Aazim Lakhani, and Alona Fyshe
- Can string kernels pass the test of time in Native Language Identification?
  Radu Tudor Ionescu and Marius Popescu
- Neural Networks and Spelling Features for Native Language Identification
  Johannes Bjerva, Gintare Grigonyte, Robert Östling, and Barbara Plank
- A study of N-gram and Embedding Representations for Native Language Identification
  Sowmya Vajjala and Sagnik Banerjee
- A Shallow Neural Network for Native Language Identification with Character N-grams
  Yunita Sari, Muhammad Rifqi Fatchurrahman, and Meisyarah Dwiastuti
- Fewer features perform well at Native Language Identification task
  Taraka Rama and Çağrı Çöltekin

14:45–15:30 **Poster Session B**

- Structured Generation of Technical Reading Lists
  Jonathan Gordon, Stephen Aguilar, Emily Sheng, and Gully Burns
- Effects of Lexical Properties on Viewing Time per Word in Autistic and Neurotypical Readers
  Sanja Štajner, Victoria Yaneva, Ruslan Mitkov, and Simone Paolo Ponzetto

- Transparent text quality assessment with convolutional neural networks
  Robert Östling and Gintare Grigonyte
- Artificial Error Generation with Machine Translation and Syntactic Patterns
  Marek Rei, Mariano Felice, Zheng Yuan, and Ted Briscoe
- Modelling semantic acquisition in second language learning
  Ekaterina Kochmar and Ekaterina Shutova
- Multiple Choice Question Generation Utilizing An Ontology
  Katherine Stasaski and Marti A. Hearst
- Simplifying metaphorical language for young readers: A corpus study on news text
  Magdalena Wolska and Yulia Clausen
- Language Based Mapping of Science Assessment Items to Skills
  Farah Nadeem and Mari Ostendorf
- Connecting the Dots: Towards Human-Level Grammatical Error Correction
  Shamil Chollampatt and Hwee Tou Ng
- Question Generation for Language Learning: From ensuring texts are read to supporting learning
  Maria Chinkina and Detmar Meurers
- Systematically Adapting Machine Translation for Grammatical Error Correction
  Courtney Napoles and Chris Callison-Burch
- Fine-grained essay scoring of a complex writing task for native speakers
  Andrea Horbach, Dirk Scholten-Akoun, Yuning Ding, and Torsten Zesch
- Exploring Optimal Voting in Native Language Identification
  Cyril Goutte and Serge Léger
- CIC-FBK Approach to Native Language Identification
  Ilia Markov, Lingzhen Chen, Carlo Strapparava, and Grigori Sidorov
- The Power of Character N-grams in Native Language Identification
  Artur Kulmizev, Bo Blankers, Johannes Bjerva, Malvina Nissim, Gertjan van Noord, Barbara Plank, and Martijn Wieling
- Classifier Stacking for Native Language Identification
  Wen Li and Liang Zou
- Native Language Identification on Text and Speech
  Marcos Zampieri, Alina Maria Ciobanu, and Liviu P. Dinu
- Native Language Identification using Phonetic Algorithms
  Charese Smiley and Sandra Kübler
- A deep-learning based native-language classification by using a latent semantic analysis for the NLI Shared Task 2017
  Yoo Rhee Oh, Hyung-Bae Jeon, Hwa Jeon Song, Yun-Kyung Lee, Jeon-Gue Park, and Yun-Keun Lee
- Fusion of Simple Models for Native Language Identification
  Fabio Kepler, Ramón Astudillo, and Alberto Abad
- Stacked Sentence-Document Classifier Approach for Improving Native Language Identification
  Andrea Cimino and Felice Dell'Orletta

15:30–16:00 **Break**

16:00–17:30 **Session 3**

16:00–16:25 Using Gaze to Predict Text Readability
Ana Valeria Gonzalez-Garduño and Anders Søgaard

16:25–16:50 Annotating Orthographic Target Hypotheses in a German L1 Learner Corpus
Ronja Laarmann-Quante, Katrin Ortmann, Anna Ehlert, Maurice Vogel, and Stefanie Dipper

16:50–17:15 A Large Scale Quantitative Exploration of Modeling Strategies for Content Scoring
Nitin Madnani, Anastassia Loukina, and Aoife Cahill

17:15–17:30 **Closing Remarks**

Workshop 11: 4th Workshop on Argument Mining

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

Location: Kastrup Airport

Friday, September 8, 2017

8:50–9:50 **Welcome session**

8:50–9:00 **Welcome (Workshop Chairs)**

9:00–9:50 **Invited talk (Christian Kock, Dept. of Media, Cognition and Communication, University of Copenhagen)**

9:50–10:30 **Paper session I**

9:50–10:10 200K+ Crowdsourced Political Arguments for a New Chilean Constitution
Constanza Fierro, Claudio Fuentes, Jorge Pérez, and Mauricio Quezada

10:10–10:30 Analyzing the Semantic Types of Claims and Premises in an Online Persuasive Forum
Christopher Hidey, Elena Musi, Alyssa Hwang, Smaranda Muresan, and Kathy McKeown

10:30–11:00 **Coffee break**

11:00–12:30 **Paper session II**

11:00–11:20 Annotation of argument structure in Japanese legal documents
Hiroaki Yamada, Simone Teufel, and Takenobu Tokunaga

11:20–11:40 Improving Claim Stance Classification with Lexical Knowledge Expansion and Context Utilization
Roy Bar-Haim, Lilach Edelstein, Charles Jochim, and Noam Slonim

11:40–12:00 Mining Argumentative Structure from Natural Language text using Automatically Generated Premise-Conclusion Topic Models
John Lawrence and Chris Reed

12:00–12:20 Building an Argument Search Engine for the Web
Henning Wachsmuth, Martin Potthast, Khalid Al Khatib, Yamen Ajjour, Jana Puschmann, Jiani Qu, Jonas Dorsch, Viorel Morari, Janek Bevendorff, and Benno Stein

12:30–14:30 **Lunch break**

14:30–15:30 **Poster session**

14:30–15:30 Argument Relation Classification Using a Joint Inference Model
Yufang Hou and Charles Jochim

14:30–15:30 Projection of Argumentative Corpora from Source to Target Languages
Ahmet Aker and Huangpan Zhang

14:30–15:30 Manual Identification of Arguments with Implicit Conclusions Using Semantic Rules for Argument Mining
Nancy Green

14:30–15:30 Unsupervised corpus—wide claim detection
Ran Levy, Shai Gretz, Benjamin Sznajder, Shay Hummel, Ranit Aharonov, and Noam Slonim

14:30–15:30 Using Question-Answering Techniques to Implement a Knowledge-Driven Argument Mining Approach
Patrick Saint-Dizier

14:30–15:30 What works and what does not: Classifier and feature analysis for argument mining
Ahmet Aker, Alfred Sliwa, Yuan Ma, Ruishen Lui, Niravkumar Borad, Seyedeh Ziyaei, and Mina Ghobadi

15:30–16:00 **Coffee break**

16:00–17:00 **Paper session III**

16:00–16:20 Unsupervised Detection of Argumentative Units though Topic Modeling Techniques
Alfio Ferrara, Stefano Montanelli, and Georgios Petasis

16:20–16:40 Using Complex Argumentative Interactions to Reconstruct the Argumentative Structure of Large-Scale Debates
John Lawrence and Chris Reed

16:40–17:00 Unit Segmentation of Argumentative Texts
Yamen Ajjour, Wei-Fan Chen, Johannes Kiesel, Henning Wachsmuth, and Benno Stein

17:00–17:30 **Wrap-up discussion**

Workshop 12: 8th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis

Organizers: *Alexandra Balahur, Saif Mohammad, and Erik van der Goot*

Location: Hovedbanen

Friday, September 8, 2017

08:30–08:40 **Opening Remarks**

08:40–10:30 **Session 1: Irony, stance and negotiating interpersonal meaning**

08:40–09:15 Detecting Sarcasm Using Different Forms Of Incongruity
Aditya Joshi

09:15–09:40 Assessing State-of-the-Art Sentiment Models on State-of-the-Art Sentiment Datasets
Jeremy Barnes, Roman Klinger, and Sabine Schulte im Walde

09:40–10:05 Annotation, Modelling and Analysis of Fine-Grained Emotions on a Stance and Sentiment Detection Corpus
Hendrik Schuff, Jeremy Barnes, Julian Mohme, Sebastian Padó, and Roman Klinger

10:05–10:30 Ranking Right-Wing Extremist Social Media Profiles by Similarity to Democratic and Extremist Groups
Matthias Hartung, Roman Klinger, Franziska Schmidtke, and Lars Vogel

10:30–11:00 **Coffee Break**

11:00–12:30 **Session 2: Emotion Intensity Task**

11:00–11:40 WASSA-2017 Shared Task on Emotion Intensity
Saif Mohammad and Felipe Bravo-Marquez

11:40–12:05 IMS at EmoInt-2017: Emotion Intensity Prediction with Affective Norms, Automatically Extended Resources and Deep Learning
Maximilian Köper, Evgeny Kim, and Roman Klinger

12:05–12:30 Prayas at EmoInt 2017: An Ensemble of Deep Neural Architectures for Emotion Intensity Prediction in Tweets
Prayas Jain, Pranav Goel, Devang Kulshreshtha, and Kaushal Kumar Shukla

12:30–14:00 **Lunch Break**

14:00–15:30 **Session 3: Sentiment, stance and emotion**

14:00–14:35 Latest News in Computational Argumentation: Surfing on the Deep Learning Wave, Scuba Diving in the Abyss of Fundamental Questions
Iryna Gurevych

14:35–15:00 Towards Syntactic Iberian Polarity Classification
David Vilares, Marcos Garcia, Miguel A. Alonso, and Carlos Gómez-Rodríguez

15:00–15:15 Toward Stance Classification Based on Claim Microstructures
Filip Boltuzic and Jan Šnajder

15:15–15:30 Linguistic Reflexes of Well-Being and Happiness in Echo
Jiaqi Wu, Marilyn Walker, Pranav Anand, and Steve Whittaker

15:30–16:00 **Coffee Break**

16:00–17:15 **Session 4: Preferences and values as determiners of sentiment and emotion**

16:00–16:35 Forecasting Consumer Spending from Purchase Intentions Expressed on Social Media
Viktor Pekar and Jane Binner

16:25–16:50 Mining fine-grained opinions on closed captions of YouTube videos with an attention-RNN
Edison Marrese-Taylor, Jorge Balazs, and Yutaka Matsuo

16:50–17:15 Understanding human values and their emotional effect
Alexandra Balahur

17:15–17:25 **Break**

17:25–18:25 **Session 5: Posters (Main Workshop and Emotion Intensity Task)**

- Did you ever read about Frogs drinking Coffee? Investigating the Compositionality of Multi-Emoji Expressions
Rebeca Padilla López and Fabienne Cap
- Investigating Redundancy in Emoji Use: Study on a Twitter Based Corpus
Giulia Donato and Patrizia Paggio
- Modeling Temporal Progression of Emotional Status in Mental Health Forum: A Recurrent Neural Net Approach
Kishaloy Halder, Lahari Poddar, and Min-Yen Kan
- Towards an integrated pipeline for aspect-based sentiment analysis in various domains
Orphee De Clercq, Els Lefever, Gilles Jacobs, Tijl Carpels, and Veronique Hoste
- Building a SentiWordNet for Odia
Gaurav Mohanty, Abishek Kannan, and Radhika Mamidi
- Lexicon Integrated CNN Models with Attention for Sentiment Analysis
Bonggun Shin, Timothy Lee, and Jinho D. Choi
- Explaining Recurrent Neural Network Predictions in Sentiment Analysis
Leila Arras, Grégoire Montavon, Klaus-Robert Müller, and Wojciech Samek
- GradAscent at EmoInt-2017: Character and Word Level Recurrent Neural Network Models for Tweet Emotion Intensity Detection
Egor Lakomkin, Chandrakant Bothe, and Stefan Wermter
- NUIG at EmoInt-2017: BiLSTM and SVR Ensemble to Detect Emotion Intensity
Vladimir Andryushechkin, Ian Wood, and James O' Neill
- Unsupervised Aspect Term Extraction with B-LSTM & CRF using Automatically Labelled Datasets
Athanasios Giannakopoulos, Claudiu Musat, Andreea Hossmann, and Michael Baeriswyl
- PLN-PUCRS at EmoInt-2017: Psycholinguistic features for emotion intensity prediction in tweets
Henrique Santos and Renata Vieira
- Textmining at EmoInt-2017: A Deep Learning Approach to Sentiment Intensity Scoring of English Tweets
Hardik Meisheri, Rupsa Saha, Priyanka Sinha, and Lipika Dey
- YNU-HPCC at EmoInt-2017: Using a CNN-LSTM Model for Sentiment Intensity Prediction
You Zhang, Hang Yuan, Jin Wang, and Xuejie Zhang
- Seernet at EmoInt-2017: Tweet Emotion Intensity Estimator
Venkatesh Duppada and Sushant Hiray

- IITP at EmoInt-2017: Measuring Intensity of Emotions using Sentence Embeddings and Optimized Features
  Md Shad Akhtar, Palaash Sawant, Asif Ekbal, Jyoti Pawar, and Pushpak Bhattacharyya
- NSEmo at EmoInt-2017: An Ensemble to Predict Emotion Intensity in Tweets
  Sreekanth Madisetty and Maunendra Sankar Desarkar
- Tecnolengua Lingmotif at EmoInt-2017: A lexicon-based approach
  Antonio Moreno-Ortiz
- EmoAtt at EmoInt-2017: Inner attention sentence embedding for Emotion Intensity
  Edison Marrese-Taylor and Yutaka Matsuo
- YZU-NLP at EmoInt-2017: Determining Emotion Intensity Using a Bi-directional LSTM-CNN Model
  Yuanye He, Liang-Chih Yu, K. Robert Lai, and Weiyi Liu
- DMGroup at EmoInt-2017: Emotion Intensity Using Ensemble Method
  Song Jiang and Xiaotian Han
- UWat-Emote at EmoInt-2017: Emotion Intensity Detection using Affect Clues, Sentiment Polarity and Word Embeddings
  Vineet John and Olga Vechtomova
- LIPN-UAM at EmoInt-2017:Combination of Lexicon-based features and Sentence-level Vector Representations for Emotion Intensity Determination
  Davide Buscaldi and Belem Priego
- deepCybErNet at EmoInt-2017: Deep Emotion Intensities in Tweets
  Vinayakumar R and Prabaharan Poornachandran

18:25–18:30 **Closing remarks**

Workshop 13: 2nd Workshop on Evaluating Vector Space Representations for NLP

Organizers: *Samuel Bowman, Yoav Goldberg, Felix Hill, Angeliki Lazaridou, Omer Levy, Roi Reichart, and Anders Søgaard*

Location: Skt. Hans Torv

Friday, September 8, 2017

09:00–09:20 **Opening Remarks**

09:20–09:55 **Shared task report**

- The RepEval 2017 Shared Task: Multi-Genre Natural Language Inference with Sentence Representations
  Nikita Nangia, Adina Williams, Angeliki Lazaridou, and Samuel Bowman

09:55–10:30 **Yejin Choi (University of Washington)**

10:30–11:00 **Coffee Break (set up posters)**

11:00–11:35 **Jakob Uszkoreit (Google Research)**

11:35–12:10 **Kyunghyun Cho (New York University)**

12:10–12:30 **Few Minutes Madness (Evaluation Proposals)**

- Traversal-Free Word Vector Evaluation in Analogy Space
  Xiaoyin Che, Nico Ring, Willi Raschkowski, Haojin Yang, and Christoph Meinel
- Hypothesis Testing based Intrinsic Evaluation of Word Embeddings
  Nishant Gurnani
- Evaluation of word embeddings against cognitive processes: primed reaction times in lexical decision and naming tasks
  Jeremy Auguste, Arnaud Rey, and Benoit Favre
- Playing with Embeddings : Evaluating embeddings for Robot Language Learning through MUD Games
  Anmol Gulati and Kumar Krishna Agrawal
- Recognizing Textual Entailment in Twitter Using Word Embeddings
  Octavia-Maria Şulea

12:30–14:00 **Lunch (somewhere together if pos)**

14:00–14:30 **Contributed Talks (shared task systems)**

14:00–14:15 Recurrent Neural Network-Based Sentence Encoder with Gated Attention for Natural Language Inference
Qian Chen, Xiaodan Zhu, Zhen-Hua Ling, Si Wei, Hui Jiang, and Diana Inkpen

14:15–14:30 Shortcut-Stacked Sentence Encoders for Multi-Domain Inference
Yixin Nie and Mohit Bansal

14:30–15:30 **Posters and discussion**

- Character-level Intra Attention Network for Natural Language Inference
  Han Yang, Marta R. Costa-jussà, and José A. R. Fonollosa

- Refining Raw Sentence Representations for Textual Entailment Recognition via Attention
  Jorge Balazs, Edison Marrese-Taylor, Pablo Loyola, and Yutaka Matsuo
- LCT-MALTA's Submission to RepEval 2017 Shared Task
  Hoa Vu

15:30–16:00 **Working Coffee Break**

16:00–17:30 **Presentation of Findings and Panel Discussion**

Workshop 14: Building Linguistically Generalizable NLP Systems

Organizers: *Emily M. Bender, Hal Daumé III, Allyson Ettinger, and Sudha Rao*

Location: Enghave Plads & Kødbyen

Friday, September 8, 2017

09:00–09:15 **Welcome Note**

- Towards Linguistically Generalizable NLP Systems: A Workshop and Shared Task
  Allyson Ettinger, Sudha Rao, Hal Daumé III, and Emily M. Bender

09:15–10:00 **Invited Talk (Aurelie Herbelot)**

10:00–12:10 **Session 1: Research Contribution Papers**

10:00–10:25 Analysing Errors of Open Information Extraction Systems
Rudolf Schneider, Tom Oberhauser, Tobias Klatt, Felix A. Gers, and Alexander Löser

10:30–11:00 **Coffee Break**

11:00–11:45 **Invited Talk (Grzegorz Chrupała)**

11:45–12:10 Massively Multilingual Neural Grapheme-to-Phoneme Conversion
Ben Peters, Jon Dehdari, and Josef van Genabith

12:10–12:30 **"Build It Break It, Language Edition"Shared Task Overview**

12:30–14:00 **Lunch Break**

14:00–14:45 **Invited Talk (Martha Palmer)**

14:45–15:35 **Session 2: Shared Task Description Papers**

14:45–15:10 BIBI System Description: Building with CNNs and Breaking with Deep Reinforcement Learning
Yitong Li, Trevor Cohn, and Timothy Baldwin

15:10–15:35 Breaking NLP: Using Morphosyntax, Semantics, Pragmatics and World Knowledge to Fool Sentiment Analysis Systems
Taylor Mahler, Willy Cheung, Micha Elsner, David King, Marie-Catherine de Marneffe, Cory Shain, Symon Stevens-Guille, and Michael White

15:35–16:00 **Coffee Break**

16:00–17:15 **Poster Session**

- An Adaptable Lexical Simplification Architecture for Major Ibero-Romance Languages
  Daniel Ferrés, Horacio Saggion, and Xavier Gómez Guinovart
- Cross-genre Document Retrieval: Matching between Conversational and Formal Writings
  Tomasz Jurczyk and Jinho D. Choi
- ACTSA: Annotated Corpus for Telugu Sentiment Analysis
  Sandeep Sricharan Mukku and Radhika Mamidi

- Strawman: An Ensemble of Deep Bag-of-Ngrams for Sentiment Analysis
  Kyunghyun Cho
- Breaking Sentiment Analysis of Movie Reviews
  Ieva Staliūnaitė and Ben Bonfil

17:15–17:30 **Closing Remarks**

5

Main Conference: Saturday, September 9

Overview

07:30 – 17:30	**Registration Day 1**			*Foyer*
08:00 – 08:30	**Morning Coffee**			*Foyer*
08:30 – 09:00	**Opening Remarks (General Chair, PC Co-Chairs)**			*Jutland*
09:00 – 10:00	**Invited Talk.**			*Jutland*
	Physical simulation, learning and language (Nando de Freitas).			
10:00 – 10:30	**Coffee Break**			*Foyer*
	Session 1			
10:30 – 12:10	Syntax 1	Information Extraction 1	Multilingual NLP	
	Jutland	*Funen*	*Zealand*	
10:30 – 12:10	**Demo Session (Posters)**			*Aarhus*
12:10 – 13:40	**Lunch**			
	Session 2			
13:40 – 15:20	Machine Translation 1	Language Grounding	Discourse and Summarization	
	Jutland	*Funen*	*Zealand*	
	Poster Session. Embeddings	Poster Session. Machine Learning 1	Poster Session. Sentiment Analysis 1	
	Aarhus	*Odense*	*Copenhagen*	
15:20 – 15:50	**Coffee Break**			*Foyer*
	Session 3			
15:50 – 17:30	Machine Learning 2	Generation	Semantics 1	
	Jutland	*Funen*	*Zealand*	
	Poster Session. Syntax 2	Poster Session. Question Answering and Machine Comprehension	Poster Session. Multimodal NLP 1	
	Aarhus	*Odense*	*Copenhagen*	

Invited Talk: Nando de Freitas

Physical simulation, learning and language

Saturday, September 9, 2017, 9:00–10:00

Jutland

Abstract: Simulated physical environments, with common physical laws, objects and agents with bodies, provide us with consistency to facilitate transfer and continual learning. In such environments, research topics such as learning to experiment, learning to learn and emergent communication can be easily explored. Given the relevance of these topics to language, it is natural to ask ourselves whether research in language would benefit from the development of such environments, and whether language can contribute toward improving such environments and agents within them. This talk will provide an overview of some of these environments, discuss learning to learn and its potential relevance to language, and present some deep reinforcement learning agents that capitalize on formal language instructions to develop disentangled interpretable representations that allow them to generalize to a wide variety of zero-shot semantic tasks. The talk will pose more questions than answers in the hope of stimulating discussion.

Biography: I was born in Zimbabwe, with malaria. I was a refugee from the war in Mocambique and thanks to my parents getting in debt to buy me a passport from a corrupt official, I grew up in Portugal without water and electricity, before the EU got there, and without my parents who were busy making money to pay their debt. At 8, I joined my parents in Venezuela and began school in the hood; see City of God. I moved to South Africa after highschool and sold beer illegally in black-townships for a living until 1991. Apartheid was the worst thing I ever experienced. I did my BSc in electrical engineering and MSc in control at the University of the Witwatersrand, where I strived to be the best student to prove to racists that anyone can do it. I did my PhD on Bayesian methods for neural networks at Trinity College, Cambridge University. I did a postdoc in Artificial Intelligence at UC Berkeley. I became a Full Professor at the University of British Columbia, before joining the University of Oxford in 2013. I quit Oxford in 2017 to join DeepMind full-time, where I lead the Machine Learning team. I aim to solve intelligence so that future generations have a better life. I have been a Senior Fellow of the Canadian Institute for Advanced Research for a long time. Some of my recent awards, mostly thanks to my collaborators, include: Best Paper Award at the International Conference on Machine Learning (2016), Best Paper Award at the International Conference on Learning Representations (2016), Winner of round 5 of the Yelp Dataset Challenge (2015), Distinguished Paper Award at the International Joint Conference on Artificial Intelligence (2013), Charles A. McDowell Award for Excellence in Research (2012), and Mathematics of Information Technology and Complex Systems Young Researcher Award (2010).

Session 1 Overview – Saturday, September 9, 2017

Oral tracks

Track A *Syntax 1* Jutland	**Track B** *Information Extraction 1* Funen	**Track C** *Multilingual NLP* Zealand	
Monolingual Phrase Alignment on Parse Forests *Arase and Tsujii*	Position-aware Attention and Supervised Data Improve Slot Filling *Zhang, Zhong, Chen, Angeli, and Manning*	Train-O-Matic: Large-Scale Supervised Word Sense Disambiguation in Multiple Languages without Manual Training Data *Pasini and Navigli*	10:30
Fast(er) Exact Decoding and Global Training for Transition-Based Dependency Parsing via a Minimal Feature Set *Shi, Huang, and Lee*	Heterogeneous Supervision for Relation Extraction: A Representation Learning Approach *Liu, Ren, Zhu, Zhi, Gui, Ji, and Han*	Universal Semantic Parsing *Reddy, Täckström, Petrov, Steedman, and Lapata*	10:55
[TACL] Parsing with Traces: An O(n^4) Algorithm and a Structural Representation *Kummerfeld and Klein*	Integrating Order Information and Event Relation for Script Event Prediction *Wang, Zhang, and Chang*	Mimicking Word Embeddings using Subword RNNs *Pinter, Guthrie, and Eisenstein*	11:20
Quasi-Second-Order Parsing for 1-Endpoint-Crossing, Pagenumber-2 Graphs *Cao, Huang, Sun, and Wan*	Entity Linking for Queries by Searching Wikipedia Sentences *Tan, Wei, Ren, Lv, and Zhou*	Past, Present, Future: A Computational Investigation of the Typology of Tense in 1000 Languages *Asgari and Schütze*	11:45

Poster tracks

10:30–12:10

Track D: *Demo Session (Posters)* — Aarhus

Parallel Session 1

Session 1A: Syntax 1

Jutland — Chair: *Joakim Nivre*

Monolingual Phrase Alignment on Parse Forests
Yuki Arase and Jun'ichi Tsujii 10:30–10:55

We propose an efficient method to conduct phrase alignment on parse forests for paraphrase detection. Unlike previous studies, our method identifies syntactic paraphrases under linguistically motivated grammar. In addition, it allows phrases to non-compositionally align to handle paraphrases with non-homographic phrase correspondences. A dataset that provides gold parse trees and their phrase alignments is created. The experimental results confirm that the proposed method conducts highly accurate phrase alignment compared to human performance.

Fast(er) Exact Decoding and Global Training for Transition-Based Dependency Parsing via a Minimal Feature Set
Tianze Shi, Liang Huang, and Lillian Lee 10:55–11:20

We first present a minimal feature set for transition-based dependency parsing, continuing a recent trend started by Kiperwasser and Goldberg (2016a) and Cross and Huang (2016a) of using bi-directional LSTM features. We plug our minimal feature set into the dynamic-programming framework of Huang and Sagae (2010) and Kuhlmann et al. (2011) to produce the first implementation of worst-case O(n^3) exact decoders for arc-hybrid and arc-eager transition systems. With our minimal features, we also present O(n^3) global training methods. Finally, using ensembles including our new parsers, we achieve the best unlabeled attachment score reported (to our knowledge) on the Chinese Treebank and the "second-best-in-class" result on the English Penn Treebank.

[TACL] Parsing with Traces: An O(n^4) Algorithm and a Structural Representation
Jonathan K. Kummerfeld and Dan Klein 11:20–11:45

General treebank analyses are graph structured, but parsers are typically restricted to tree structures for efficiency and modeling reasons. We propose a new representation and algorithm for a class of graph structures that is flexible enough to cover almost all treebank structures, but still admit efficient learning and inference. In particular, we consider directed, acyclic, one-endpoint-crossing graph structures, which cover most long-distance dislocation, shared argumentation, and similar tree-violating linguistic phenomena. We describe how to convert phrase structure parses, including traces, to our new representation, in a reversible manner. Our dynamic program uniquely decomposes structures, is sound and complete, and covers 97.3% of the Penn English treebank. We also implement a proof-of-concept parser that recovers a range of null elements and trace types.

Quasi-Second-Order Parsing for 1-Endpoint-Crossing, Pagenumber-2 Graphs
Junjie Cao, Sheng Huang, Weiwei Sun, and Xiaojun Wan 11:45–12:10

We propose a new Maximum Subgraph algorithm for first-order parsing to 1-endpoint-crossing, pagenumber-2 graphs. Our algorithm has two characteristics: (1) it separates the construction for noncrossing edges and crossing edges; (2) in a single construction step, whether to create a new arc is deterministic. These two characteristics make our algorithm relatively easy to be extended to incorporiate crossing-sensitive second-order features. We then introduce a new algorithm for quasi-second-order parsing. Experiments demonstrate that second-order features are helpful for Maximum Subgraph parsing.

Session 1B: Information Extraction 1

Funen — Chair: *Ming-Wei Chang*

Position-aware Attention and Supervised Data Improve Slot Filling
Yuhao Zhang, Victor Zhong, Danqi Chen, Gabor Angeli, and Christopher D. Manning 10:30–10:55

Organized relational knowledge in the form of "knowledge graphs" is important for many applications. However, the ability to populate knowledge bases with facts automatically extracted from documents has improved frustratingly slowly. This paper simultaneously addresses two issues that have held back prior work. We first propose an effective new model, which combines an LSTM sequence model with a form of entity position-aware attention that is better suited to relation extraction. Then we build TACRED, a large (119,474 examples) supervised relation extraction dataset obtained via crowdsourcing and targeted towards TAC KBP relations. The combination of better supervised data and a more appropriate high-capacity model enables much better relation extraction performance. When the model trained on this new dataset replaces the previous relation extraction component of the best TAC KBP 2015 slot filling system, its F1 score increases markedly from 22.2% to 26.7%.

Heterogeneous Supervision for Relation Extraction: A Representation Learning Approach
Liyuan Liu, Xiang Ren, Qi Zhu, Shi Zhi, Huan Gui, Heng Ji, and Jiawei Han 10:55–11:20

Relation extraction is a fundamental task in information extraction. Most existing methods have heavy reliance on annotations labeled by human experts, which are costly and time-consuming. To overcome this drawback, we propose a novel framework, REHession, to conduct relation extractor learning using annotations from heterogeneous information source, e.g., knowledge base and domain heuristics. These annotations, referred as heterogeneous supervision, often conflict with each other, which brings a new challenge to the original relation extraction task: how to infer the true label from noisy labels for a given instance. Identifying context information as the backbone of both relation extraction and true label discovery, we adopt embedding techniques to learn the distributed representations of context, which bridges all components with mutual enhancement in an iterative fashion. Extensive experimental results demonstrate the superiority of REHession over the state-of-the-art.

Integrating Order Information and Event Relation for Script Event Prediction
Zhongqing Wang, Yue Zhang, and Ching-Yun Chang 11:20–11:45

There has been a recent line of work automatically learning scripts from unstructured texts, by modeling narrative event chains. While the dominant approach group events using event pair relations, LSTMs have been used to encode full chains of narrative events. The latter has the advantage of learning long-range temporal orders, yet the former is more adaptive to partial orders. We propose a neural model that leverages the advantages of both methods, by using LSTM hidden states as features for event pair modelling. A dynamic memory network is utilized to automatically induce weights on existing events for inferring a subsequent event. Standard evaluation shows that our method significantly outperforms both methods above, giving the best results reported so far.

Entity Linking for Queries by Searching Wikipedia Sentences
Chuanqi Tan, Furu Wei, Pengjie Ren, Weifeng Lv, and Ming Zhou 11:45–12:10

We present a simple yet effective approach for linking entities in queries. The key idea is to search sentences similar to a query from Wikipedia articles and directly use the human-annotated entities in the similar sentences as candidate entities for the query. Then, we employ a rich set of features, such as link-probability, context-matching, word embeddings, and relatedness among candidate entities as well as their related entities, to rank the candidates under a regression based framework. The advantages of our approach lie in two aspects, which contribute to the ranking process and final linking result. First, it can greatly reduce the number of candidate entities by filtering out irrelevant entities with the words in the query. Second, we can obtain the query sensitive prior probability in addition to the static link-probability derived from all Wikipedia articles. We conduct experiments on two benchmark datasets on entity linking for queries, namely the ERD14 dataset and the GERDAQ dataset. Experimental results show that our method outperforms state-of-the-art systems and yields 75.0% in F1 on the ERD14 dataset and 56.9% on the GERDAQ dataset.

Session 1C: Multilingual NLP

Zealand Chair: *Ivan Titov*

Train-O-Matic: Large-Scale Supervised Word Sense Disambiguation in Multiple Languages without Manual Training Data
Tommaso Pasini and Roberto Navigli 10:30–10:55

Annotating large numbers of sentences with senses is the heaviest requirement of current Word Sense Disambiguation. We present Train-O-Matic, a language-independent method for generating millions of sense-annotated training instances for virtually all meanings of words in a language's vocabulary. The approach is fully automatic: no human intervention is required and the only type of human knowledge used is a WordNet-like resource. Train-O-Matic achieves consistently state-of-the-art performance across gold standard datasets and languages, while at the same time removing the burden of manual annotation. All the training data is available for research purposes at http://trainomatic.org.

Universal Semantic Parsing
Siva Reddy, Oscar Täckström, Slav Petrov, Mark Steedman, and Mirella Lapata 10:55–11:20

Universal Dependencies (UD) offer a uniform cross-lingual syntactic representation, with the aim of advancing multilingual applications. Recent work shows that semantic parsing can be accomplished by transforming syntactic dependencies to logical forms. However, this work is limited to English, and cannot process dependency graphs, which allow handling complex phenomena such as control. In this work, we introduce UDepLambda, a semantic interface for UD, which maps natural language to logical forms in an almost language-independent fashion and can process dependency graphs. We perform experiments on question answering against Freebase and provide German and Spanish translations of the WebQuestions and GraphQuestions datasets to facilitate multilingual evaluation. Results show that UDepLambda outperforms strong baselines across languages and datasets. For English, it achieves a 4.9 F1 point improvement over the state-of-the-art on GraphQuestions.

Mimicking Word Embeddings using Subword RNNs
Yuval Pinter, Robert Guthrie, and Jacob Eisenstein 11:20–11:45

Word embeddings improve generalization over lexical features by placing each word in a lower-dimensional space, using distributional information obtained from unlabeled data. However, the effectiveness of word embeddings for downstream NLP tasks is limited by out-of-vocabulary (OOV) words, for which embeddings do not exist. In this paper, we present MIMICK, an approach to generating OOV word embeddings compositionally, by learning a function from spellings to distributional embeddings. Unlike prior work, MIMICK does not require re-training on the original word embedding corpus; instead, learning is performed at the type level. Intrinsic and extrinsic evaluations demonstrate the power of this simple approach. On 23 languages, MIMICK improves performance over a word-based baseline for tagging part-of-speech and morphosyntactic attributes. It is competitive with (and complementary to) a supervised character-based model in low resource settings.

Past, Present, Future: A Computational Investigation of the Typology of Tense in 1000 Languages
Ehsaneddin Asgari and Hinrich Schütze 11:45–12:10

We present SuperPivot, an analysis method for low-resource languages that occur in a superparallel corpus, i.e., in a corpus that contains an order of magnitude more languages than parallel corpora currently in use. We show that SuperPivot performs well for the crosslingual analysis of the linguistic phenomenon of tense. We produce analysis results for more than 1000 languages, conducting – to the best of our knowledge – the largest crosslingual computational study performed to date. We extend existing methodology for leveraging parallel corpora for typological analysis by overcoming a limiting assumption of earlier work: We only require that a linguistic feature is overtly marked in a few of thousands of languages as opposed to requiring that it be marked in all languages under investigation.

Session 1D: Demo Session (Posters)

10:30–12:10
Aarhus
Chair: *Michael Paul*

The NLTK FrameNet API: Designing for Discoverability with a Rich Linguistic Resource
Nathan Schneider and Chuck Wooters

A new Python API, integrated within the NLTK suite, offers access to the FrameNet 1.7 lexical database. The lexicon (structured in terms of frames) as well as annotated sentences can be processed programatically, or browsed with human-readable displays via the interactive Python prompt.

Argotario: Computational Argumentation Meets Serious Games
Ivan Habernal, Raffael Hannemann, Christian Pollak, Christopher Klamm, Patrick Pauli, and Iryna Gurevych

An important skill in critical thinking and argumentation is the ability to spot and recognize fallacies. Fallacious arguments, omnipresent in argumentative discourse, can be deceptive, manipulative, or simply leading to 'wrong moves' in a discussion. Despite their importance, argumentation scholars and NLP researchers with focus on argumentation quality have not yet investigated fallacies empirically. The nonexistence of resources dealing with fallacious argumentation calls for scalable approaches to data acquisition and annotation, for which the serious games methodology offers an appealing, yet unexplored, alternative. We present Argotario, a serious game that deals with fallacies in everyday argumentation. Argotario is a multilingual, open-source, platform-independent application with strong educational aspects, accessible at www.argotario.net.

An Analysis and Visualization Tool for Case Study Learning of Linguistic Concepts
Cecilia Ovesdotter Alm, Benjamin Meyers, and Emily Prud'hommeaux

We present an educational tool that integrates computational linguistics resources for use in non-technical undergraduate language science courses. By using the tool in conjunction with evidence-driven pedagogical case studies, we strive to provide opportunities for students to gain an understanding of linguistic concepts and analysis through the lens of realistic problems in feasible ways. Case studies tend to be used in legal, business, and health education contexts, but less in the teaching and learning of linguistics. The approach introduced also has potential to encourage students across training backgrounds to continue on to computational language analysis coursework.

GraphDocExplore: A Framework for the Experimental Comparison of Graph-based Document Exploration Techniques
Tobias Falke and Iryna Gurevych

Graphs have long been proposed as a tool to browse and navigate in a collection of documents in order to support exploratory search. Many techniques to automatically extract different types of graphs, showing for example entities or concepts and different relationships between them, have been suggested. While experimental evidence that they are indeed helpful exists for some of them, it is largely unknown which type of graph is most helpful for a specific exploratory task. However, carrying out experimental comparisons with human subjects is challenging and time-consuming. Towards this end, we present the *GraphDocExplore* framework. It provides an intuitive web interface for graph-based document exploration that is optimized for experimental user studies. Through a generic graph interface, different methods to extract graphs from text can be plugged into the system. Hence, they can be compared at minimal implementation effort in an environment that ensures controlled comparisons. The system is publicly available under an open-source license.

SGNMT – A Flexible NMT Decoding Platform for Quick Prototyping of New Models and Search Strategies
Felix Stahlberg, Eva Hasler, Danielle Saunders, and Bill Byrne

This paper introduces SGNMT, our experimental platform for machine translation research. SGNMT provides a generic interface to neural and symbolic scoring modules (predictors) with left-to-right semantic such as translation models like NMT, language models, translation lattices, n-best lists or other kinds of scores and constraints. Predictors can be combined with other predictors to form complex decoding tasks. SGNMT implements a number of search strategies for traversing the space spanned by the predictors which are appropriate for different predictor constellations. Adding new predictors or decoding strategies is particularly easy, making it a very efficient tool for prototyping new research ideas. SGNMT is actively being used by students in the MPhil program in Machine Learning, Speech and Language Technology at the University of Cambridge for course work and theses, as well as for most of the research work

in our group.

StruAP: A Tool for Bundling Linguistic Trees through Structure-based Abstract Pattern
Kohsuke Yanai, Misa Sato, Toshihiko Yanase, Kenzo Kurotsuchi, Yuta Koreeda, and Yoshiki Niwa

We present a tool for developing tree structure patterns that makes it easy to define the relations among textual phrases and create a search index for these newly defined relations. By using the proposed tool, users develop tree structure patterns through abstracting syntax trees. The tool features (1) intuitive pattern syntax, (2) unique functions such as recursive call of patterns and use of lexicon dictionaries, and (3) whole workflow support for relation development and validation. We report the current implementation of the tool and its effectiveness.

KnowYourNyms? A Game of Semantic Relationships
Ross Mechanic, Dean Fulgoni, Hannah Cutler, Sneha Rajana, Zheyuan Liu, Bradley Jackson, Anne Cocos, Chris Callison-Burch, and Marianna Apidianaki

Semantic relation knowledge is crucial for natural language understanding. We introduce "KnowYourNyms?", a web-based game for learning semantic relations. While providing users with an engaging experience, the application collects large amounts of data that can be used to improve semantic relation classifiers. The data also broadly informs us of how people perceive the relationships between words, providing useful insights for research in psychology and linguistics.

The Projector: An Interactive Annotation Projection Visualization Tool
Alan Akbik and Roland Vollgraf

Previous works proposed annotation projection in parallel corpora to inexpensively generate treebanks or propbanks for new languages. In this approach, linguistic annotation is automatically transferred from a resource-rich source language (SL) to translations in a target language (TL). However, annotation projection may be adversely affected by translational divergences between specific language pairs. For this reason, previous work often required careful qualitative analysis of projectability of specific annotation in order to define strategies to address quality and coverage issues. In this demonstration, we present THE PROJECTOR, an interactive GUI designed to assist researchers in such analysis: it allows users to execute and visually inspect annotation projection in a range of different settings. We give an overview of the GUI, discuss use cases and illustrate how the tool can facilitate discussions with the research community.

Interactive Visualization for Linguistic Structure
Aaron Sarnat, Vidur Joshi, Cristian Petrescu-Prahova, Alvaro Herrasti, Brandon Stilson, and Mark Hopkins

We provide a visualization library and web interface for interactively exploring a parse tree or a forest of parses. The library is not tied to any particular linguistic representation, but provides a general-purpose API for the interactive exploration of hierarchical linguistic structure. To facilitate rapid understanding of a complex structure, the API offers several important features, including expand/collapse functionality, positional and color cues, explicit visual support for sequential structure, and dynamic highlighting to convey node-to-text correspondence.

DLATK: Differential Language Analysis ToolKit
H. Andrew Schwartz, Salvatore Giorgi, Maarten Sap, Patrick Crutchley, Lyle Ungar, and Johannes Eichstaedt

We present Differential Language Analysis Toolkit (DLATK), an open-source python package and command-line tool developed for conducting social-scientific language analyses. While DLATK provides standard NLP pipeline steps such as tokenization or SVM-classification, its novel strengths lie in analyses useful for psychological, health, and social science: (1) incorporation of extra-linguistic structured information, (2) specified levels and units of analysis (e.g. document, user, community), (3) statistical metrics for continuous outcomes, and (4) robust, proven, and accurate pipelines for social-scientific prediction problems. DLATK integrates multiple popular packages (SKLearn, Mallet), enables interactive usage (Jupyter Notebooks), and generally follows object oriented principles to make it easy to tie in additional libraries or storage technologies.

QUINT: Interpretable Question Answering over Knowledge Bases
Abdalghani Abujabal, Rishiraj Saha Roy, Mohamed Yahya, and Gerhard Weikum

We present QUINT, a live system for question answering over knowledge bases. QUINT automatically learns role-aligned utterance-query templates from user questions paired with their answers. When

QUINT answers a question, it visualizes the complete derivation sequence from the natural language utterance to the final answer. The derivation provides an explanation of how the syntactic structure of the question was used to derive the structure of a SPARQL query, and how the phrases in the question were used to instantiate different parts of the query. When an answer seems unsatisfactory, the derivation provides valuable insights towards reformulating the question.

Function Assistant: A Tool for NL Querying of APIs
Kyle Richardson and Jonas Kuhn

In this paper, we describe Function Assistant, a lightweight Python-based toolkit for querying and exploring source code repositories using natural language. The toolkit is designed to help end-users of a target API quickly find information about functions through high-level natural language queries, or descriptions. For a given text query and background API, the tool finds candidate functions by performing a translation from the text to known representations in the API using the semantic parsing approach of (Richardson and Kuhn, 2017). Translations are automatically learned from example text-code pairs in example APIs. The toolkit includes features for building translation pipelines and query engines for arbitrary source code projects. To explore this last feature, we perform new experiments on 27 well-known Python projects hosted on Github.

MoodSwipe: A Soft Keyboard that Suggests MessageBased on User-Specified Emotions
Chieh-Yang Huang, Tristan Labetoulle, Ting-Hao Huang, Yi-Pei Chen, Hung-Chen Chen, Vallari Srivastava, and Lun-Wei Ku

We present MoodSwipe, a soft keyboard that suggests text messages given the user-specified emotions utilizing the real dialog data. The aim of MoodSwipe is to create a convenient user interface to enjoy the technology of emotion classification and text suggestion, and at the same time to collect labeled data automatically for developing more advanced technologies. While users select the MoodSwipe keyboard, they can type as usual but sense the emotion conveyed by their text and receive suggestions for their message as a benefit. In MoodSwipe, the detected emotions serve as the medium for suggested texts, where viewing the latter is the incentive to correcting the former. We conduct several experiments to show the superiority of the emotion classification models trained on the dialog data, and further to verify good emotion cues are important context for text suggestion.

ParlAI: A Dialog Research Software Platform
Alexander Miller, Will Feng, Dhruv Batra, Antoine Bordes, Adam Fisch, Jiasen Lu, Devi Parikh, and Jason Weston

We introduce ParlAI (pronounced "par-lay"), an open-source software platform for dialog research implemented in Python, available at http://parl.ai. Its goal is to provide a unified framework for sharing, training and testing dialog models; integration of Amazon Mechanical Turk for data collection, human evaluation, and online/reinforcement learning; and a repository of machine learning models for comparing with others' models, and improving upon existing architectures. Over 20 tasks are supported in the first release, including popular datasets such as SQuAD, bAbI tasks, MCTest, WikiQA, QACNN, QADailyMail, CBT, bAbI Dialog, Ubuntu, OpenSubtitles and VQA. Several models are integrated, including neural models such as memory networks, seq2seq and attentive LSTMs.

HeidelPlace: An Extensible Framework for Geoparsing
Ludwig Richter, Johanna Geiß, Andreas Spitz, and Michael Gertz

Geographic information extraction from textual data sources, called geoparsing, is a key task in text processing and central to subsequent spatial analysis approaches. Several geoparsers are available that support this task, each with its own (often limited or specialized) gazetteer and its own approaches to toponym detection and resolution. In this demonstration paper, we present HeidelPlace, an extensible framework in support of geoparsing. Key features of HeidelPlace include a generic gazetteer model that supports the integration of place information from different knowledge bases, and a pipeline approach that enables an effective combination of diverse modules tailored to specific geoparsing tasks. This makes HeidelPlace a valuable tool for testing and evaluating different gazetteer sources and geoparsing methods. In the demonstration, we show how to set up a geoparsing workflow with HeidelPlace and how it can be used to compare and consolidate the output of different geoparsing approaches.

Unsupervised, Knowledge-Free, and Interpretable Word Sense Disambiguation
Alexander Panchenko, Fide Marten, Eugen Ruppert, Stefano Faralli, Dmitry Ustalov, Simone Paolo Ponzetto, and Chris Biemann

Interpretability of a predictive model is a powerful feature that gains the trust of users in the correctness of the predictions. In word sense disambiguation (WSD), knowledge-based systems tend to be much more interpretable than knowledge-free counterparts as they rely on the wealth of manually-encoded elements representing word senses, such as hypernyms, usage examples, and images. We present a WSD system that bridges the gap between these two so far disconnected groups of methods. Namely, our system, providing access to several state-of-the-art WSD models, aims to be interpretable as a knowledge-based system while it remains completely unsupervised and knowledge-free. The presented tool features a Web interface for all-word disambiguation of texts that makes the sense predictions human readable by providing interpretable word sense inventories, sense representations, and disambiguation results. We provide a public API, enabling seamless integration.

NeuroNER: an easy-to-use program for named-entity recognition based on neural networks
Franck Dernoncourt, Ji Young Lee, and Peter Szolovits

Named-entity recognition (NER) aims at identifying entities of interest in a text. Artificial neural networks (ANNs) have recently been shown to outperform existing NER systems. However, ANNs remain challenging to use for non-expert users. In this paper, we present NeuroNER, an easy-to-use named-entity recognition tool based on ANNs. Users can annotate entities using a graphical web-based user interface (BRAT): the annotations are then used to train an ANN, which in turn predict entities' locations and categories in new texts. NeuroNER makes this annotation-training-prediction flow smooth and accessible to anyone.

SupWSD: A Flexible Toolkit for Supervised Word Sense Disambiguation
Simone Papandrea, Alessandro Raganato, and Claudio Delli Bovi

In this demonstration we present SupWSD, a Java API for supervised Word Sense Disambiguation (WSD). This toolkit includes the implementation of a state-of-the-art supervised WSD system, together with a Natural Language Processing pipeline for preprocessing and feature extraction. Our aim is to provide an easy-to-use tool for the research community, designed to be modular, fast and scalable for training and testing on large datasets. The source code of SupWSD is available at http://github.com/SI3P/SupWSD.

Interactive Abstractive Summarization for Event News Tweets
Ori Shapira, Hadar Ronen, Meni Adler, Yael Amsterdamer, Judit Bar-Ilan, and Ido Dagan

We present a novel interactive summarization system that is based on abstractive summarization, derived from a recent consolidated knowledge representation for multiple texts. We incorporate a couple of interaction mechanisms, providing a bullet-style summary while allowing to attain the most important information first and interactively drill down to more specific details. A usability study of our implementation, for event news tweets, suggests the utility of our approach for text exploration.

LangPro: Natural Language Theorem Prover
Lasha Abzianidze

LangPro is an automated theorem prover for natural language. Given a set of premises and a hypothesis, it is able to prove semantic relations between them. The prover is based on a version of analytic tableau method specially designed for natural logic. The proof procedure operates on logical forms that preserve linguistic expressions to a large extent. %This property makes the logical forms easily obtainable from syntactic trees. %, in particular, Combinatory Categorial Grammar derivation trees. The nature of proofs is deductive and transparent. On the FraCaS and SICK textual entailment datasets, the prover achieves high results comparable to state-of-the-art.

Interactive Visualization and Manipulation of Attention-based Neural Machine Translation
Jaesong Lee, Joong-Hwi Shin, and Jun-Seok Kim

While neural machine translation (NMT) provides high-quality translation, it is still hard to interpret and analyze its behavior. We present an interactive interface for visualizing and intervening behavior of NMT, specifically concentrating on the behavior of beam search mechanism and attention component. The tool (1) visualizes search tree and attention and (2) provides interface to adjust search tree and attention weight (manually or automatically) at real-time. We show the tool gives various methods to understand NMT.

Session 2 Overview – Saturday, September 9, 2017

Oral tracks

Track A *Machine Translation 1* Jutland	**Track B** *Language Grounding* Funen	**Track C** *Discourse and Summarization* Zealand	
Neural Machine Translation with Source-Side Latent Graph Parsing *Hashimoto and Tsuruoka*	Where is Misty? Interpreting Spatial Descriptors by Modeling Regions in Space *Kitaev and Klein*	End-to-end Neural Coreference Resolution *Lee, He, Lewis, and Zettlemoyer*	13:40
Neural Machine Translation with Word Predictions *Weng, Huang, Zheng, Dai, and Chen*	Continuous Representation of Location for Geolocation and Lexical Dialectology using Mixture Density Networks *Rahimi, Baldwin, and Cohn*	Neural Net Models of Open-domain Discourse Coherence *Li and Jurafsky*	14:05
Towards Decoding as Continuous Optimisation in Neural Machine Translation *Hoang, Haffari, and Cohn*	[TACL] Colors in Context: A Pragmatic Neural Model for Grounded Language Understanding *Monroe, Hawkins, Goodman, and Potts*	Affinity-Preserving Random Walk for Multi-Document Summarization *Wang, Liu, Sui, and Chang*	14:30
[TACL] Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation *Johnson, Schuster, Le, Krikun, Wu, Chen, Thorat, Viégas, Wattenberg, Corrado, Hughes, and Dean*	Obj2Text: Generating Visually Descriptive Language from Object Layouts *Yin and Ordonez*	A Mention-Ranking Model for Abstract Anaphora Resolution *Marasovic, Born, Opitz, and Frank*	14:55

Poster tracks

13:40–15:20

Track D: *Poster Session. Embeddings* — Aarhus

Track E: *Poster Session. Machine Learning 1* — Odense

Track F: *Poster Session. Sentiment Analysis 1* — Copenhagen

Parallel Session 2

Session 2A: Machine Translation 1

Jutland Chair: *Graham Neubig*

Neural Machine Translation with Source-Side Latent Graph Parsing
Kazuma Hashimoto and Yoshimasa Tsuruoka 13:40–14:05

This paper presents a novel neural machine translation model which jointly learns translation and source-side latent graph representations of sentences. Unlike existing pipelined approaches using syntactic parsers, our end-to-end model learns a latent graph parser as part of the encoder of an attention-based neural machine translation model, and thus the parser is optimized according to the translation objective. In experiments, we first show that our model compares favorably with state-of-the-art sequential and pipelined syntax-based NMT models. We also show that the performance of our model can be further improved by pre-training it with a small amount of treebank annotations. Our final ensemble model significantly outperforms the previous best models on the standard English-to-Japanese translation dataset.

Neural Machine Translation with Word Predictions
Rongxiang Weng, Shujian Huang, Zaixiang Zheng, Xin-Yu Dai, and Jiajun Chen 14:05–14:30

In the encoder-decoder architecture for neural machine translation (NMT), the hidden states of the recurrent structures in the encoder and decoder carry the crucial information about the sentence. These vectors are generated by parameters which are updated by back-propagation of translation errors through time.We argue that propagating errors through the end-to-end recurrent structures are not a direct way of control the hidden vectors. In this paper, we propose to use word predictions as a mechanism for direct supervision. More specifically, we require these vectors to be able to predict the vocabulary in target sentence. Our simple mechanism ensures better representations in the encoder and decoder without using any extra data or annotation. It is also helpful in reducing the target side vocabulary and improving the decoding efficiency. Experiments on Chinese-English machine translation task show an average BLEU improvement by 4.53, respectively.

Towards Decoding as Continuous Optimisation in Neural Machine Translation
Cong Duy Vu Hoang, Gholamreza Haffari, and Trevor Cohn 14:30–14:55

We propose a novel decoding approach for neural machine translation (NMT) based on continuous optimisation. We reformulate decoding, a discrete optimization problem, into a continuous problem, such that optimization can make use of efficient gradient-based techniques. Our powerful decoding framework allows for more accurate decoding for standard neural machine translation models, as well as enabling decoding in intractable models such as intersection of several different NMT models. Our empirical results show that our decoding framework is effective, and can leads to substantial improvements in translations, especially in situations where greedy search and beam search are not feasible. Finally, we show how the technique is highly competitive with, and complementary to, reranking.

[TACL] Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation
Melvin Johnson, Mike Schuster, Quoc V. Le, Maxim Krikun, Yonghui Wu, Zhifeng Chen, Nikhil Thorat, Fernanda Viégas, Martin Wattenberg, Greg Corrado, Macduff Hughes, and Jeffrey Dean 14:55–15:20

We propose a simple solution to use a single Neural Machine Translation (NMT) model to translate between multiple languages. Our solution requires no changes to the model architecture from a standard NMT system but instead introduces an artificial token at the beginning of the input sentence to specify the required target language. Using a shared wordpiece vocabulary, our approach enables Multilingual NMT using a single model. On the WMT'14 benchmarks, a single multilingual model achieves comparable performance for English→French and surpasses state-of-the-art results for English→German. Similarly, a single multilingual model surpasses state-of-the-art results for French→English and German→English on WMT'14 and WMT'15 benchmarks, respectively. On production corpora, multilingual models of up to twelve language pairs allow for better translation of many individual pairs. Our models can also learn to perform implicit bridging between language pairs never seen explicitly during training, showing that transfer learning and zero-shot translation is possible for neural translation. Finally, we show analyses that hints at a universal interlingua representation in our models and show some interesting examples when mixing languages.

Session 2B: Language Grounding

Funen — Chair: *Yejin Choi*

Where is Misty? Interpreting Spatial Descriptors by Modeling Regions in Space
Nikita Kitaev and Dan Klein 13:40–14:05

We present a model for locating regions in space based on natural language descriptions. Starting with a 3D scene and a sentence, our model is able to associate words in the sentence with regions in the scene, interpret relations such as 'on top of' or 'next to,' and finally locate the region described in the sentence. All components form a single neural network that is trained end-to-end without prior knowledge of object segmentation. To evaluate our model, we construct and release a new dataset consisting of Minecraft scenes with crowdsourced natural language descriptions. We achieve a 32% relative error reduction compared to a strong neural baseline.

Continuous Representation of Location for Geolocation and Lexical Dialectology using Mixture Density Networks
Afshin Rahimi, Timothy Baldwin, and Trevor Cohn 14:05–14:30

We propose a method for embedding two-dimensional locations in a continuous vector space using a neural network-based model incorporating mixtures of Gaussian distributions, presenting two model variants for text-based geolocation and lexical dialectology. Evaluated over Twitter data, the proposed model outperforms conventional regression-based geolocation and provides a better estimate of uncertainty. We also show the effectiveness of the representation for predicting words from location in lexical dialectology, and evaluate it using the DARE dataset.

[TACL] Colors in Context: A Pragmatic Neural Model for Grounded Language Understanding
Will Monroe, Robert X. D. Hawkins, Noah D. Goodman, and Christopher Potts 14:30–14:55

We present a model of pragmatic referring expression interpretation in a grounded communication task (identifying colors from descriptions) that draws upon predictions from two recurrent neural network classifiers, a speaker and a listener, unified by a recursive pragmatic reasoning framework. Experiments show that this combined pragmatic model interprets color descriptions more accurately than the classifiers from which it is built, and that much of this improvement results from combining the speaker and listener perspectives. We observe that pragmatic reasoning helps primarily in the hardest cases: when the model must distinguish very similar colors, or when few utterances adequately express the target color. Our findings make use of a newly-collected corpus of human utterances in color reference games, which exhibit a variety of pragmatic behaviors. We also show that the embedded speaker model reproduces many of these pragmatic behaviors.

Obj2Text: Generating Visually Descriptive Language from Object Layouts
Xuwang Yin and Vicente Ordonez 14:55–15:20

Generating captions for images is a task that has recently received considerable attention. Another type of visual inputs are abstract scenes or object layouts where the only information provided is a set of objects and their locations. This type of imagery is commonly found in many applications in computer graphics, virtual reality, and storyboarding. We explore in this paper OBJ2TEXT, a sequence-to-sequence model that encodes a set of objects and their locations as an input sequence using an LSTM network, and decodes this representation using an LSTM language model. We show in our paper that this model despite using a sequence encoder can effectively represent complex spatial object-object relationships and produce descriptions that are globally coherent and semantically relevant. We test our approach for the task of describing object layouts in the MS-COCO dataset by producing sentences given only object annotations. We additionally show that our model combined with a state-of-the-art object detector can improve the accuracy of an image captioning model.

Session 2C: Discourse and Summarization

Zealand — Chair: *Lu Wang*

End-to-end Neural Coreference Resolution
Kenton Lee, Luheng He, Mike Lewis, and Luke Zettlemoyer 13:40–14:05

We introduce the first end-to-end coreference resolution model and show that it significantly outperforms all previous work without using a syntactic parser or hand-engineered mention detector. The key idea is to directly consider all spans in a document as potential mentions and learn distributions over possible antecedents for each. The model computes span embeddings that combine context-dependent boundary representations with a head-finding attention mechanism. It is trained to maximize the marginal likelihood of gold antecedent spans from coreference clusters and is factored to enable aggressive pruning of potential mentions. Experiments demonstrate state-of-the-art performance, with a gain of 1.5 F1 on the OntoNotes benchmark and by 3.1 F1 using a 5-model ensemble, despite the fact that this is the first approach to be successfully trained with no external resources.

Neural Net Models of Open-domain Discourse Coherence
Jiwei Li and Dan Jurafsky 14:05–14:30

Discourse coherence is strongly associated with text quality, making it important to natural language generation and understanding. Yet existing models of coherence focus on measuring individual aspects of coherence (lexical overlap, rhetorical structure, entity centering) in narrow domains. In this paper, we describe domain-independent neural models of discourse coherence that are capable of measuring multiple aspects of coherence in existing sentences and can maintain coherence while generating new sentences. We study both discriminative models that learn to distinguish coherent from incoherent discourse, and generative models that produce coherent text, including a novel neural latent-variable Markovian generative model that captures the latent discourse dependencies between sentences in a text. Our work achieves state-of-the-art performance on multiple coherence evaluations, and marks an initial step in generating coherent texts given discourse contexts.

Affinity-Preserving Random Walk for Multi-Document Summarization
Kexiang Wang, Tianyu Liu, Zhifang Sui, and Baobao Chang 14:30–14:55

Multi-document summarization provides users with a short text that summarizes the information in a set of related documents. This paper introduces affinity-preserving random walk to the summarization task, which preserves the affinity relations of sentences by an absorbing random walk model. Meanwhile, we put forward adjustable affinity-preserving random walk to enforce the diversity constraint of summarization in the random walk process. The ROUGE evaluations on DUC 2003 topic-focused summarization task and DUC 2004 generic summarization task show the good performance of our method, which has the best ROUGE-2 recall among the graph-based ranking methods.

A Mention-Ranking Model for Abstract Anaphora Resolution
Ana Marasovic, Leo Born, Juri Opitz, and Anette Frank 14:55–15:20

Resolving abstract anaphora is an important, but difficult task for text understanding. Yet, with recent advances in representation learning this task becomes a more tangible aim. A central property of abstract anaphora is that it establishes a relation between the anaphor embedded in the anaphoric sentence and its (typically non-nominal) antecedent. We propose a mention-ranking model that learns how abstract anaphors relate to their antecedents with an LSTM-Siamese Net. We overcome the lack of training data by generating artificial anaphoric sentence–antecedent pairs. Our model outperforms state-of-the-art results on shell noun resolution. We also report first benchmark results on an abstract anaphora subset of the ARRAU corpus. This corpus presents a greater challenge due to a mixture of nominal and pronominal anaphors and a greater range of confounders. We found model variants that outperform the baselines for nominal anaphors, without training on individual anaphor data, but still lag behind for pronominal anaphors. Our model selects syntactically plausible candidates and – if disregarding syntax – discriminates candidates using deeper features.

Session 2D: Poster Session. Embeddings

13:40–15:20
Aarhus
Chair: *Heike Adel*

Hierarchical Embeddings for Hypernymy Detection and Directionality
Kim Anh Nguyen, Maximilian Köper, Sabine Schulte im Walde, and Ngoc Thang Vu

We present a novel neural model HyperVec to learn hierarchical embeddings for hypernymy detection and directionality. While previous embeddings have shown limitations on prototypical hypernyms, HyperVec represents an unsupervised measure where embeddings are learned in a specific order and capture the hypernym—hyponym distributional hierarchy. Moreover, our model is able to generalize over unseen hypernymy pairs, when using only small sets of training data, and by mapping to other languages. Results on benchmark datasets show that HyperVec outperforms both state-of-the- art unsupervised measures and embedding models on hypernymy detection and directionality, and on predicting graded lexical entailment.

Ngram2vec: Learning Improved Word Representations from Ngram Co-occurrence Statistics
Zhe Zhao, Tao Liu, Shen Li, Bofang Li, and Xiaoyong Du

The existing word representation methods mostly limit their information source to word co-occurrence statistics. In this paper, we introduce ngrams into four representation methods: SGNS, GloVe, PPMI matrix, and its SVD factorization. Comprehensive experiments are conducted on word analogy and similarity tasks. The results show that improved word representations are learned from ngram co-occurrence statistics. We also demonstrate that the trained ngram representations are useful in many aspects such as finding antonyms and collocations. Besides, a novel approach of building co-occurrence matrix is proposed to alleviate the hardware burdens brought by ngrams.

Dict2vec : Learning Word Embeddings using Lexical Dictionaries
Julien Tissier, Christopher Gravier, and Amaury Habrard

Learning word embeddings on large unlabeled corpus has been shown to be successful in improving many natural language tasks. The most efficient and popular approaches learn or retrofit such representations using additional external data. Resulting embeddings are generally better than their corpus-only counterparts, although such resources cover a fraction of words in the vocabulary. In this paper, we propose a new approach, Dict2vec, based on one of the largest yet refined datasource for describing words — natural language dictionaries. Dict2vec builds new word pairs from dictionary entries so that semantically-related words are moved closer, and negative sampling filters out pairs whose words are unrelated in dictionaries. We evaluate the word representations obtained using Dict2vec on eleven datasets for the word similarity task and on four datasets for a text classification task.

Learning Chinese Word Representations From Glyphs Of Characters
Tzu-ray Su and Hung-yi Lee

In this paper, we propose new methods to learn Chinese word representations. Chinese characters are composed of graphical components, which carry rich semantics. It is common for a Chinese learner to comprehend the meaning of a word from these graphical components. As a result, we propose models that enhance word representations by character glyphs. The character glyph features are directly learned from the bitmaps of characters by convolutional auto-encoder(convAE), and the glyph features improve Chinese word representations which are already enhanced by character embeddings. Another contribution in this paper is that we created several evaluation datasets in traditional Chinese and made them public.

Learning Paraphrastic Sentence Embeddings from Back-Translated Bitext
John Wieting, Jonathan Mallinson, and Kevin Gimpel

We consider the problem of learning general-purpose, paraphrastic sentence embeddings in the setting of Wieting et al. (2016b). We use neural machine translation to generate sentential paraphrases via back-translation of bilingual sentence pairs. We evaluate the paraphrase pairs by their ability to serve as training data for learning paraphrastic sentence embeddings. We find that the data quality is stronger than prior work based on bitext and on par with manually-written English paraphrase pairs, with the advantage that our approach can scale up to generate large training sets for many languages and domains. We experiment with several language pairs and data sources, and develop a variety of data filtering techniques. In the process, we explore how neural machine translation output differs from human-written sentences, finding clear differences in length, the amount of repetition, and the use of rare words.

Joint Embeddings of Chinese Words, Characters, and Fine-grained Subcharacter Components
Jinxing Yu, Xun Jian, Hao Xin, and Yangqiu Song

Word embeddings have attracted much attention recently. Different from alphabetic writing systems, Chinese characters are often composed of subcharacter components which are also semantically informative. In this work, we propose an approach to jointly embed Chinese words as well as their characters and fine-grained subcharacter components. We use three likelihoods to evaluate whether the context words, characters, and components can predict the current target word, and collected 13,253 subcharacter components to demonstrate the existing approaches of decomposing Chinese characters are not enough. Evaluation on both word similarity and word analogy tasks demonstrates the superior performance of our model.

Exploiting Morphological Regularities in Distributional Word Representations
Arihant Gupta, Syed Sarfaraz Akhtar, Avijit Vajpayee, Arjit Srivastava, Madan Gopal Jhanwar, and Manish Shrivastava

We present an unsupervised, language agnostic approach for exploiting morphological regularities present in high dimensional vector spaces. We propose a novel method for generating embeddings of words from their morphological variants using morphological transformation operators. We evaluate this approach on MSR word analogy test set with an accuracy of 85% which is 12% higher than the previous best known system.

Exploiting Word Internal Structures for Generic Chinese Sentence Representation
Shaonan Wang, Jiajun Zhang, and Chengqing Zong

We introduce a novel mixed characterword architecture to improve Chinese sentence representations, by utilizing rich semantic information of word internal structures. Our architecture uses two key strategies. The first is a mask gate on characters, learning the relation among characters in a word. The second is a maxpooling operation on words, adaptively finding the optimal mixture of the atomic and compositional word representations. Finally, the proposed architecture is applied to various sentence composition models, which achieves substantial performance gains over baseline models on sentence similarity task.

High-risk learning: acquiring new word vectors from tiny data
Aurélie Herbelot and Marco Baroni

Distributional semantics models are known to struggle with small data. It is generally accepted that in order to learn 'a good vector' for a word, a model must have sufficient examples of its usage. This contradicts the fact that humans can guess the meaning of a word from a few occurrences only. In this paper, we show that a neural language model such as Word2Vec only necessitates minor modifications to its standard architecture to learn new terms from tiny data, using background knowledge from a previously learnt semantic space. We test our model on word definitions and on a nonce task involving 2-6 sentences' worth of context, showing a large increase in performance over state-of-the-art models on the definitional task.

Word Embeddings based on Fixed-Size Ordinally Forgetting Encoding
Joseph Sanu, Mingbin Xu, Hui Jiang, and Quan Liu

In this paper, we propose to learn word embeddings based on the recent fixed-size ordinally forgetting encoding (FOFE) method, which can almost uniquely encode any variable-length sequence into a fixed-size representation. We use FOFE to fully encode the left and right context of each word in a corpus to construct a novel word-context matrix, which is further weighted and factorized using truncated SVD to generate low-dimension word embedding vectors. We evaluate this alternate method in encoding word-context statistics and show the new FOFE method has a notable effect on the resulting word embeddings. Experimental results on several popular word similarity tasks have demonstrated that the proposed method outperforms other SVD models that use canonical count based techniques to generate word context matrices.

VecShare: A Framework for Sharing Word Representation Vectors
Jared Fernandez, Zhaocheng Yu, and Doug Downey

Many Natural Language Processing (NLP) models rely on distributed vector representations of words. Because the process of training word vectors can require large amounts of data and computation, NLP researchers and practitioners often utilize pre-trained embeddings downloaded from the Web. However, finding the best embeddings for a given task is difficult, and can be computationally prohibitive. We present a framework, called VecShare, that makes it easy to share and retrieve word embeddings on the

Web. The framework leverages a public data-sharing infrastructure to host embedding sets, and provides automated mechanisms for retrieving the embeddings most similar to a given corpus. We perform an experimental evaluation of VecShare's similarity strategies, and show that they are effective at efficiently retrieving embeddings that boost accuracy in a document classification task. Finally, we provide an open-source Python library for using the VecShare framework.

Word Re-Embedding via Manifold Dimensionality Retention
Souleiman Hasan and Edward Curry

Word embeddings seek to recover a Euclidean metric space by mapping words into vectors, starting from words co-occurrences in a corpus. Word embeddings may underestimate the similarity between nearby words, and overestimate it between distant words in the Euclidean metric space. In this paper, we re-embed pre-trained word embeddings with a stage of manifold learning which retains dimensionality. We show that this approach is theoretically founded in the metric recovery paradigm, and empirically show that it can improve on state-of-the-art embeddings in word similarity tasks 0.5 - 5.0% points depending on the original space.

MUSE: Modularizing Unsupervised Sense Embeddings
Guang-He Lee and Yun-Nung Chen

This paper proposes to address the word sense ambiguity issue in an unsupervised manner, where word sense representations are learned along a word sense selection mechanism given contexts. Prior work focused on designing a single model to deliver both mechanisms, and thus suffered from either coarse-grained representation learning or inefficient sense selection. The proposed modular approach, MUSE, implements flexible modules to optimize distinct mechanisms, achieving the first purely sense-level representation learning system with linear-time sense selection. We leverage reinforcement learning to enable joint training on the proposed modules, and introduce various exploration techniques on sense selection for better robustness. The experiments on benchmark data show that the proposed approach achieves the state-of-the-art performance on synonym selection as well as on contextual word similarities in terms of MaxSimC.

Session 2E: Poster Session. Machine Learning 1

13:40–15:20
Odense
Chair: *Pontus Stenetorp*

Reporting Score Distributions Makes a Difference: Performance Study of LSTM-networks for Sequence Tagging
Nils Reimers and Iryna Gurevych

In this paper we show that reporting a single performance score is insufficient to compare non-deterministic approaches. We demonstrate for common sequence tagging tasks that the seed value for the random number generator can result in statistically significant $p < 10^-4$ differences for state-of-the-art systems. For two recent systems for NER, we observe an absolute difference of one percentage point F_1-score depending on the selected seed value, making these systems perceived either as state-of-the-art or mediocre. Instead of publishing and reporting single performance scores, we propose to compare score distributions based on multiple executions. Based on the evaluation of 50.000 LSTM-networks for five sequence tagging tasks, we present network architectures that produce both superior performance as well as are more stable with respect to the remaining hyperparameters.

Learning What's Easy: Fully Differentiable Neural Easy-First Taggers
André F. T. Martins and Julia Kreutzer

We introduce a novel neural easy-first decoder that learns to solve sequence tagging tasks in a flexible order. In contrast to previous easy-first decoders, our models are end-to-end differentiable. The decoder iteratively updates a "sketch" of the predictions over the sequence. At its core is an attention mechanism that controls which parts of the input are strategically the best to process next. We present a new constrained softmax transformation that ensures the same cumulative attention to every word, and show how to efficiently evaluate and backpropagate over it. Our models compare favourably to BILSTM taggers on three sequence tagging tasks.

Incremental Skip-gram Model with Negative Sampling
Nobuhiro Kaji and Hayato Kobayashi

This paper explores an incremental training strategy for the skip-gram model with negative sampling (SGNS) from both empirical and theoretical perspectives. Existing methods of neural word embeddings, including SGNS, are multi-pass algorithms and thus cannot perform incremental model update. To address this problem, we present a simple incremental extension of SGNS and provide a thorough theoretical analysis to demonstrate its validity. Empirical experiments demonstrated the correctness of the theoretical analysis as well as the practical usefulness of the incremental algorithm.

Learning to select data for transfer learning with Bayesian Optimization
Sebastian Ruder and Barbara Plank

Domain similarity measures can be used to gauge adaptability and select suitable data for transfer learning, but existing approaches define ad hoc measures that are deemed suitable for respective tasks. Inspired by work on curriculum learning, we propose to learn data selection measures using Bayesian Optimization and evaluate them across models, domains and tasks. Our learned measures outperform existing domain similarity measures significantly on three tasks: sentiment analysis, part-of-speech tagging, and parsing. We show the importance of complementing similarity with diversity, and that learned measures are–to some degree–transferable across models, domains, and even tasks.

Unsupervised Pretraining for Sequence to Sequence Learning
Prajit Ramachandran, Peter Liu, and Quoc Le

This work presents a general unsupervised learning method to improve the accuracy of sequence to sequence (seq2seq) models. In our method, the weights of the encoder and decoder of a seq2seq model are initialized with the pretrained weights of two language models and then fine-tuned with labeled data. We apply this method to challenging benchmarks in machine translation and abstractive summarization and find that it significantly improves the subsequent supervised models. Our main result is that pretraining improves the generalization of seq2seq models. We achieve state-of-the-art results on the WMT English→German task, surpassing a range of methods using both phrase-based machine translation and neural machine translation. Our method achieves a significant improvement of 1.3 BLEU from th previous best models on both WMT'14 and WMT'15 English→German. We also conduct human evaluations on abstractive summarization and find that our method outperforms a purely supervised learning baseline in a statistically significant manner.

Efficient Attention using a Fixed-Size Memory Representation
Denny Britz, Melody Guan, and Minh-Thang Luong

The standard content-based attention mechanism typically used in sequence-to-sequence models is computationally expensive as it requires the comparison of large encoder and decoder states at each time step. In this work, we propose an alternative attention mechanism based on a fixed size memory representation that is more efficient. Our technique predicts a compact set of K attention contexts during encoding and lets the decoder compute an efficient lookup that does not need to consult the memory. We show that our approach performs on-par with the standard attention mechanism while yielding inference speedups of 20% for real-world translation tasks and more for tasks with longer sequences. By visualizing attention scores we demonstrate that our models learn distinct, meaningful alignments.

Rotated Word Vector Representations and their Interpretability
Sungjoon Park, JinYeong Bak, and Alice Oh

Vector representation of words improves performance in various NLP tasks, but the high dimensional word vectors are very difficult to interpret. We apply several rotation algorithms to the vector representation of words to improve the interpretability. Unlike previous approaches that induce sparsity, the rotated vectors are interpretable while preserving the expressive performance of the original vectors. Furthermore, any prebuilt word vector representation can be rotated for improved interpretability. We apply rotation to skipgrams and glove and compare the expressive power and interpretability with the original vectors and the sparse overcomplete vectors. The results show that the rotated vectors outperform the original and the sparse overcomplete vectors for interpretability and expressiveness tasks.

A causal framework for explaining the predictions of black-box sequence-to-sequence models
David Alvarez-Melis and Tommi Jaakkola

We interpret the predictions of any black-box structured input-structured output model around a specific input-output pair. Our method returns an "explanation" consisting of groups of input-output tokens that are causally related. These dependencies are inferred by querying the model with perturbed inputs, generating a graph over tokens from the responses, and solving a partitioning problem to select the most relevant components. We focus the general approach on sequence-to-sequence problems, adopting a variational autoencoder to yield meaningful input perturbations. We test our method across several NLP sequence generation tasks.

Piecewise Latent Variables for Neural Variational Text Processing
Iulian Vlad Serban, Alexander G. Ororbia, Joelle Pineau, and Aaron Courville

Advances in neural variational inference have facilitated the learning of powerful directed graphical models with continuous latent variables, such as variational autoencoders. The hope is that such models will learn to represent rich, multi-modal latent factors in real-world data, such as natural language text. However, current models often assume simplistic priors on the latent variables - such as the uni-modal Gaussian distribution - which are incapable of representing complex latent factors efficiently. To overcome this restriction, we propose the simple, but highly flexible, piecewise constant distribution. This distribution has the capacity to represent an exponential number of modes of a latent target distribution, while remaining mathematically tractable. Our results demonstrate that incorporating this new latent distribution into different models yields substantial improvements in natural language processing tasks such as document modeling and natural language generation for dialogue.

Learning the Structure of Variable-Order CRFs: a finite-state perspective
Thomas Lavergne and François Yvon

The computational complexity of linear-chain Conditional Random Fields (CRFs) makes it difficult to deal with very large label sets and long range dependencies. Such situations are not rare and arise when dealing with morphologically rich languages or joint labelling tasks. We extend here recent proposals to consider variable order CRFs. Using an effective finite-state representation of variable-length dependencies, we propose new ways to perform feature selection at large scale and report experimental results where we outperform strong baselines on a tagging task.

Sparse Communication for Distributed Gradient Descent
Alham Fikri Aji and Kenneth Heafield

We make distributed stochastic gradient descent faster by exchanging sparse updates instead of dense updates. Gradient updates are positively skewed as most updates are near zero, so we map the 99%

smallest updates (by absolute value) to zero then exchange sparse matrices. This method can be combined with quantization to further improve the compression. We explore different configurations and apply them to neural machine translation and MNIST image classification tasks. Most configurations work on MNIST, whereas different configurations reduce convergence rate on the more complex translation task. Our experiments show that we can achieve up to 49% speed up on MNIST and 22% on NMT without damaging the final accuracy or BLEU.

Why ADAGRAD Fails for Online Topic Modeling
You Lu, Jeffrey Lund, and Jordan Boyd-Graber

Online topic modeling, i.e., topic modeling with stochastic variational inference, is a powerful and efficient technique for analyzing large datasets, and ADAGRAD is a widely-used technique for tuning learning rates during online gradient optimization. However, these two techniques do not work well together. We show that this is because ADAGRAD uses accumulation of previous gradients as the learning rates' denominators. For online topic modeling, the magnitude of gradients is very large. It causes learning rates to shrink very quickly, so the parameters cannot fully converge until the training ends

Session 2F: Poster Session. Sentiment Analysis 1

13:40–15:20
Copenhagen
Chair: *Diyi Yang*

Recurrent Attention Network on Memory for Aspect Sentiment Analysis
Peng Chen, Zhongqian Sun, Lidong Bing, and Wei Yang

We propose a novel framework based on neural networks to identify the sentiment of opinion targets in a comment/review. Our framework adopts multiple-attention mechanism to capture sentiment features separated by a long distance, so that it is more robust against irrelevant information. The results of multiple attentions are non-linearly combined with a recurrent neural network, which strengthens the expressive power of our model for handling more complications. The weighted-memory mechanism not only helps us avoid the labor-intensive feature engineering work, but also provides a tailor-made memory for different opinion targets of a sentence. We examine the merit of our model on four datasets: two are from SemEval2014, i.e. reviews of restaurants and laptops; a twitter dataset, for testing its performance on social media data; and a Chinese news comment dataset, for testing its language sensitivity. The experimental results show that our model consistently outperforms the state-of-the-art methods on different types of data.

A Cognition Based Attention Model for Sentiment Analysis
Yunfei Long, Lu Qin, Rong Xiang, Minglei Li, and Chu-Ren Huang

Attention models are proposed in sentiment analysis because some words are more important than others. However,most existing methods either use local context based text information or user preference information. In this work, we propose a novel attention model trained by cognition grounded eye-tracking data. A reading prediction model is first built using eye-tracking data as dependent data and other features in the context as independent data. The predicted reading time is then used to build a cognition based attention (CBA) layer for neural sentiment analysis. As a comprehensive model, We can capture attentions of words in sentences as well as sentences in documents. Different attention mechanisms can also be incorporated to capture other aspects of attentions. Evaluations show the CBA based method outperforms the state-of-the-art local context based attention methods significantly. This brings insight to how cognition grounded data can be brought into NLP tasks.

Author-aware Aspect Topic Sentiment Model to Retrieve Supporting Opinions from Reviews
Lahari Poddar, Wynne Hsu, and Mong Li Lee

User generated content about products and services in the form of reviews are often diverse and even contradictory. This makes it difficult for users to know if an opinion in a review is prevalent or biased. We study the problem of searching for supporting opinions in the context of reviews. We propose a framework called SURF, that first identifies opinions expressed in a review, and then finds similar opinions from other reviews. We design a novel probabilistic graphical model that captures opinions as a combination of aspect, topic and sentiment dimensions, takes into account the preferences of individual authors, as well as the quality of the entity under review, and encodes the flow of thoughts in a review by constraining the aspect distribution dynamically among successive review segments. We derive a similarity measure that considers both lexical and semantic similarity to find supporting opinions. Experiments on TripAdvisor hotel reviews and Yelp restaurant reviews show that our model outperforms existing methods for modeling opinions, and the proposed framework is effective in finding supporting opinions.

Magnets for Sarcasm: Making Sarcasm Detection Timely, Contextual and Very Personal
Aniruddha Ghosh and Tony Veale

Sarcasm is a pervasive phenomenon in social media, permitting the concise communication of meaning, affect and attitude. Concision requires wit to produce and wit to understand, which demands from each party knowledge of norms, context and a speaker's mindset. Insight into a speaker's psychological profile at the time of production is a valuable source of context for sarcasm detection. Using a neural architecture, we show significant gains in detection accuracy when knowledge of the speaker's mood at the time of production can be inferred. Our focus is on sarcasm detection on Twitter, and show that the mood exhibited by a speaker over tweets leading up to a new post is as useful a cue for sarcasm as the topical context of the post itself. The work opens the door to an empirical exploration not just of sarcasm in text but of the sarcastic state of mind.

Identifying Humor in Reviews using Background Text Sources
Alex Morales and Chengxiang Zhai

We study the problem of automatically identifying humorous text from a new kind of text data, i.e., online reviews. We propose a generative language model, based on the theory of incongruity, to model humorous text, which allows us to leverage background text sources, such as Wikipedia entry descriptions, and enables construction of multiple features for identifying humorous reviews. Evaluation of these features using supervised learning for classifying reviews into humorous and non-humorous reviews shows that the features constructed based on the proposed generative model are much more effective than the major features proposed in the existing literature, allowing us to achieve almost 86% accuracy. These humorous review predictions can also supply good indicators for identifying helpful reviews.

Sentiment Lexicon Construction with Representation Learning Based on Hierarchical Sentiment Supervision
Leyi Wang and Rui Xia

Sentiment lexicon is an important tool for identifying the sentiment polarity of words and texts. How to automatically construct sentiment lexicons has become a research topic in the field of sentiment analysis and opinion mining. Recently there were some attempts to employ representation learning algorithms to construct a sentiment lexicon with sentiment-aware word embedding. However, these methods were normally trained under document-level sentiment supervision. In this paper, we develop a neural architecture to train a sentiment-aware word embedding by integrating the sentiment supervision at both document and word levels, to enhance the quality of word embedding as well as the sentiment lexicon. Experiments on the SemEval 2013-2016 datasets indicate that the sentiment lexicon generated by our approach achieves the state-of-the-art performance in both supervised and unsupervised sentiment classification, in comparison with several strong sentiment lexicon construction methods.

Towards a Universal Sentiment Classifier in Multiple languages
Kui Xu and Xiaojun Wan

Existing sentiment classifiers usually work for only one specific language, and different classification models are used in different languages. In this paper we aim to build a universal sentiment classifier with a single classification model in multiple different languages. In order to achieve this goal, we propose to learn multilingual sentiment-aware word embeddings simultaneously based only on the labeled reviews in English and unlabeled parallel data available in a few language pairs. It is not required that the parallel data exist between English and any other language, because the sentiment information can be transferred into any language via pivot languages. We present the evaluation results of our universal sentiment classifier in five languages, and the results are very promising even when the parallel data between English and the target languages are not used. Furthermore, the universal single classifier is compared with a few cross-language sentiment classifiers relying on direct parallel data between the source and target languages, and the results show that the performance of our universal sentiment classifier is very promising compared to that of different cross-language classifiers in multiple target languages.

Capturing User and Product Information for Document Level Sentiment Analysis with Deep Memory Network
Zi-Yi Dou

Document-level sentiment classification is a fundamental problem which aims to predict a user's overall sentiment about a product in a document. Several methods have been proposed to tackle the problem whereas most of them fail to consider the influence of users who express the sentiment and products which are evaluated. To address the issue, we propose a deep memory network for document-level sentiment classification which could capture the user and product information at the same time. To prove the effectiveness of our algorithm, we conduct experiments on IMDB and Yelp datasets and the results indicate that our model can achieve better performance than several existing methods.

Identifying and Tracking Sentiments and Topics from Social Media Texts during Natural Disasters
Min Yang, Jincheng Mei, Heng Ji, Zhao Wei, Zhou Zhao, and Xiaojun Chen

We study the problem of identifying the topics and sentiments and tracking their shifts from social media texts in different geographical regions during emergencies and disasters. We propose a location-based dynamic sentiment-topic model (LDST) which can jointly model topic, sentiment, time and Geolocation information. The experimental results demonstrate that LDST performs very well at discovering topics and sentiments from social media and tracking their shifts in different geographical regions during emergencies and disasters. We will release the data and source code after this work is published.

Refining Word Embeddings for Sentiment Analysis
Liang-Chih Yu, Jin Wang, K. Robert Lai, and Xuejie Zhang

Word embeddings that can capture semantic and syntactic information from contexts have been extensively used for various natural language processing tasks. However, existing methods for learning context-based word embeddings typically fail to capture sufficient sentiment information. This may result in words with similar vector representations having an opposite sentiment polarity (e.g., good and bad), thus degrading sentiment analysis performance. Therefore, this study proposes a word vector refinement model that can be applied to any pre-trained word vectors (e.g., Word2vec and GloVe). The refinement model is based on adjusting the vector representations of words such that they can be closer to both semantically and sentimentally similar words and further away from sentimentally dissimilar words. Experimental results show that the proposed method can improve conventional word embeddings and outperform previously proposed sentiment embeddings for both binary and fine-grained classification on Stanford Sentiment Treebank (SST).

A Multilayer Perceptron based Ensemble Technique for Fine-grained Financial Sentiment Analysis
Md Shad Akhtar, Abhishek Kumar, Deepanway Ghosal, Asif Ekbal, and Pushpak Bhattacharyya

In this paper, we propose a novel method for combining deep learning and classical feature based models using a Multi-Layer Perceptron (MLP) network for financial sentiment analysis. We develop various deep learning models based on Convolutional Neural Network (CNN), Long Short Term Memory (LSTM) and Gated Recurrent Unit (GRU). These are trained on top of pre-trained, autoencoder-based, financial word embeddings and lexicon features. An ensemble is constructed by combining these deep learning models and a classical supervised model based on Support Vector Regression (SVR). We evaluate our proposed technique on a benchmark dataset of SemEval-2017 shared task on financial sentiment analysis. The propose model shows impressive results on two datasets, i.e. microblogs and news headlines datasets. Comparisons show that our proposed model performs better than the existing state-of-the-art systems for the above two datasets by 2.0 and 4.1 cosine points, respectively.

Sentiment Intensity Ranking among Adjectives Using Sentiment Bearing Word Embeddings
Raksha Sharma, Arpan Somani, Lakshya Kumar, and Pushpak Bhattacharyya

Identification of intensity ordering among polar (positive or negative) words which have the same semantics can lead to a fine-grained sentiment analysis. For example, 'master', 'seasoned' and 'familiar' point to different intensity levels, though they all convey the same meaning (semantics), i.e., expertise: having a good knowledge of. In this paper, we propose a semi-supervised technique that uses sentiment bearing word embeddings to produce a continuous ranking among adjectives that share common semantics. Our system demonstrates a strong Spearman's rank correlation of 0.83 with the gold standard ranking. We show that sentiment bearing word embeddings facilitate a more accurate intensity ranking system than other standard word embeddings (word2vec and GloVe). Word2vec is the state-of-the-art for intensity ordering task.

Sentiment Lexicon Expansion Based on Neural PU Learning, Double Dictionary Lookup, and Polarity Association
Yasheng Wang, Yang Zhang, and Bing Liu

Although many sentiment lexicons in different languages exist, most are not comprehensive. In a recent sentiment analysis application, we used a large Chinese sentiment lexicon and found that it missed a large number of sentiment words in social media. This prompted us to make a new attempt to study sentiment lexicon expansion. This paper first poses the problem as a PU learning problem, which is a new formulation. It then proposes a new PU learning method suitable for our problem using a neural network. The results are enhanced further with a new dictionary-based technique and a novel polarity classification technique. Experimental results show that the proposed approach outperforms baseline methods greatly.

Session 3 Overview – Saturday, September 9, 2017

Oral tracks

	Track A *Machine Learning 2* Jutland	**Track B** *Generation* Funen	**Track C** *Semantics 1* Zealand
15:50	DeepPath: A Reinforcement Learning Method for Knowledge Graph Reasoning *Xiong, Hoang, and Wang*	Split and Rephrase *Narayan, Gardent, Cohen, and Shimorina*	Measuring Thematic Fit with Distributional Feature Overlap *Santus, Chersoni, Lenci, and Blache*
16:15	Task-Oriented Query Reformulation with Reinforcement Learning *Nogueira and Cho*	Neural Response Generation via GAN with an Approximate Embedding Layer *Xu, Liu, Wang, Sun, Wang, Wang, and Qi*	SCDV : Sparse Composite Document Vectors using soft clustering over distributional representations *Mekala, Gupta, Paranjape, and Karnick*
16:40	Sentence Simplification with Deep Reinforcement Learning *Zhang and Lapata*	A Hybrid Convolutional Variational Autoencoder for Text Generation *Semeniuta, Severyn, and Barth*	Supervised Learning of Universal Sentence Representations from Natural Language Inference Data *Conneau, Kiela, Schwenk, Barrault, and Bordes*
17:05	Learning how to Active Learn: A Deep Reinforcement Learning Approach *Fang, Li, and Cohn*	Filling the Blanks (hint: plural noun) for Mad Libs Humor *Hossain, Krumm, Vanderwende, Horvitz, and Kautz*	Determining Semantic Textual Similarity using Natural Deduction Proofs *Yanaka, Mineshima, Martínez-Gómez, and Bekki*

Poster tracks

15:50–17:30

Track D: *Poster Session. Syntax 2* — Aarhus

Track E: *Poster Session. Question Answering and Machine Comprehension* — Odense

Track F: *Poster Session. Multimodal NLP 1* — Copenhagen

Parallel Session 3

Session 3A: Machine Learning 2

Jutland Chair: *Karl Moritz Hermann*

DeepPath: A Reinforcement Learning Method for Knowledge Graph Reasoning
Wenhan Xiong, Thien Hoang, and William Yang Wang 15:50–16:15

We study the problem of learning to reason in large scale knowledge graphs (KGs). More specifically, we describe a novel reinforcement learning framework for learning multi-hop relational paths: we use a policy-based agent with continuous states based on knowledge graph embeddings, which reasons in a KG vector-space by sampling the most promising relation to extend its path. In contrast to prior work, our approach includes a reward function that takes the accuracy, diversity, and efficiency into consideration. Experimentally, we show that our proposed method outperforms a path-ranking based algorithm and knowledge graph embedding methods on Freebase and Never-Ending Language Learning datasets.

Task-Oriented Query Reformulation with Reinforcement Learning
Rodrigo Nogueira and Kyunghyun Cho 16:15–16:40

Search engines play an important role in our everyday lives by assisting us in finding the information we need. When we input a complex query, however, results are often far from satisfactory. In this work, we introduce a query reformulation system based on a neural network that rewrites a query to maximize the number of relevant documents returned. We train this neural network with reinforcement learning. The actions correspond to selecting terms to build a reformulated query, and the reward is the document recall. We evaluate our approach on three datasets against strong baselines and show a relative improvement of 5-20% in terms of recall. Furthermore, we present a simple method to estimate a conservative upper-bound performance of a model in a particular environment and verify that there is still large room for improvements.

Sentence Simplification with Deep Reinforcement Learning
Xingxing Zhang and Mirella Lapata 16:40–17:05

Sentence simplification aims to make sentences easier to read and understand. Most recent approaches draw on insights from machine translation to learn simplification rewrites from monolingual corpora of complex and simple sentences. We address the simplification problem with an encoder-decoder model coupled with a deep reinforcement learning framework. Our model, which we call DRESS (as shorthand for **D**eep **RE**inforcement **S**entence **S**implification), explores the space of possible simplifications while learning to optimize a reward function that encourages outputs which are simple, fluent, and preserve the meaning of the input. Experiments on three datasets demonstrate that our model outperforms competitive simplification systems.

Learning how to Active Learn: A Deep Reinforcement Learning Approach
Meng Fang, Yuan Li, and Trevor Cohn 17:05–17:30

Active learning aims to select a small subset of data for annotation such that a classifier learned on the data is highly accurate. This is usually done using heuristic selection methods, however the effectiveness of such methods is limited and moreover, the performance of heuristics varies between datasets. To address these shortcomings, we introduce a novel formulation by reframing the active learning as a reinforcement learning problem and explicitly learning a data selection policy, where the policy takes the role of the active learning heuristic. Importantly, our method allows the selection policy learned using simulation to one language to be transferred to other languages. We demonstrate our method using cross-lingual named entity recognition, observing uniform improvements over traditional active learning algorithms.

Session 3B: Generation

Funen — Chair: *Wei Xu*

Split and Rephrase
Shashi Narayan, Claire Gardent, Shay B. Cohen, and Anastasia Shimorina 15:50–16:15

We propose a new sentence simplification task (Split-and-Rephrase) where the aim is to split a complex sentence into a meaning preserving sequence of shorter sentences. Like sentence simplification, splitting-and-rephrasing has the potential of benefiting both natural language processing and societal applications. Because shorter sentences are generally better processed by NLP systems, it could be used as a preprocessing step which facilitates and improves the performance of parsers, semantic role labellers and machine translation systems. It should also be of use for people with reading disabilities because it allows the conversion of longer sentences into shorter ones. This paper makes two contributions towards this new task. First, we create and make available a benchmark consisting of 1,066,115 tuples mapping a single complex sentence to a sequence of sentences expressing the same meaning. Second, we propose five models (vanilla sequence-to-sequence to semantically-motivated models) to understand the difficulty of the proposed task.

Neural Response Generation via GAN with an Approximate Embedding Layer
Zhen Xu, Bingquan Liu, Baoxun Wang, Chengjie Sun, Xiaolong Wang, Zhuoran Wang, and Chao Qi 16:15–16:40

This paper presents a Generative Adversarial Network (GAN) to model single-turn short-text conversations, which trains a sequence-to-sequence (Seq2Seq) network for response generation simultaneously with a discriminative classifier that measures the differences between human-produced responses and machine-generated ones. In addition, the proposed method introduces an approximate embedding layer to solve the non-differentiable problem caused by the sampling-based output decoding procedure in the Seq2Seq generative model. The GAN setup provides an effective way to avoid noninformative responses (a.k.a "safe responses"), which are frequently observed in traditional neural response generators. The experimental results show that the proposed approach significantly outperforms existing neural response generation models in diversity metrics, with slight increases in relevance scores as well, when evaluated on both a Mandarin corpus and an English corpus.

A Hybrid Convolutional Variational Autoencoder for Text Generation
Stanislau Semeniuta, Aliaksei Severyn, and Erhardt Barth 16:40–17:05

In this paper we explore the effect of architectural choices on learning a variational autoencoder (VAE) for text generation. In contrast to the previously introduced VAE model for text where both the encoder and decoder are RNNs, we propose a novel hybrid architecture that blends fully feed-forward convolutional and deconvolutional components with a recurrent language model. Our architecture exhibits several attractive properties such as faster run time and convergence, ability to better handle long sequences and, more importantly, it helps to avoid the issue of the VAE collapsing to a deterministic model.

Filling the Blanks (hint: plural noun) for Mad Libs Humor
Nabil Hossain, John Krumm, Lucy Vanderwende, Eric Horvitz, and Henry Kautz 17:05–17:30

Computerized generation of humor is a notoriously difficult AI problem. We develop an algorithm called Libitum that helps humans generate humor in a Mad Lib, which is a popular fill-in-the-blank game. The algorithm is based on a machine learned classifier that determines whether a potential fill-in word is funny in the context of the Mad Lib story. We use Amazon Mechanical Turk to create ground truth data and to judge humor for our classifier to mimic, and we make this data freely available. Our testing shows that Libitum successfully aids humans in filling in Mad Libs that are usually judged funnier than those filled in by humans with no computerized help. We go on to analyze why some words are better than others at making a Mad Lib funny.

Session 3C: Semantics 1

Zealand — Chair: *Felix Hill*

Measuring Thematic Fit with Distributional Feature Overlap
Enrico Santus, Emmanuele Chersoni, Alessandro Lenci, and Philippe Blache 15:50–16:15

In this paper, we introduce a new distributional method for modeling predicate-argument thematic fit judgments. We use a syntax-based DSM to build a prototypical representation of verb-specific roles: for every verb, we extract the most salient second order contexts for each of its roles (i.e. the most salient dimensions of typical role fillers), and then we compute thematic fit as a weighted overlap between the top features of candidate fillers and role prototypes. Our experiments show that our method consistently outperforms a baseline re-implementing a state-of-the-art system, and achieves better or comparable results to those reported in the literature for the other unsupervised systems. Moreover, it provides an explicit representation of the features characterizing verb-specific semantic roles.

SCDV : Sparse Composite Document Vectors using soft clustering over distributional representations
Dheeraj Mekala, Vivek Gupta, Bhargavi Paranjape, and Harish Karnick 16:15–16:40

We present a feature vector formation technique for documents - Sparse Composite Document Vector (SCDV) - which overcomes several shortcomings of the current distributional paragraph vector representations that are widely used for text representation. In SCDV, word embeddings are clustered to capture multiple semantic contexts in which words occur. They are then chained together to form document topic-vectors that can express complex, multi-topic documents. Through extensive experiments on multi-class and multi-label classification tasks, we outperform the previous state-of-the-art method, NTSG. We also show that SCDV embeddings perform well on heterogeneous tasks like Topic Coherence, context-sensitive Learning and Information Retrieval. Moreover, we achieve a significant reduction in training and prediction times compared to other representation methods. SCDV achieves best of both worlds - better performance with lower time and space complexity.

Supervised Learning of Universal Sentence Representations from Natural Language Inference Data
Alexis Conneau, Douwe Kiela, Holger Schwenk, Loïc Barrault, and Antoine Bordes 16:40–17:05

Many modern NLP systems rely on word embeddings, previously trained in an unsupervised manner on large corpora, as base features. Efforts to obtain embeddings for larger chunks of text, such as sentences, have however not been so successful. Several attempts at learning unsupervised representations of sentences have not reached satisfactory enough performance to be widely adopted. In this paper, we show how universal sentence representations trained using the supervised data of the Stanford Natural Language Inference datasets can consistently outperform unsupervised methods like SkipThought vectors on a wide range of transfer tasks. Much like how computer vision uses ImageNet to obtain features, which can then be transferred to other tasks, our work tends to indicate the suitability of natural language inference for transfer learning to other NLP tasks. Our encoder is publicly available.

Determining Semantic Textual Similarity using Natural Deduction Proofs
Hitomi Yanaka, Koji Mineshima, Pascual Martínez-Gómez, and Daisuke Bekki 17:05–17:30

Determining semantic textual similarity is a core research subject in natural language processing. Since vector-based models for sentence representation often use shallow information, capturing accurate semantics is difficult. By contrast, logical semantic representations capture deeper levels of sentence semantics, but their symbolic nature does not offer graded notions of textual similarity. We propose a method for determining semantic textual similarity by combining shallow features with features extracted from natural deduction proofs of bidirectional entailment relations between sentence pairs. For the natural deduction proofs, we use ccg2lambda, a higher-order automatic inference system, which converts Combinatory Categorial Grammar (CCG) derivation trees into semantic representations and conducts natural deduction proofs. Experiments show that our system was able to outperform other logic-based systems and that features derived from the proofs are effective for learning textual similarity.

Session 3D: Poster Session. Syntax 2

15:50–17:30
Aarhus
Chair: *Yuval Pinter*

Multi-Grained Chinese Word Segmentation
Chen Gong, Zhenghua Li, Min Zhang, and Xinzhou Jiang

Traditionally, word segmentation (WS) adopts the single-grained formalism, where a sentence corresponds to a single word sequence. However, Sproat et al. (1997) show that the inter-native-speaker consistency ratio over Chinese word boundaries is only 76%, indicating single-grained WS (SWS) imposes unnecessary challenges on both manual annotation and statistical modeling. Moreover, WS results of different granularities can be complementary and beneficial for high-level applications. This work proposes and addresses multi-grained WS (MWS). We build a large-scale pseudo MWS dataset for model training and tuning by leveraging the annotation heterogeneity of three SWS datasets. Then we manually annotate 1,500 test sentences with true MWS annotations. Finally, we propose three benchmark approaches by casting MWS as constituent parsing and sequence labeling. Experiments and analysis lead to many interesting findings.

Don't Throw Those Morphological Analyzers Away Just Yet: Neural Morphological Disambiguation for Arabic
Nasser Zalmout and Nizar Habash

This paper presents a model for Arabic morphological disambiguation based on Recurrent Neural Networks (RNN). We train Long Short-Term Memory (LSTM) cells in several configurations and embedding levels to model the various morphological features. Our experiments show that these models outperform state-of-the-art systems without explicit use of feature engineering. However, adding learning features from a morphological analyzer to model the space of possible analyses provides additional improvement. We make use of the resulting morphological models for scoring and ranking the analyses of the morphological analyzer for morphological disambiguation. The results show significant gains in accuracy across several evaluation metrics. Our system results in 4.4% absolute increase over the state-of-the-art in full morphological analysis accuracy (30.6% relative error reduction), and 10.6% (31.5% relative error reduction) for out-of-vocabulary words.

Paradigm Completion for Derivational Morphology
Ryan Cotterell, Ekaterina Vylomova, Huda Khayrallah, Christo Kirov, and David Yarowsky

The generation of complex derived word forms has been an overlooked problem in NLP; we fill this gap by applying neural sequence-to-sequence models to the task. We overview the theoretical motivation for a paradigmatic treatment of derivational morphology, and introduce the task of derivational paradigm completion as a parallel to inflectional paradigm completion. State-of-the-art neural models adapted from the inflection task are able to learn the range of derivation patterns, and outperform a non-neural baseline by 16.4%. However, due to semantic, historical, and lexical considerations involved in derivational morphology, future work will be needed to achieve performance parity with inflection-generating systems.

A Sub-Character Architecture for Korean Language Processing
Karl Stratos

We introduce a novel sub-character architecture that exploits a unique compositional structure of the Korean language. Our method decomposes each character into a small set of primitive phonetic units called jamo letters from which character- and word-level representations are induced. The jamo letters divulge syntactic and semantic information that is difficult to access with conventional character-level units. They greatly alleviate the data sparsity problem, reducing the observation space to 1.6% of the original while increasing accuracy in our experiments. We apply our architecture to dependency parsing and achieve dramatic improvement over strong lexical baselines.

Do LSTMs really work so well for PoS tagging? – A replication study
Tobias Horsmann and Torsten Zesch

A recent study by Plank et al. (2016) found that LSTM-based PoS taggers considerably improve over the current state-of-the-art when evaluated on the corpora of the Universal Dependencies project that use a coarse-grained tagset. We replicate this study using a fresh collection of 27 corpora of 21 languages that are annotated with fine-grained tagsets of varying size. Our replication confirms the result in general, and we additionally find that the advantage of LSTMs is even bigger for larger tagsets. However, we also find that for the very large tagsets of morphologically rich languages, hand-crafted morphological lexicons are still necessary to reach state-of-the-art performance.

The Labeled Segmentation of Printed Books
Lara McConnaughey, Jennifer Dai, and David Bamman

We introduce the task of book structure labeling: segmenting and assigning a fixed category (such as Table of Contents, Preface, Index) to the document structure of printed books. We manually annotate the page-level structural categories for a large dataset totaling 294,816 pages in 1,055 books evenly sampled from 1750-1922, and present empirical results comparing the performance of several classes of models. The best-performing model, a bidirectional LSTM with rich features, achieves an overall accuracy of 95.8 and a class-balanced macro F-score of 71.4.

Cross-lingual Character-Level Neural Morphological Tagging
Ryan Cotterell and Georg Heigold

Even for common NLP tasks, sufficient supervision is not available in many languages – morphological tagging is no exception. In the work presented here, we explore a transfer learning scheme, whereby we train character-level recurrent neural taggers to predict morphological taggings for high-resource languages and low-resource languages together. Learning joint character representations among multiple related languages successfully enables knowledge transfer from the high-resource languages to the low-resource ones.

Word-Context Character Embeddings for Chinese Word Segmentation
Hao Zhou, Zhenting Yu, Yue Zhang, Shujian Huang, Xin-Yu Dai, and Jiajun Chen

Neural parsers have benefited from automatically labeled data via dependency-context word embeddings. We investigate training character embeddings on a word-based context in a similar way, showing that the simple method improves state-of-the-art neural word segmentation models significantly, beating tri-training baselines for leveraging auto-segmented data.

Segmentation-Free Word Embedding for Unsegmented Languages
Takamasa Oshikiri

In this paper, we propose a new pipeline of word embedding for unsegmented languages, called segmentation-free word embedding, which does not require word segmentation as a preprocessing step. Unlike space-delimited languages, unsegmented languages, such as Chinese and Japanese, require word segmentation as a preprocessing step. However, word segmentation, that often requires manually annotated resources, is difficult and expensive, and unavoidable errors in word segmentation affect downstream tasks. To avoid these problems in learning word vectors of unsegmented languages, we consider word co-occurrence statistics over all possible candidates of segmentations based on frequent character n-grams instead of segmented sentences provided by conventional word segmenters. Our experiments of noun category prediction tasks on raw Twitter, Weibo, and Wikipedia corpora show that the proposed method outperforms the conventional approaches that require word segmenters.

Session 3E: Poster Session. Question Answering and Machine Comprehension

15:50–17:30
Odense
Chair: *Jay Pujara*

From Textbooks to Knowledge: A Case Study in Harvesting Axiomatic Knowledge from Textbooks to Solve Geometry Problems
Mrinmaya Sachan, Kumar Dubey, and Eric Xing

Textbooks are rich sources of information. Harvesting structured knowledge from textbooks is a key challenge in many educational applications. As a case study, we present an approach for harvesting structured axiomatic knowledge from math textbooks. Our approach uses rich contextual and typographical features extracted from raw textbooks. It leverages the redundancy and shared ordering across multiple textbooks to further refine the harvested axioms. These axioms are then parsed into rules that are used to improve the state-of-the-art in solving geometry problems.

RACE: Large-scale ReAding Comprehension Dataset From Examinations
Guokun Lai, Qizhe Xie, Hanxiao Liu, Yiming Yang, and Eduard Hovy

We present RACE, a new dataset for benchmark evaluation of methods in the reading comprehension task. Collected from the English exams for middle and high school Chinese students in the age range between 12 to 18, RACE consists of near 28,000 passages and near 100,000 questions generated by human experts (English instructors), and covers a variety of topics which are carefully designed for evaluating the students' ability in understanding and reasoning. In particular, the proportion of questions that requires reasoning is much larger in RACE than that in other benchmark datasets for reading comprehension, and there is a significant gap between the performance of the state-of-the-art models (43%) and the ceiling human performance (95%). We hope this new dataset can serve as a valuable resource for research and evaluation in machine comprehension. The dataset is freely available at http://www.cs.cmu.edu/~glai1/data/race/ and the code is available at https://github.com/qizhex/RACE_AR_baselines.

Beyond Sentential Semantic Parsing: Tackling the Math SAT with a Cascade of Tree Transducers
Mark Hopkins, Cristian Petrescu-Prahova, Roie Levin, Ronan Le Bras, Alvaro Herrasti, and Vidur Joshi

We present an approach for answering questions that span multiple sentences and exhibit sophisticated cross-sentence anaphoric phenomena, evaluating on a rich source of such questions – the math portion of the Scholastic Aptitude Test (SAT). By using a tree transducer cascade as its basic architecture, our system propagates uncertainty from multiple sources (e.g. coreference resolution or verb interpretation) until it can be confidently resolved. Experiments show the first-ever results 43% recall and 91% precision) on SAT algebra word problems. We also apply our system to the public Dolphin algebra question set, and improve the state-of-the-art F1-score from 73.9% to 77.0%.

Learning Fine-Grained Expressions to Solve Math Word Problems
Danqing Huang, Shuming Shi, Chin-Yew Lin, and Jian Yin

This paper presents a novel template-based method to solve math word problems. This method learns the mappings between math concept phrases in math word problems and their math expressions from training data. For each equation template, we automatically construct a rich template sketch by aggregating information from various problems with the same template. Our approach is implemented in a two-stage system. It first retrieves a few relevant equation system templates and aligns numbers in math word problems to those templates for candidate equation generation. It then does a fine-grained inference to obtain the final answer. Experiment results show that our method achieves an accuracy of 28.4% on the linear Dolphin18K benchmark, which is 10% (54% relative) higher than previous state-of-the-art systems while achieving an accuracy increase of 12% (59% relative) on the TS6 benchmark subset.

Structural Embedding of Syntactic Trees for Machine Comprehension
Rui Liu, Junjie Hu, Wei Wei, Zi Yang, and Eric Nyberg

Deep neural networks for machine comprehension typically utilizes only word or character embeddings without explicitly taking advantage of structured linguistic information such as constituency trees and dependency trees. In this paper, we propose structural embedding of syntactic trees (SEST), an algorithm framework to utilize structured information and encode them into vector representations that can boost the performance of algorithms for the machine comprehension. We evaluate our approach using a state-

of-the-art neural attention model on the SQuAD dataset. Experimental results demonstrate that our model can accurately identify the syntactic boundaries of the sentences and extract answers that are syntactically coherent over the baseline methods.

World Knowledge for Reading Comprehension: Rare Entity Prediction with Hierarchical LSTMs Using External Descriptions
Teng Long, Emmanuel Bengio, Ryan Lowe, Jackie Chi Kit Cheung, and Doina Precup

Humans interpret texts with respect to some background information, or world knowledge, and we would like to develop automatic reading comprehension systems that can do the same. In this paper, we introduce a task and several models to drive progress towards this goal. In particular, we propose the task of rare entity prediction: given a web document with several entities removed, models are tasked with predicting the correct missing entities conditioned on the document context and the lexical resources. This task is challenging due to the diversity of language styles and the extremely large number of rare entities. We propose two recurrent neural network architectures which make use of external knowledge in the form of entity descriptions. Our experiments show that our hierarchical LSTM model performs significantly better at the rare entity prediction task than those that do not make use of external resources.

Two-Stage Synthesis Networks for Transfer Learning in Machine Comprehension
David Golub, Po-Sen Huang, Xiaodong He, and Li Deng

We develop a technique for transfer learning in machine comprehension (MC) using a novel two-stage synthesis network. Given a high performing MC model in one domain, our technique aims to answer questions about documents in another domain, where we use no labeled data of question-answer pairs. Using the proposed synthesis network with a pretrained model on the SQuAD dataset, we achieve an F1 measure of 46.6% on the challenging NewsQA dataset, approaching performance of in-domain models (F1 measure of 50.0%) and outperforming the out-of-domain baseline by 7.6%, without use of provided annotations.

Deep Neural Solver for Math Word Problems
Yan Wang, Xiaojiang Liu, and Shuming Shi

This paper presents a deep neural solver to automatically solve math word problems. In contrast to previous statistical learning approaches, we directly translate math word problems to equation templates using a recurrent neural network (RNN) model, without sophisticated feature engineering. We further design a hybrid model that combines the RNN model and a similarity-based retrieval model to achieve additional performance improvement. Experiments conducted on a large dataset show that the RNN model and the hybrid model significantly outperform state-of-the-art statistical learning methods for math word problem solving.

Latent Space Embedding for Retrieval in Question-Answer Archives
Deepak P, Dinesh Garg, and Shirish Shevade

Community-driven Question Answering (CQA) systems such as Yahoo! Answers have become valuable sources of reusable information. CQA retrieval enables usage of historical CQA archives to solve new questions posed by users. This task has received much recent attention, with methods building upon literature from translation models, topic models, and deep learning. In this paper, we devise a CQA retrieval technique, LASER-QA, that embeds question-answer pairs within a unified latent space preserving the local neighborhood structure of question and answer spaces. The idea is that such a space mirrors semantic similarity among questions as well as answers, thereby enabling high quality retrieval. Through an empirical analysis on various real-world QA datasets, we illustrate the improved effectiveness of LASER-QA over state-of-the-art methods.

Question Generation for Question Answering
Nan Duan, Duyu Tang, Peng Chen, and Ming Zhou

This paper presents how to generate questions from given passages using neural networks, where large scale QA pairs are automatically crawled and processed from Community-QA website, and used as training data. The contribution of the paper is 2-fold: First, two types of question generation approaches are proposed, one is a retrieval-based method using convolution neural network (CNN), the other is a generation-based method using recurrent neural network (RNN); Second, we show how to leverage the generated questions to improve existing question answering systems. We evaluate our question generation method for the answer sentence selection task on three benchmark datasets, including SQuAD, MS MARCO, and WikiQA. Experimental results show that, by using generated questions as an extra signal,

significant QA improvement can be achieved.

Learning to Paraphrase for Question Answering
Li Dong, Jonathan Mallinson, Siva Reddy, and Mirella Lapata

Question answering (QA) systems are sensitive to the many different ways natural language expresses the same information need. In this paper we turn to paraphrases as a means of capturing this knowledge and present a general framework which learns felicitous paraphrases for various QA tasks. Our method is trained end-to-end using question-answer pairs as a supervision signal. A question and its paraphrases serve as input to a neural scoring model which assigns higher weights to linguistic expressions most likely to yield correct answers. We evaluate our approach on QA over Freebase and answer sentence selection. Experimental results on three datasets show that our framework consistently improves performance, achieving competitive results despite the use of simple QA models.

Temporal Information Extraction for Question Answering Using Syntactic Dependencies in an LSTM-based Architecture
Yuanliang Meng, Anna Rumshisky, and Alexey Romanov

In this paper, we propose to use a set of simple, uniform in architecture LSTM-based models to recover different kinds of temporal relations from text. Using the shortest dependency path between entities as input, the same architecture is used to extract intra-sentence, cross-sentence, and document creation time relations. A "double-checking" technique reverses entity pairs in classification, boosting the recall of positive cases and reducing misclassifications between opposite classes. An efficient pruning algorithm resolves conflicts globally. Evaluated on QA-TempEval (SemEval2015 Task 5), our proposed technique outperforms state-of-the-art methods by a large margin. We also conduct intrinsic evaluation and post state-of-the-art results on Timebank-Dense.

Ranking Kernels for Structures and Embeddings: A Hybrid Preference and Classification Model
Kateryna Tymoshenko, Daniele Bonadiman, and Alessandro Moschitti

Recent work has shown that Tree Kernels (TKs) and Convolutional Neural Networks (CNNs) obtain the state of the art in answer sentence reranking. Additionally, their combination used in Support Vector Machines (SVMs) is promising as it can exploit both the syntactic patterns captured by TKs and the embeddings learned by CNNs. However, the embeddings are constructed according to a classification function, which is not directly exploitable in the preference ranking algorithm of SVMs. In this work, we propose a new hybrid approach combining preference ranking applied to TKs and pointwise ranking applied to CNNs. We show that our approach produces better results on two well-known and rather different datasets: WikiQA for answer sentence selection and SemEval cQA for comment selection in Community Question Answering.

Recovering Question Answering Errors via Query Revision
Semih Yavuz, Izzeddin Gur, Yu Su, and Xifeng Yan

The existing factoid QA systems often lack a post-inspection component that can help models recover from their own mistakes. In this work, we propose to crosscheck the corresponding KB relations behind the predicted answers and identify potential inconsistencies. Instead of developing a new model that accepts evidences collected from these relations, we choose to plug them back to the original questions directly and check if the revised question makes sense or not. A bidirectional LSTM is applied to encode revised questions. We develop a scoring mechanism over the revised question encodings to refine the predictions of a base QA system. This approach can improve the F1 score of STAGG (Yih et al., 2015), one of the leading QA systems, from 52.5% to 53.9% on WEBQUESTIONS data.

Session 3F: Poster Session. Multimodal NLP 1

15:50–17:30
Copenhagen
Chair: *Tianze Shi*

An empirical study on the effectiveness of images in Multimodal Neural Machine Translation
Jean-Benoit Delbrouck and Stéphane Dupont

In state-of-the-art Neural Machine Translation (NMT), an attention mechanism is used during decoding to enhance the translation. At every step, the decoder uses this mechanism to focus on different parts of the source sentence to gather the most useful information before outputting its target word. Recently, the effectiveness of the attention mechanism has also been explored for multi-modal tasks, where it becomes possible to focus both on sentence parts and image regions that they describe. In this paper, we compare several attention mechanism on the multi-modal translation task (English, image → German) and evaluate the ability of the model to make use of images to improve translation. We surpass state-of-the-art scores on the Multi30k data set, we nevertheless identify and report different misbehavior of the machine while translating.

Sound-Word2Vec: Learning Word Representations Grounded in Sounds
Ashwin Vijayakumar, Ramakrishna Vedantam, and Devi Parikh

To be able to interact better with humans, it is crucial for machines to understand sound — a primary modality of human perception. Previous works have used sound to learn embeddings for improved generic semantic similarity assessment. In this work, we treat sound as a first-class citizen, studying downstream 6 textual tasks which require aural grounding. To this end, we propose sound-word2vec — a new embedding scheme that learns specialized word embeddings grounded in sounds. For example, we learn that two seemingly (semantically) unrelated concepts, like leaves and paper are similar due to the similar rustling sounds they make. Our embeddings prove useful in textual tasks requiring aural reasoning like text-based sound retrieval and discovering Foley sound effects (used in movies). Moreover, our embedding space captures interesting dependencies between words and onomatopoeia and outperforms prior work on aurally-relevant word relatedness datasets such as AMEN and ASLex.

The Promise of Premise: Harnessing Question Premises in Visual Question Answering
Aroma Mahendru, Viraj Prabhu, Akrit Mohapatra, Dhruv Batra, and Stefan Lee

In this paper, we make a simple observation that questions about images often contain premises — objects and relationships implied by the question — and that reasoning about premises can help Visual Question Answering (VQA) models respond more intelligently to irrelevant or previously unseen questions. When presented with a question that is irrelevant to an image, state-of-the-art VQA models will still answer purely based on learned language biases, resulting in non-sensical or even misleading answers. We note that a visual question is irrelevant to an image if at least one of its premises is false (i.e. not depicted in the image). We leverage this observation to construct a dataset for Question Relevance Prediction and Explanation (QRPE) by searching for false premises. We train novel question relevance detection models and show that models that reason about premises consistently outperform models that do not. We also find that forcing standard VQA models to reason about premises during training can lead to improvements on tasks requiring compositional reasoning.

Guided Open Vocabulary Image Captioning with Constrained Beam Search
Peter Anderson, Basura Fernando, Mark Johnson, and Stephen Gould

Existing image captioning models do not generalize well to out-of-domain images containing novel scenes or objects. This limitation severely hinders the use of these models in real world applications dealing with images in the wild. We address this problem using a flexible approach that enables existing deep captioning architectures to take advantage of image taggers at test time, without re-training. Our method uses constrained beam search to force the inclusion of selected tag words in the output, and fixed, pretrained word embeddings to facilitate vocabulary expansion to previously unseen tag words. Using this approach we achieve state of the art results for out-of-domain captioning on MSCOCO (and improved results for in-domain captioning). Perhaps surprisingly, our results significantly outperform approaches that incorporate the same tag predictions into the learning algorithm. We also show that we can significantly improve the quality of generated ImageNet captions by leveraging ground-truth labels.

Zero-Shot Activity Recognition with Verb Attribute Induction
Rowan Zellers and Yejin Choi

In this paper, we investigate large-scale zero-shot activity recognition by modeling the visual and linguistic attributes of action verbs. For example, the verb "salute" has several properties, such as being a light movement, a social act, and short in duration. We use these attributes as the internal mapping between visual and textual representations to reason about a previously unseen action. In contrast to much prior work that assumes access to gold standard attributes for zero-shot classes and focuses primarily on object attributes, our model uniquely learns to infer action attributes from dictionary definitions and distributed word representations. Experimental results confirm that action attributes inferred from language can provide a predictive signal for zero-shot prediction of previously unseen activities.

Deriving continous grounded meaning representations from referentially structured multimodal contexts
Sina Zarrieß and David Schlangen

Corpora of referring expressions paired with their visual referents are a good source for learning word meanings directly grounded in visual representations. Here, we explore additional ways of extracting from them word representations linked to multi-modal context: through expressions that refer to the same object, and through expressions that refer to different objects in the same scene. We show that continuous meaning representations derived from these contexts capture complementary aspects of similarity, , even if not outperforming textual embeddings trained on very large amounts of raw text when tested on standard similarity benchmarks. We propose a new task for evaluating grounded meaning representations—detection of potentially co-referential phrases—and show that it requires precise denotational representations of attribute meanings, which our method provides.

Hierarchically-Attentive RNN for Album Summarization and Storytelling
Licheng Yu, Mohit Bansal, and Tamara Berg

We address the problem of end-to-end visual storytelling. Given a photo album, our model first selects the most representative (summary) photos, and then composes a natural language story for the album. For this task, we make use of the Visual Storytelling dataset and a model composed of three hierarchically-attentive Recurrent Neural Nets (RNNs) to: encode the album photos, select representative (summary) photos, and compose the story. Automatic and human evaluations show our model achieves better performance on selection, generation, and retrieval than baselines.

Video Highlight Prediction Using Audience Chat Reactions
Cheng-Yang Fu, Joon Lee, Mohit Bansal, and Alexander Berg

Sports channel video portals offer an exciting domain for research on multimodal, multilingual analysis. We present methods addressing the problem of automatic video highlight prediction based on joint visual features and textual analysis of the real-world audience discourse with complex slang, in both English and traditional Chinese. We present a novel dataset based on League of Legends championships recorded from North American and Taiwanese Twitch.tv channels (will be released for further research), and demonstrate strong results on these using multimodal, character-level CNN-RNN model architectures.

Reinforced Video Captioning with Entailment Rewards
Ramakanth Pasunuru and Mohit Bansal

Sequence-to-sequence models have shown promising improvements on the temporal task of video captioning, but they optimize word-level cross-entropy loss during training. First, using policy gradient and mixed-loss methods for reinforcement learning, we directly optimize sentence-level task-based metrics (as rewards), achieving significant improvements over the baseline, based on both automatic metrics and human evaluation on multiple datasets. Next, we propose a novel entailment-enhanced reward (CIDEnt) that corrects phrase-matching based metrics (such as CIDEr) to only allow for logically-implied partial matches and avoid contradictions, achieving further significant improvements over the CIDEr-reward model. Overall, our CIDEnt-reward model achieves the new state-of-the-art on the MSR-VTT dataset.

Evaluating Hierarchies of Verb Argument Structure with Hierarchical Clustering
Jesse Mu, Joshua K. Hartshorne, and Timothy O'Donnell

Verbs can only be used with a few specific arrangements of their arguments (syntactic frames). Most theorists note that verbs can be organized into a hierarchy of verb classes based on the frames they admit. Here we show that such a hierarchy is objectively well-supported by the patterns of verbs and frames in English, since a systematic hierarchical clustering algorithm converges on the same structure as the handcrafted taxonomy of VerbNet, a broad-coverage verb lexicon. We also show that the hierarchies capture meaningful psychological dimensions of generalization by predicting novel verb coercions by

human participants. We discuss limitations of a simple hierarchical representation and suggest similar approaches for identifying the representations underpinning verb argument structure.

Incorporating Global Visual Features into Attention-based Neural Machine Translation.
Iacer Calixto and Qun Liu

We introduce multi-modal, attention-based neural machine translation (NMT) models which incorporate visual features into different parts of both the encoder and the decoder. Global image features are extracted using a pre-trained convolutional neural network and are incorporated (i) as words in the source sentence, (ii) to initialise the encoder hidden state, and (iii) as additional data to initialise the decoder hidden state. In our experiments, we evaluate translations into English and German, how different strategies to incorporate global image features compare and which ones perform best. We also study the impact that adding synthetic multi-modal, multilingual data brings and find that the additional data have a positive impact on multi-modal NMT models. We report new state-of-the-art results and our best models also significantly improve on a comparable phrase-based Statistical MT (PBSMT) model trained on the Multi30k data set according to all metrics evaluated. To the best of our knowledge, it is the first time a purely neural model significantly improves over a PBSMT model on all metrics evaluated on this data set.

Mapping Instructions and Visual Observations to Actions with Reinforcement Learning
Dipendra Misra, John Langford, and Yoav Artzi

We propose to directly map raw visual observations and text input to actions for instruction execution. While existing approaches assume access to structured environment representations or use a pipeline of separately trained models, we learn a single model to jointly reason about linguistic and visual input. We use reinforcement learning in a contextual bandit setting to train a neural network agent. To guide the agent's exploration, we use reward shaping with different forms of supervision. Our approach does not require intermediate representations, planning procedures, or training different models. We evaluate in a simulated environment, and show significant improvements over supervised learning and common reinforcement learning variants.

An analysis of eye-movements during reading for the detection of mild cognitive impairment
Kathleen C. Fraser, Kristina Lundholm Fors, Dimitrios Kokkinakis, and Arto Nordlund

We present a machine learning analysis of eye-tracking data for the detection of mild cognitive impairment, a decline in cognitive abilities that is associated with an increased risk of developing dementia. We compare two experimental configurations (reading aloud versus reading silently), as well as two methods of combining information from the two trials (concatenation and merging). Additionally, we annotate the words being read with information about their frequency and syntactic category, and use these annotations to generate new features. Ultimately, we are able to distinguish between participants with and without cognitive impairment with up to 86% accuracy.

[TACL] Evaluating Low-Level Speech Features Against Human Perceptual Data
Naomi H Feldman, Caitlin Richter, Harini Salgado, and Aren Jansen

We introduce a method for measuring the correspondence between low-level speech features and human perception, using a cognitive model of speech perception implemented directly on speech recordings. We evaluate two speaker normalization techniques using this method and find that in both cases, speech features that are normalized across speakers predict human data better than unnormalized speech features, consistent with previous research. Results further reveal differences across normalization methods in how well each predicts human data. This work provides a new framework for evaluating low-level representations of speech on their match to human perception, and lays the groundwork for creating more ecologically valid models of speech perception.

6

Main Conference: Sunday, September 10

Overview

07:30 – 17:30 **Registration Day 2** *Foyer*

08:00 – 09:00 **Morning Coffee** *Foyer*

09:00 – 10:00 **Invited Talk.** *Jutland*

Towards more universal language technology: unsupervised learning from speech (Sharon Goldwater)

10:00 – 10:30 **Coffee Break** *Foyer*

Session 4

10:30 – 12:10

Reading and Retrieving *Jutland*	Multimodal NLP 2 *Funen*	Human Centered NLP and Linguistic Theory *Zealand*
Poster Session. Semantics 2 *Aarhus*	Poster Session. Discourse *Odense*	Poster Session. Machine Translation and Multilingual NLP 1 *Copenhagen*

12:10 – 13:40 **Lunch**

12:40 – 13:40 **SIGDAT Business Meeting** *Jutland*

Session 5

13:40 – 15:20

Semantics 3 *Jutland*	Computational Social Science 1 *Funen*	Sentiment Analysis 2 *Zealand*
Poster Session. Syntax 3 *Aarhus*	Poster Session. Relations *Odense*	Poster Session. Language Models, Text Mining, and Crowd Sourcing *Copenhagen*

15:20 – 15:50 **Coffee Break** *Foyer*

Session 6

15:50 – 17:30	Machine Translation 2 *Jutland*	Text Mining and NLP applications *Funen*	Machine Comprehension *Zealand*
	Poster Session. Summarization, Generation, Dialog, and Discourse 1 *Aarhus*	Poster Session. Summarization, Generation, Dialog, and Discourse 2 *Odense*	Poster Session. Computational Social Science 2 *Copenhagen*

18:00 – 22:00 **Social Event** *Øksnehallen Courtyard*

Invited Talk: Sharon Goldwater

Towards more universal language technology: unsupervised learning from speech

Sunday, September 10, 2017, 9:00–10:00

Jutland

Abstract: Speech and language processing has advanced enormously in the last decade, with successful applications in machine translation, voice-activated search, and even language-enabled personal assistants. Yet these systems typically still rely on learning from very large quantities of human-annotated data. These resource-intensive methods mean that effective technology is available for only a tiny fraction of the world's 7000 or so languages, mainly those spoken in large rich countries.

This talk describes our recent work on developing *unsupervised* speech technology, where transcripts and pronunciation dictionaries are not used. The work is inspired by considering both how young infants may begin to acquire the sounds and words of their language, and how we might develop systems to help linguists analyze and document endangered languages. I will first present work on learning from speech audio alone, where the system must learn to segment the speech stream into word tokens and cluster repeated instances of the same word together to learn a lexicon of vocabulary items. The approach combines Bayesian and neural network methods to address learning at the word and sub-word levels.

Biography: Sharon Goldwater is a Reader at the University of Edinburgh's School of Informatics, where she is a member of the Institute for Language, Cognition and Computation. She received her PhD in 2007 from Brown University and spent two years as a postdoctoral researcher at Stanford University before moving to Edinburgh. Her research interests include unsupervised learning for speech and language processing, computer modelling of language acquisition in children, and computational studies of language use. Dr. Goldwater co-chaired the 2014 Conference of the European Chapter of the Association for Computational Linguistics and is Chair-Elect of EACL. She has served on the editorial boards of the Transactions of the Association for Computational Linguistics, the Computational Linguistics journal, and OPEN MIND: Advances in Cognitive Science (a new open-access journal). In 2016, she received the Roger Needham Award from the British Computer Society, awarded for "distinguished research contribution in computer science by a UK-based researcher who has completed up to 10 years of post-doctoral research."

Session 4 Overview – Sunday, September 10, 2017

Oral tracks

	Track A *Reading and Retrieving* Jutland	**Track B** *Multimodal NLP 2* Funen	**Track C** *Human Centered NLP and Linguistic Theory* Zealand
10:30	A Structured Learning Approach to Temporal Relation Extraction *Ning, Feng, and Roth*	Speech segmentation with a neural encoder model of working memory *Elsner and Shain*	ConStance: Modeling Annotation Contexts to Improve Stance Classification *Joseph, Friedland, Hobbs, Lazer, and Tsur*
10:55	Importance sampling for unbiased on-demand evaluation of knowledge base population *Chaganty, Paranjape, Liang, and Manning*	Speaking, Seeing, Understanding: Correlating semantic models with conceptual representation in the brain *Bulat, Clark, and Shutova*	Deeper Attention to Abusive User Content Moderation *Pavlopoulos, Malakasiotis, and Androutsopoulos*
11:20	PACRR: A Position-Aware Neural IR Model for Relevance Matching *Hui, Yates, Berberich, and Melo*	Multi-modal Summarization for Asynchronous Collection of Text, Image, Audio and Video *Li, Zhu, Ma, Zhang, and Zong*	Outta Control: Laws of Semantic Change and Inherent Biases in Word Representation Models *Dubossarsky, Weinshall, and Grossman*
11:45	Globally Normalized Reader *Raiman and Miller*	Tensor Fusion Network for Multimodal Sentiment Analysis *Zadeh, Chen, Poria, Cambria, and Morency*	Human Centered NLP with User-Factor Adaptation *Lynn, Son, Kulkarni, Balasubramanian, and Schwartz*

Poster tracks

10:30–12:10

Track D: *Poster Session. Semantics 2* — Aarhus

Track E: *Poster Session. Discourse* — Odense

Track F: *Poster Session. Machine Translation and Multilingual NLP 1* — Copenhagen

Parallel Session 4

Session 4A: Reading and Retrieving

Jutland Chair: *Heng Ji*

A Structured Learning Approach to Temporal Relation Extraction
Qiang Ning, Zhili Feng, and Dan Roth 10:30–10:55

Identifying temporal relations between events is an essential step towards natural language understanding. However, the temporal relation between two events in a story depends on, and is often dictated by, relations among other events. Consequently, effectively identifying temporal relations between events is a challenging problem even for human annotators. This paper suggests that it is important to take these dependencies into account while learning to identify these relations and proposes a structured learning approach to address this challenge. As a byproduct, this provides a new perspective on handling missing relations, a known issue that hurts existing methods. As we show, the proposed approach results in significant improvements on the two commonly used data sets for this problem.

Importance sampling for unbiased on-demand evaluation of knowledge base population
Arun Chaganty, Ashwin Paranjape, Percy Liang, and Christopher D. Manning 10:55–11:20

Knowledge base population (KBP) systems take in a large document corpus and extract entities and their relations. Thus far, KBP evaluation has relied on judgements on the pooled predictions of existing systems. We show that this evaluation is problematic: when a new system predicts a previously unseen relation, it is penalized even if it is correct. This leads to significant bias against new systems, which counterproductively discourages innovation in the field. Our first contribution is a new importance-sampling based evaluation which corrects for this bias by annotating a new system's predictions on-demand via crowdsourcing. We show this eliminates bias and reduces variance using data from the 2015 TAC KBP task. Our second contribution is an implementation of our method made publicly available as an online KBP evaluation service. We pilot the service by testing diverse state-of-the-art systems on the TAC KBP 2016 corpus and obtain accurate scores in a cost effective manner.

PACRR: A Position-Aware Neural IR Model for Relevance Matching
Kai Hui, Andrew Yates, Klaus Berberich, and Gerard de Melo 11:20–11:45

In order to adopt deep learning for information retrieval, models are needed that can capture all relevant information required to assess the relevance of a document to a given user query. While previous works have successfully captured unigram term matches, how to fully employ position-dependent information such as proximity and term dependencies has been insufficiently explored. In this work, we propose a novel neural IR model named PACRR aiming at better modeling position-dependent interactions between a query and a document. Extensive experiments on six years' TREC Web Track data confirm that the proposed model yields better results under multiple benchmarks.

Globally Normalized Reader
Jonathan Raiman and John Miller 11:45–12:10

Rapid progress has been made towards question answering (QA) systems that can extract answers from text. Existing neural approaches make use of expensive bi-directional attention mechanisms or score all possible answer spans, limiting scalability. We propose instead to cast extractive QA as an iterative search problem: select the answer's sentence, start word, and end word. This representation reduces the space of each search step and allows computation to be conditionally allocated to promising search paths. We show that globally normalizing the decision process and back-propagating through beam search makes this representation viable and learning efficient. We empirically demonstrate the benefits of this approach using our model, Globally Normalized Reader (GNR), which achieves the second highest single model performance on the Stanford Question Answering Dataset (68.4 EM, 76.21 F1 dev) and is 24.7x faster than bi-attention-flow. We also introduce a data-augmentation method to produce semantically valid examples by aligning named entities to a knowledge base and swapping them with new entities of the same type. This method improves the performance of all models considered in this work and is of independent interest for a variety of NLP tasks.

Session 4B: Multimodal NLP 2

Funen Chair: *Brian Roark*

Speech segmentation with a neural encoder model of working memory
Micha Elsner and Cory Shain 10:30–10:55

We present the first unsupervised LSTM speech segmenter as a cognitive model of the acquisition of words from unsegmented input. Cognitive biases toward phonological and syntactic predictability in speech are rooted in the limitations of human memory (Baddeley et al., 1998); compressed representations are easier to acquire and retain in memory. To model the biases introduced by these memory limitations, our system uses an LSTM-based encoder-decoder with a small number of hidden units, then searches for a segmentation that minimizes autoencoding loss. Linguistically meaningful segments (e.g. words) should share regular patterns of features that facilitate decoder performance in comparison to random segmentations, and we show that our learner discovers these patterns when trained on either phoneme sequences or raw acoustics. To our knowledge, ours is the first fully unsupervised system to be able to segment both symbolic and acoustic representations of speech.

Speaking, Seeing, Understanding: Correlating semantic models with conceptual representation in the brain
Luana Bulat, Stephen Clark, and Ekaterina Shutova 10:55–11:20

Research in computational semantics is increasingly guided by our understanding of human semantic processing. However, semantic models are typically studied in the context of natural language processing system performance. In this paper, we present a systematic evaluation and comparison of a range of widely-used, state-of-the-art semantic models in their ability to predict patterns of conceptual representation in the human brain. Our results provide new insights both for the design of computational semantic models and for further research in cognitive neuroscience.

Multi-modal Summarization for Asynchronous Collection of Text, Image, Audio and Video
Haoran Li, Junnan Zhu, Cong Ma, Jiajun Zhang, and Chengqing Zong 11:20–11:45

The rapid increase of the multimedia data over the Internet necessitates multi-modal summarization from collections of text, image, audio and video. In this work, we propose an extractive Multi-modal Summarization (MMS) method which can automatically generate a textual summary given a set of documents, images, audios and videos related to a specific topic. The key idea is to bridge the semantic gaps between multi-modal contents. For audio information, we design an approach to selectively use its transcription. For vision information, we learn joint representations of texts and images using a neural network. Finally, all the multi-modal aspects are considered to generate the textual summary by maximizing the salience, non-redundancy, readability and coverage through budgeted optimization of submodular functions. We further introduce an MMS corpus in English and Chinese. The experimental results on this dataset demonstrate that our method outperforms other competitive baseline methods.

Tensor Fusion Network for Multimodal Sentiment Analysis
Amir Zadeh, Minghai Chen, Soujanya Poria, Erik Cambria, and Louis-Philippe Morency 11:45–12:10

Multimodal sentiment analysis is an increasingly popular research area, which extends the conventional language-based definition of sentiment analysis to a multimodal setup where other relevant modalities accompany language. In this paper, we pose the problem of multimodal sentiment analysis as modeling intra-modality and inter-modality dynamics. We introduce a novel model, termed Tensor Fusion Networks, which learns both such dynamics end-to-end. The proposed approach is tailored for the volatile nature of spoken language in online videos as well as accompanying gestures and voice. In the experiments, our model outperforms state-of-the-art approaches for both multimodal and unimodal sentiment analysis.

Session 4C: Human Centered NLP and Linguistic Theory

Zealand Chair: *Alan Ritter*

ConStance: Modeling Annotation Contexts to Improve Stance Classification
Kenneth Joseph, Lisa Friedland, William Hobbs, David Lazer, and Oren Tsur 10:30–10:55

Manual annotations are a prerequisite for many applications of machine learning. However, weaknesses in the annotation process itself are easy to overlook. In particular, scholars often choose what information to give to annotators without examining these decisions empirically. For subjective tasks such as sentiment analysis, sarcasm, and stance detection, such choices can impact results. Here, for the task of political stance detection on Twitter, we show that providing too little context can result in noisy and uncertain annotations, whereas providing too strong a context may cause it to outweigh other signals. To characterize and reduce these biases, we develop ConStance, a general model for reasoning about annotations across information conditions. Given conflicting labels produced by multiple annotators seeing the same instances with different contexts, ConStance simultaneously estimates gold standard labels and also learns a classifier for new instances. We show that the classifier learned by ConStance outperforms a variety of baselines at predicting political stance, while the model's interpretable parameters shed light on the effects of each context.

Deeper Attention to Abusive User Content Moderation
John Pavlopoulos, Prodromos Malakasiotis, and Ion Androutsopoulos 10:55–11:20

Experimenting with a new dataset of 1.6M user comments from a news portal and an existing dataset of 115K Wikipedia talk page comments, we show that an RNN operating on word embeddings outpeforms the previous state of the art in moderation, which used logistic regression or an MLP classifier with character or word n-grams. We also compare against a CNN operating on word embeddings, and a word-list baseline. A novel, deep, classificationspecific attention mechanism improves the performance of the RNN further, and can also highlight suspicious words for free, without including highlighted words in the training data. We consider both fully automatic and semi-automatic moderation.

Outta Control: Laws of Semantic Change and Inherent Biases in Word Representation Models
Haim Dubossarsky, Daphna Weinshall, and Eitan Grossman 11:20–11:45

This article evaluates three proposed laws of semantic change. Our claim is that in order to validate a putative law of semantic change, the effect should be observed in the genuine condition but absent or reduced in a suitably matched control condition, in which no change can possibly have taken place. Our analysis shows that the effects reported in recent literature must be substantially revised: (i) the proposed negative correlation between meaning change and word frequency is shown to be largely an artefact of the models of word representation used; (ii) the proposed negative correlation between meaning change and prototypicality is shown to be much weaker than what has been claimed in prior art; and (iii) the proposed positive correlation between meaning change and polysemy is largely an artefact of word frequency. These empirical observations are corroborated by analytical proofs that show that count representations introduce an inherent dependence on word frequency, and thus word frequency cannot be evaluated as an independent factor with these representations.

Human Centered NLP with User-Factor Adaptation
Veronica Lynn, Youngseo Son, Vivek Kulkarni, Niranjan Balasubramanian, and H. Andrew Schwartz 11:45–12:10

We pose the general task of user-factor adaptation – adapting supervised learning models to real-valued user factors inferred from a background of their language, reflecting the idea that a piece of text should be understood within the context of the user that wrote it. We introduce a continuous adaptation technique, suited for real-valued user factors that are common in social science and bringing us closer to personalized NLP, adapting to each user uniquely. We apply this technique with known user factors including age, gender, and personality traits, as well as latent factors, evaluating over five tasks: POS tagging, PP-attachment, sentiment analysis, sarcasm detection, and stance detection. Adaptation provides statistically significant benefits for 3 of the 5 tasks: up to +1.2 points for PP-attachment, +3.4 points for sarcasm, and +3.0 points for stance.

Session 4D: Poster Session. Semantics 2

10:30–12:10
Aarhus
Chair: *Ivan Vulić*

Neural Sequence Learning Models for Word Sense Disambiguation
Alessandro Raganato, Claudio Delli Bovi, and Roberto Navigli

Word Sense Disambiguation models exist in many flavors. Even though supervised ones tend to perform best in terms of accuracy, they often lose ground to more flexible knowledge-based solutions, which do not require training by a word expert for every disambiguation target. To bridge this gap we adopt a different perspective and rely on sequence learning to frame the disambiguation problem: we propose and study in depth a series of end-to-end neural architectures directly tailored to the task, from bidirectional Long Short-Term Memory to encoder-decoder models. Our extensive evaluation over standard benchmarks and in multiple languages shows that sequence learning enables more versatile all-words models that consistently lead to state-of-the-art results, even against word experts with engineered features.

Learning Word Relatedness over Time
Guy D. Rosin, Eytan Adar, and Kira Radinsky

Search systems are often focused on providing relevant results for the "now", assuming both corpora and user needs that focus on the present. However, many corpora today reflect significant longitudinal collections ranging from 20 years of the Web to hundreds of years of digitized newspapers and books. Understanding the temporal intent of the user and retrieving the most relevant historical content has become a significant challenge. Common search features, such as query expansion, leverage the relationship between terms but cannot function well across all times when relationships vary temporally. In this work, we introduce a temporal relationship model that is extracted from longitudinal data collections. The model supports the task of identifying, given two words, when they relate to each other. We present an algorithmic framework for this task and show its application for the task of query expansion, achieving high gain.

Inter-Weighted Alignment Network for Sentence Pair Modeling
Gehui Shen, Yunlun Yang, and Zhi-Hong Deng

Sentence pair modeling is a crucial problem in the field of natural language processing. In this paper, we propose a model to measure the similarity of a sentence pair focusing on the interaction information. We utilize the word level similarity matrix to discover fine-grained alignment of two sentences. It should be emphasized that each word in a sentence has a different importance from the perspective of semantic composition, so we exploit two novel and efficient strategies to explicitly calculate a weight for each word. Although the proposed model only use a sequential LSTM for sentence modeling without any external resource such as syntactic parser tree and additional lexicon features, experimental results show that our model achieves state-of-the-art performance on three datasets of two tasks.

A Short Survey on Taxonomy Learning from Text Corpora: Issues, Resources and Recent Advances
Chengyu Wang, Xiaofeng He, and Aoying Zhou

A taxonomy is a semantic hierarchy, consisting of concepts linked by is-a relations. While a large number of taxonomies have been constructed from human-compiled resources (e.g., Wikipedia), learning taxonomies from text corpora has received a growing interest and is essential for long-tailed and domain-specific knowledge acquisition. In this paper, we overview recent advances on taxonomy construction from free texts, reorganizing relevant subtasks into a complete framework. We also overview resources for evaluation and discuss challenges for future research.

Idiom-Aware Compositional Distributed Semantics
Pengfei Liu, Kaiyu Qian, Xipeng Qiu, and Xuanjing Huang

Idioms are peculiar linguistic constructions that impose great challenges for representing the semantics of language, especially in current prevailing end-to-end neural models, which assume that the semantics of a phrase or sentence can be literally composed from its constitutive words. In this paper, we propose an idiom-aware distributed semantic model to build representation of sentences on the basis of understanding their contained idioms. Our models are grounded in the literal-first psycholinguistic hypothesis, which can adaptively learn semantic compositionality of a phrase literally or idiomatically. To better evaluate our models, we also construct an idiom-enriched sentiment classification dataset with considerable scale and abundant peculiarities of idioms. The qualitative and quantitative experimental analyses demonstrate the

efficacy of our models.

Macro Grammars and Holistic Triggering for Efficient Semantic Parsing
Yuchen Zhang, Panupong Pasupat, and Percy Liang

To learn a semantic parser from denotations, a learning algorithm must search over a combinatorially large space of logical forms for ones consistent with the annotated denotations. We propose a new online learning algorithm that searches faster as training progresses. The two key ideas are using macro grammars to cache the abstract patterns of useful logical forms found thus far, and holistic triggering to efficiently retrieve the most relevant patterns based on sentence similarity. On the WikiTableQuestions dataset, we first expand the search space of an existing model to improve the state-of-the-art accuracy from 38.7% to 42.7%, and then use macro grammars and holistic triggering to achieve an 11x speedup and an accuracy of 43.7%.

A Continuously Growing Dataset of Sentential Paraphrases
Wuwei Lan, Siyu Qiu, Hua He, and Wei Xu

A major challenge in paraphrase research is the lack of parallel corpora. In this paper, we present a new method to collect large-scale sentential paraphrases from Twitter by linking tweets through shared URLs. The main advantage of our method is its simplicity, as it gets rid of the classifier or human in the loop needed to select data before annotation and subsequent application of paraphrase identification algorithms in the previous work. We present the largest human-labeled paraphrase corpus to date of 51,524 sentence pairs and the first cross-domain benchmarking for automatic paraphrase identification. In addition, we show that more than 30,000 new sentential paraphrases can be easily and continuously captured every month at ~70% precision, and demonstrate their utility for downstream NLP tasks through phrasal paraphrase extraction. We make our code and data freely available.

Cross-domain Semantic Parsing via Paraphrasing
Yu Su and Xifeng Yan

Existing studies on semantic parsing mainly focus on the in-domain setting. We formulate cross-domain semantic parsing as a domain adaptation problem: train a semantic parser on some source domains and then adapt it to the target domain. Due to the diversity of logical forms in different domains, this problem presents unique and intriguing challenges. By converting logical forms into canonical utterances in natural language, we reduce semantic parsing to paraphrasing, and develop an attentive sequence-to-sequence paraphrase model that is general and flexible to adapt to different domains. We discover two problems, small micro variance and large macro variance, of pre-trained word embeddings that hinder their direct use in neural networks, and propose standardization techniques as a remedy. On the popular Overnight dataset, which contains eight domains, we show that both cross-domain training and standardized pre-trained word embeddings can bring significant improvement.

A Joint Sequential and Relational Model for Frame-Semantic Parsing
Bishan Yang and Tom Mitchell

We introduce a new method for frame-semantic parsing that significantly improves the prior state of the art. Our model leverages the advantages of a deep bidirectional LSTM network which predicts semantic role labels word by word and a relational network which predicts semantic roles for individual text expressions in relation to a predicate. The two networks are integrated into a single model via knowledge distillation, and a unified graphical model is employed to jointly decode frames and semantic roles during inference. Experiments on the standard FrameNet data show that our model significantly outperforms existing neural and non-neural approaches, achieving a 5.7 F1 gain over the current state of the art, for full frame structure extraction.

Getting the Most out of AMR Parsing
Chuan Wang and Nianwen Xue

This paper proposes to tackle the AMR parsing bottleneck by improving two components of an AMR parser: concept identification and alignment. We first build a Bidirectional LSTM based concept identifier that is able to incorporate richer contextual information to learn sparse AMR concept labels. We then extend an HMM-based word-to-concept alignment model with graph distance distortion and a rescoring method during decoding to incorporate the structural information in the AMR graph. We show integrating the two components into an existing AMR parser results in consistently better performance over the state of the art on various datasets.

AMR Parsing using Stack-LSTMs
Miguel Ballesteros and Yaser Al-Onaizan

We present a transition-based AMR parser that directly generates AMR parses from plain text. We use Stack-LSTMs to represent our parser state and make decisions greedily. In our experiments, we show that our parser achieves very competitive scores on English using only AMR training data. Adding additional information, such as POS tags and dependency trees, improves the results further.

An End-to-End Deep Framework for Answer Triggering with a Novel Group-Level Objective
Jie Zhao, Yu Su, Ziyu Guan, and Huan Sun

Given a question and a set of answer candidates, answer triggering determines whether the candidate set contains any correct answers. If yes, it then outputs a correct one. In contrast to existing pipeline methods which first consider individual candidate answers separately and then make a prediction based on a threshold, we propose an end-to-end deep neural network framework, which is trained by a novel group-level objective function that directly optimizes the answer triggering performance. Our objective function penalizes three potential types of error and allows training the framework in an end-to-end manner. Experimental results on the WikiQA benchmark show that our framework outperforms the state of the arts by a 6.6% absolute gain under F1 measure.

Predicting Word Association Strengths
Andrew Cattle and Xiaojuan Ma

This paper looks at the task of predicting word association strengths across three datasets; WordNet Evocation (Boyd-Graber et al., 2006), University of Southern Florida Free Association norms (Nelson et al., 2004), and Edinburgh Associative Thesaurus (Kiss et al., 1973). We achieve results of r=0.357 and p=0.379, r=0.344 and p=0.300, and r=0.292 and p=0.363, respectively. We find Word2Vec (Mikolov et al., 2013) and GloVe (Pennington et al., 2014) cosine similarities, as well as vector offsets, to be the highest performing features. Furthermore, we examine the usefulness of Gaussian embeddings (Vilnis and McCallum, 2014) for predicting word association strength, the first work to do so.

Session 4E: Poster Session. Discourse

10:30–12:10
Odense
Chair: *Sam Wiseman*

Learning Contextually Informed Representations for Linear-Time Discourse Parsing
Yang Liu and Mirella Lapata

Recent advances in RST discourse parsing have focused on two modeling paradigms: (a) high order parsers which jointly predict the tree structure of the discourse and the relations it encodes; or (b) linear-time parsers which are efficient but mostly based on local features. In this work, we propose a linear-time parser with a novel way of representing discourse constituents based on neural networks which takes into account global contextual information and is able to capture long-distance dependencies. Experimental results show that our parser obtains state-of-the art performance on benchmark datasets, while being efficient (with time complexity linear in the number of sentences in the document) and requiring minimal feature engineering.

Multi-task Attention-based Neural Networks for Implicit Discourse Relationship Representation and Identification
Man Lan, Jianxiang Wang, Yuanbin Wu, Zheng-Yu Niu, and Haifeng Wang

We present a novel multi-task attention based neural network model to address implicit discourse relationship representation and identification through two types of representation learning, an attention based neural network for learning discourse relationship representation with two arguments and a multi-task framework for learning knowledge from annotated and unannotated corpora. The extensive experiments have been performed on two benchmark corpora (i.e., PDTB and CoNLL-2016 datasets). Experimental results show that our proposed model outperforms the state-of-the-art systems on benchmark corpora.

Chinese Zero Pronoun Resolution with Deep Memory Network
Qingyu Yin, Yu Zhang, Weinan Zhang, and Ting Liu

Existing approaches for Chinese zero pronoun resolution typically utilize only syntactical and lexical features while ignoring semantic information. The fundamental reason is that zero pronouns have no descriptive information, which brings difficulty in explicitly capturing their semantic similarities with antecedents. Meanwhile, representing zero pronouns is challenging since they are merely gaps that convey no actual content. In this paper, we address this issue by building a deep memory network that is capable of encoding zero pronouns into vector representations with information obtained from their contexts and potential antecedents. Consequently, our resolver takes advantage of semantic information by using these continuous distributed representations. Experiments on the OntoNotes 5.0 dataset show that the proposed memory network could substantially outperform the state-of-the-art systems in various experimental settings.

How much progress have we made on RST discourse parsing? A replication study of recent results on the RST-DT
Mathieu Morey, Philippe Muller, and Nicholas Asher

This article evaluates purported progress over the past years in RST discourse parsing. Several studies report a relative error reduction of 24 to 51% on all metrics that authors attribute to the introduction of distributed representations of discourse units. We replicate the standard evaluation of 9 parsers, 5 of which use distributed representations, from 8 studies published between 2013 and 2017, using their predictions on the test set of the RST-DT. Our main finding is that most recently reported increases in RST discourse parser performance are an artefact of differences in implementations of the evaluation procedure. We evaluate all these parsers with the standard Parseval procedure to provide a more accurate picture of the actual RST discourse parsers performance in standard evaluation settings. Under this more stringent procedure, the gains attributable to distributed representations represent at most a 16% relative error reduction on fully-labelled structures.

What is it? Disambiguating the different readings of the pronoun 'it'
Sharid Loáiciga, Liane Guillou, and Christian Hardmeier

In this paper, we address the problem of predicting one of three functions for the English pronoun 'it': anaphoric, event reference or pleonastic. This disambiguation is valuable in the context of machine translation and coreference resolution. We present experiments using a MAXENT classifier trained on gold-standard data and self-training experiments of an RNN trained on silver-standard data, annotated using the MAXENT classifier. Lastly, we report on an analysis of the strengths of these two models.

Revisiting Selectional Preferences for Coreference Resolution
Benjamin Heinzerling, Nafise Sadat Moosavi, and Michael Strube

Selectional preferences have long been claimed to be essential for coreference resolution. However, they are modeled only implicitly by current coreference resolvers. We propose a dependency-based embedding model of selectional preferences which allows fine-grained compatibility judgments with high coverage. Incorporating our model improves performance, matching state-of-the-art results of a more complex system. However, it comes with a cost that makes it debatable how worthwhile are such improvements.

Learning to Rank Semantic Coherence for Topic Segmentation
Liang Wang, Sujian Li, Yajuan Lv, and Houfeng Wang

Topic segmentation plays an important role for discourse parsing and information retrieval. Due to the absence of training data, previous work mainly adopts unsupervised methods to rank semantic coherence between paragraphs for topic segmentation. In this paper, we present an intuitive and simple idea to automatically create a "quasi" training dataset, which includes a large amount of text pairs from the same or different documents with different semantic coherence. With the training corpus, we design a symmetric CNN neural network to model text pairs and rank the semantic coherence within the learning to rank framework. Experiments show that our algorithm is able to achieve competitive performance over strong baselines on several real-world datasets.

GRASP: Rich Patterns for Argumentation Mining
Eyal Shnarch, Ran Levy, Vikas Raykar, and Noam Slonim

GRASP (GReedy Augmented Sequential Patterns) is an algorithm for automatically extracting patterns that characterize subtle linguistic phenomena. To that end, GRASP augments each term of input text with multiple layers of linguistic information. These different facets of the text terms are systematically combined to reveal rich patterns. We report highly promising experimental results in several challenging text analysis tasks within the field of Argumentation Mining. We believe that GRASP is general enough to be useful for other domains too. For example, each of the following sentences includes a claim for a [topic]: 1. Opponents often argue that the open primary is unconstitutional. [Open Primaries] 2. Prof. Smith suggested that affirmative action devalues the accomplishments of the chosen. [Affirmative Action] 3. The majority stated that the First Amendment does not guarantee the right to offend others. [Freedom of Speech] These sentences share almost no words in common, however, they are similar at a more abstract level. A human observer may notice the following underlying common structure, or pattern: [someone][argue/suggest/state][that][topic term][sentiment term]. GRASP aims to automatically capture such underlying structures of the given data. For the above examples it finds the pattern [noun][express][that][noun,topic][sentiment], where [express] stands for all its (in)direct hyponyms, and [noun,topic] means a noun which is also related to the topic.

Patterns of Argumentation Strategies across Topics
Khalid Al Khatib, Henning Wachsmuth, Matthias Hagen, and Benno Stein

This paper presents an analysis of argumentation strategies in news editorials within and across topics. Given nearly 29,000 argumentative editorials from the New York Times, we develop two machine learning models, one for determining an editorial's topic, and one for identifying evidence types in the editorial. Based on the distribution and structure of the identified types, we analyze the usage patterns of argumentation strategies among 12 different topics. We detect several common patterns that provide insights into the manifestation of argumentation strategies. Also, our experiments reveal clear correlations between the topics and the detected patterns.

Using Argument-based Features to Predict and Analyse Review Helpfulness
Haijing Liu, Yang Gao, Pin Lv, Mengxue Li, Shiqiang Geng, Minglan Li, and Hao Wang

We study the helpful product reviews identification problem in this paper. We observe that the evidence-conclusion discourse relations, also known as arguments, often appear in product reviews, and we hypothesise that some argument-based features, e.g. the percentage of argumentative sentences, the evidences-conclusions ratios, are good indicators of helpful reviews. To validate this hypothesis, we manually annotate arguments in 110 hotel reviews, and investigate the effectiveness of several combinations of argument-based features. Experiments suggest that, when being used together with the argument-based features, the state-of-the-art baseline features can enjoy a performance boost (in terms of F1) of 11.01% in average.

Here's My Point: Joint Pointer Architecture for Argument Mining
Peter Potash, Alexey Romanov, and Anna Rumshisky

In order to determine argument structure in text, one must understand how individual components of the overall argument are linked. This work presents the first neural network-based approach to link extraction in argument mining. Specifically, we propose a novel architecture that applies Pointer Network sequence-to-sequence attention modeling to structural prediction in discourse parsing tasks. We then develop a joint model that extends this architecture to simultaneously address the link extraction task and the classification of argument components. The proposed joint model achieves state-of-the-art results on two separate evaluation corpora, showing far superior performance than the previously proposed corpus-specific and heavily feature-engineered models. Furthermore, our results demonstrate that jointly optimizing for both tasks is crucial for high performance.

Identifying attack and support argumentative relations using deep learning
Oana Cocarascu and Francesca Toni
We propose a deep learning architecture to capture argumentative relations of attack and support from one piece of text to another, of the kind that naturally occur in a debate. The architecture uses two (unidirectional or bidirectional) Long Short-Term Memory networks and (trained or non-trained) word embeddings, and allows to considerably improve upon existing techniques that use syntactic features and supervised classifiers for the same form of (relation-based) argument mining.

Session 4F: Poster Session. Machine Translation and Multilingual NLP 1

10:30–12:10

Copenhagen

Chair: *Anahita Mansouri Bigvand*

Neural Lattice-to-Sequence Models for Uncertain Inputs
Matthias Sperber, Graham Neubig, Jan Niehues, and Alex Waibel

The input to a neural sequence-to-sequence model is often determined by an up-stream system, e.g. a word segmenter, part of speech tagger, or speech recognizer. These up-stream models are potentially error-prone. Representing inputs through word lattices allows making this uncertainty explicit by capturing alternative sequences and their posterior probabilities in a compact form. In this work, we extend the TreeLSTM (Tai et al., 2015) into a LatticeLSTM that is able to consume word lattices, and can be used as encoder in an attentional encoder-decoder model. We integrate lattice posterior scores into this architecture by extending the TreeLSTM's child-sum and forget gates and introducing a bias term into the attention mechanism. We experiment with speech translation lattices and report consistent improvements over baselines that translate either the 1-best hypothesis or the lattice without posterior scores.

Memory-augmented Neural Machine Translation
Yang Feng, Shiyue Zhang, Andi Zhang, Dong Wang, and Andrew Abel

Neural machine translation (NMT) has achieved notable success in recent times, however it is also widely recognized that this approach has limitations with handling infrequent words and word pairs. This paper presents a novel memory-augmented NMT (M-NMT) architecture, which stores knowledge about how words (usually infrequently encountered ones) should be translated in a memory and then utilizes them to assist the neural model. We use this memory mechanism to combine the knowledge learned from a conventional statistical machine translation system and the rules learned by an NMT system, and also propose a solution for out-of-vocabulary (OOV) words based on this framework. Our experiments on two Chinese-English translation tasks demonstrated that the M-NMT architecture outperformed the NMT baseline by 9.0 and 2.7 BLEU points on the two tasks, respectively. Additionally, we found this architecture resulted in a much more effective OOV treatment compared to competitive methods.

Dynamic Data Selection for Neural Machine Translation
Marlies van der Wees, Arianna Bisazza, and Christof Monz

Intelligent selection of training data has proven a successful technique to simultaneously increase training efficiency and translation performance for phrase-based machine translation (PBMT). With the recent increase in popularity of neural machine translation (NMT), we explore in this paper to what extent and how NMT can also benefit from data selection. While state-of-the-art data selection (Axelrod et al., 2011) consistently performs well for PBMT, we show that gains are substantially lower for NMT. Next, we introduce 'dynamic data selection' for NMT, a method in which we vary the selected subset of training data between different training epochs. Our experiments show that the best results are achieved when applying a technique we call 'gradual fine-tuning', with improvements up to +2.6 BLEU over the original data selection approach and up to +3.1 BLEU over a general baseline.

Neural Machine Translation Leveraging Phrase-based Models in a Hybrid Search
Leonard Dahlmann, Evgeny Matusov, Pavel Petrushkov, and Shahram Khadivi

In this paper, we introduce a hybrid search for attention-based neural machine translation (NMT). A target phrase learned with statistical MT models extends a hypothesis in the NMT beam search when the attention of the NMT model focuses on the source words translated by this phrase. Phrases added in this way are scored with the NMT model, but also with SMT features including phrase-level translation probabilities and a target language model. Experimental results on German-to-English news domain and English-to-Russian e-commerce domain translation tasks show that using phrase-based models in NMT search improves MT quality by up to 2.3% BLEU absolute as compared to a strong NMT baseline.

Translating Phrases in Neural Machine Translation
Xing Wang, Zhaopeng Tu, Deyi Xiong, and Min Zhang

Phrases play an important role in natural language understanding and machine translation (Sag et al., 2002; Villavicencio et al., 2005). However, it is difficult to integrate them into current neural machine translation (NMT) which reads and generates sentences word by word. In this work, we propose a method to translate phrases in NMT by integrating a phrase memory storing target phrases from a phrase-based statistical machine translation (SMT) system into the encoder-decoder architecture of NMT. At each decoding

step, the phrase memory is first re-written by the SMT model, which dynamically generates relevant target phrases with contextual information provided by the NMT model. Then the proposed model reads the phrase memory to make probability estimations for all phrases in the phrase memory. If phrase generation is carried on, the NMT decoder selects an appropriate phrase from the memory to perform phrase translation and updates its decoding state by consuming the words in the selected phrase. Otherwise, the NMT decoder generates a word from the vocabulary as the general NMT decoder does. Experiment results on the Chinese to English translation show that the proposed model achieves significant improvements over the baseline on various test sets.

Towards Bidirectional Hierarchical Representations for Attention-based Neural Machine Translation
Baosong Yang, Derek F. Wong, Tong Xiao, Lidia S. Chao, and Jingbo Zhu

This paper proposes a hierarchical attentional neural translation model which focuses on enhancing source-side hierarchical representations by covering both local and global semantic information using a bidirectional tree-based encoder. To maximize the predictive likelihood of target words, a weighted variant of an attention mechanism is used to balance the attentive information between lexical and phrase vectors. Using a tree-based rare word encoding, the proposed model is extended to sub-word level to alleviate the out-of-vocabulary (OOV) problem. Empirical results reveal that the proposed model significantly outperforms sequence-to-sequence attention-based and tree-based neural translation models in English-Chinese translation tasks.

Exploring Hyperparameter Sensitivity in Neural Machine Translation Architectures
Denny Britz, Anna Goldie, Minh-Thang Luong, and Quoc Le

Neural Machine Translation (NMT) has shown remarkable progress over the past few years, with production systems now being deployed to end-users. As the field is moving rapidly, it has become unclear which elements of NMT architectures have a significant impact on translation quality. In this work, we present a large-scale analysis of the sensitivity of NMT architectures to common hyperparameters. We report empirical results and variance numbers for several hundred experimental runs, corresponding to over 250,000 GPU hours on a WMT English to German translation task. Our experiments provide practical insights into the relative importance of factors such as embedding size, network depth, RNN cell type, residual connections, attention mechanism, and decoding heuristics. As part of this contribution, we also release an open-source NMT framework in TensorFlow to make it easy for others to reproduce our results and perform their own experiments.

Learning Translations via Matrix Completion
Derry Tanti Wijaya, Brendan Callahan, John Hewitt, Jie Gao, Xiao Ling, Marianna Apidianaki, and Chris Callison-Burch

Bilingual Lexicon Induction is the task of learning word translations without bilingual parallel corpora. We model this task as a matrix completion problem, and present an effective and extendable framework for completing the matrix. This method harnesses diverse bilingual and monolingual signals, each of which may be incomplete or noisy. Our model achieves state-of-the-art performance for both high and low resource languages.

Reinforcement Learning for Bandit Neural Machine Translation with Simulated Human Feedback
Khanh Nguyen, Hal Daumé III, and Jordan Boyd-Graber

Machine translation is a natural candidate problem for reinforcement learning from human feedback: users provide quick, dirty ratings on candidate translations to guide a system to improve. Yet, current neural machine translation training focuses on expensive human-generated reference translations. We describe a reinforcement learning algorithm that improves neural machine translation systems from simulated human feedback. Our algorithm combines the advantage actor-critic algorithm (Mnih et al., 2016) with the attention-based neural encoder-decoder architecture (Luong et al., 2015). This algorithm (a) is well-designed for problems with a large action space and delayed rewards, (b) effectively optimizes traditional corpus-level machine translation metrics, and (c) is robust to skewed, high-variance, granular feedback modeled after actual human behaviors.

Towards Compact and Fast Neural Machine Translation Using a Combined Method
Xiaowei Zhang, Wei Chen, Feng Wang, Shuang Xu, and Bo Xu

Neural Machine Translation (NMT) lays intensive burden on computation and memory cost. It is a chal-

lenge to deploy NMT models on the devices with limited computation and memory budgets. This paper presents a four stage pipeline to compress model and speed up the decoding for NMT. Our method first introduces a compact architecture based on convolutional encoder and weight shared embeddings. Then weight pruning is applied to obtain a sparse model. Next, we propose a fast sequence interpolation approach which enables the greedy decoding to achieve performance on par with the beam search. Hence, the time-consuming beam search can be replaced by simple greedy decoding. Finally, vocabulary selection is used to reduce the computation of softmax layer. Our final model achieves 10 times speedup, 17 times parameters reduction, less than 35MB storage size and comparable performance compared to the baseline model.

Instance Weighting for Neural Machine Translation Domain Adaptation
Rui Wang, Masao Utiyama, Lemao Liu, Kehai Chen, and Eiichiro Sumita

Instance weighting has been widely applied to phrase-based machine translation domain adaptation. However, it is challenging to be applied to Neural Machine Translation (NMT) directly, because NMT is not a linear model. In this paper, two instance weighting technologies, i.e., sentence weighting and domain weighting with a dynamic weight learning strategy, are proposed for NMT domain adaptation. Empirical results on the IWSLT English-German/French tasks show that the proposed methods can substantially improve NMT performance by up to 2.7-6.7 BLEU points, outperforming the existing baselines by up to 1.6-3.6 BLEU points.

Regularization techniques for fine-tuning in neural machine translation
Antonio Valerio Miceli Barone, Barry Haddow, Ulrich Germann, and Rico Sennrich

We investigate techniques for supervised domain adaptation for neural machine translation where an existing model trained on a large out-of-domain dataset is adapted to a small in-domain dataset. In this scenario, overfitting is a major challenge. We investigate a number of techniques to reduce overfitting and improve transfer learning, including regularization techniques such as dropout and L2-regularization towards an out-of-domain prior. In addition, we introduce tuneout, a novel regularization technique inspired by dropout. We apply these techniques, alone and in combination, to neural machine translation, obtaining improvements on IWSLT datasets for English->German and English$->Russian. We also investigate the amounts of in-domain training data needed for domain adaptation in NMT, and find a logarithmic relationship between the amount of training data and gain in BLEU score.

Source-Side Left-to-Right or Target-Side Left-to-Right? An Empirical Comparison of Two Phrase-Based Decoding Algorithms
Yin-Wen Chang and Michael Collins

This paper describes an empirical study of the phrase-based decoding algorithm proposed by Chang and Collins (2017). The algorithm produces a translation by processing the source-language sentence in strictly left-to-right order, differing from commonly used approaches that build the target-language sentence in left-to-right order. Our results show that the new algorithm is competitive with Moses (Koehn et al., 2007) in terms of both speed and BLEU scores.

Using Target-side Monolingual Data for Neural Machine Translation through Multi-task Learning
Tobias Domhan and Felix Hieber

The performance of Neural Machine Translation (NMT) models relies heavily on the availability of sufficient amounts of parallel data, and an efficient and effective way of leveraging the vastly available amounts of monolingual data has yet to be found. We propose to modify the decoder in a neural sequence-to-sequence model to enable multi-task learning for two strongly related tasks: target-side language modeling and translation. The decoder predicts the next target word through two channels, a target-side language model on the lowest layer, and an attentional recurrent model which is conditioned on the source representation. This architecture allows joint training on both large amounts of monolingual and moderate amounts of bilingual data to improve NMT performance. Initial results in the news domain for three language pairs show moderate but consistent improvements over a baseline trained on bilingual data only.

Session 5 Overview – Sunday, September 10, 2017

Oral tracks

Track A *Semantics 3* Jutland	**Track B** *Computational Social Science 1* Funen	**Track C** *Sentiment Analysis 2* Zealand	
Encoding Sentences with Graph Convolutional Networks for Semantic Role Labeling *Marcheggiani and Titov*	Identifying civilians killed by police with distantly supervised entity-event extraction *Keith, Handler, Pinkham, Magliozzi, McDuffie, and O'Connor*	A Question Answering Approach for Emotion Cause Extraction *Gui, Hu, He, Xu, Qin, and Du*	13:40
Neural Semantic Parsing with Type Constraints for Semi-Structured Tables *Krishnamurthy, Dasigi, and Gardner*	Asking too much? The rhetorical role of questions in political discourse *Zhang, Spirling, and Danescu-Niculescu-Mizil*	Story Comprehension for Predicting What Happens Next *Chaturvedi, Peng, and Roth*	14:05
Joint Concept Learning and Semantic Parsing from Natural Language Explanations *Srivastava, Labutov, and Mitchell*	Detecting Perspectives in Political Debates *Vilares and He*	Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm *Felbo, Mislove, Søgaard, Rahwan, and Lehmann*	14:30
Grasping the Finer Point: A Supervised Similarity Network for Metaphor Detection *Rei, Bulat, Kiela, and Shutova*	"i have a feeling trump will win.....": Forecasting Winners and Losers from User Predictions on Twitter *Swamy, Ritter, and Marneffe*	Opinion Recommendation Using A Neural Model *Wang and Zhang*	14:55

Poster tracks

13:40–15:20

Track D: *Poster Session. Syntax 3* — Aarhus

Track E: *Poster Session. Relations* — Odense

Track F: *Poster Session. Language Models, Text Mining, and Crowd Sourcing* — Copenhagen

Parallel Session 5

Session 5A: Semantics 3

Jutland Chair: *Roberto Navigli*

Encoding Sentences with Graph Convolutional Networks for Semantic Role Labeling
Diego Marcheggiani and Ivan Titov 13:40–14:05

Semantic role labeling (SRL) is the task of identifying the predicate-argument structure of a sentence. It is typically regarded as an important step in the standard NLP pipeline. As the semantic representations are closely related to syntactic ones, we exploit syntactic information in our model. We propose a version of graph convolutional networks (GCNs), a recent class of neural networks operating on graphs, suited to model syntactic dependency graphs. GCNs over syntactic dependency trees are used as sentence encoders, producing latent feature representations of words in a sentence. We observe that GCN layers are complementary to LSTM ones: when we stack both GCN and LSTM layers, we obtain a substantial improvement over an already state-of-the-art LSTM SRL model, resulting in the best reported scores on the standard benchmark (CoNLL-2009) both for Chinese and English.

Neural Semantic Parsing with Type Constraints for Semi-Structured Tables
Jayant Krishnamurthy, Pradeep Dasigi, and Matt Gardner 14:05–14:30

We present a new semantic parsing model for answering compositional questions on semi-structured Wikipedia tables. Our parser is an encoder-decoder neural network with two key technical innovations: (1) a grammar for the decoder that only generates well-typed logical forms; and (2) an entity embedding and linking module that identifies entity mentions while generalizing across tables. We also introduce a novel method for training our neural model with question-answer supervision. On the WikiTableQuestions data set, our parser achieves a state-of-the-art accuracy of 43.3% for a single model and 45.9% for a 5-model ensemble, improving on the best prior score of 38.7% set by a 15-model ensemble. These results suggest that type constraints and entity linking are valuable components to incorporate in neural semantic parsers.

Joint Concept Learning and Semantic Parsing from Natural Language Explanations
Shashank Srivastava, Igor Labutov, and Tom Mitchell 14:30–14:55

Natural language constitutes a predominant medium for much of human learning and pedagogy. We consider the problem of concept learning from natural language explanations, and a small number of labeled examples of the concept. For example, in learning the concept of a phishing email, one might say 'this is a phishing email because it asks for your bank account number'. Solving this problem involves both learning to interpret open ended natural language statements, and learning the concept itself. We present a joint model for (1) language interpretation (semantic parsing) and (2) concept learning (classification) that does not require labeling statements with logical forms. Instead, the model prefers discriminative interpretations of statements in context of observable features of the data as a weak signal for parsing. On a dataset of email-related concepts, our approach yields across-the-board improvements in classification performance, with a 30% relative improvement in F1 score over competitive methods in the low data regime.

Grasping the Finer Point: A Supervised Similarity Network for Metaphor Detection
Marek Rei, Luana Bulat, Douwe Kiela, and Ekaterina Shutova 14:55–15:20

The ubiquity of metaphor in our everyday communication makes it an important problem for natural language understanding. Yet, the majority of metaphor processing systems to date rely on hand-engineered features and there is still no consensus in the field as to which features are optimal for this task. In this paper, we present the first deep learning architecture designed to capture metaphorical composition. Our results demonstrate that it outperforms the existing approaches in the metaphor identification task.

Session 5B: Computational Social Science 1

Funen Chair: *Noah A. Smith*

Identifying civilians killed by police with distantly supervised entity-event extraction
Katherine Keith, Abram Handler, Michael Pinkham, Cara Magliozzi, Joshua McDuffie, and Brendan O'Connor 13:40–14:05

We propose a new, socially-impactful task for natural language processing: from a news corpus, extract names of persons who have been killed by police. We present a newly collected police fatality corpus, which we release publicly, and present a model to solve this problem that uses EM-based distant supervision with logistic regression and convolutional neural network classifiers. Our model outperforms two off-the-shelf event extractor systems, and it can suggest candidate victim names in some cases faster than one of the major manually-collected police fatality databases.

Asking too much? The rhetorical role of questions in political discourse
Justine Zhang, Arthur Spirling, and Cristian Danescu-Niculescu-Mizil 14:05–14:30

Questions play a prominent role in social interactions, performing rhetorical functions that go beyond that of simple informational exchange. The surface form of a question can signal the intention and background of the person asking it, as well as the nature of their relation with the interlocutor. While the informational nature of questions has been extensively examined in the context of question-answering applications, their rhetorical aspects have been largely understudied. In this work we introduce an unsupervised methodology for extracting surface motifs that recur in questions, and for grouping them according to their latent rhetorical role. By applying this framework to the setting of question sessions in the UK parliament, we show that the resulting typology encodes key aspects of the political discourse—such as the bifurcation in questioning behavior between government and opposition parties—and reveals new insights into the effects of a legislator's tenure and political career ambitions.

Detecting Perspectives in Political Debates
David Vilares and Yulan He 14:30–14:55

We explore how to detect people's perspectives that occupy a certain proposition. We propose a Bayesian modelling approach where topics (or propositions) and their associated perspectives (or viewpoints) are modeled as latent variables. Words associated with topics or perspectives follow different generative routes. Based on the extracted perspectives, we can extract the top associated sentences from text to generate a succinct summary which allows a quick glimpse of the main viewpoints in a document. The model is evaluated on debates from the House of Commons of the UK Parliament, revealing perspectives from the debates without the use of labelled data and obtaining better results than previous related solutions under a variety of evaluations.

"i have a feeling trump will win.....": Forecasting Winners and Losers from User Predictions on Twitter
Sandesh Swamy, Alan Ritter, and Marie-Catherine de Marneffe 14:55–15:20

Social media users often make explicit predictions about upcoming events. Such statements vary in the degree of certainty the author expresses toward the outcome: "Leonardo DiCaprio will win Best Actor" vs. "Leonardo DiCaprio may win" or "No way Leonardo wins!". Can popular beliefs on social media predict who will win? To answer this question, we build a corpus of tweets annotated for veridicality on which we train a log-linear classifier that detects positive veridicality with high precision. We then forecast uncertain outcomes using the wisdom of crowds, by aggregating users' explicit predictions. Our method for forecasting winners is fully automated, relying only on a set of contenders as input. It requires no training data of past outcomes and outperforms sentiment and tweet volume baselines on a broad range of contest prediction tasks. We further demonstrate how our approach can be used to measure the reliability of individual accounts' predictions and retrospectively identify surprise outcomes.

Session 5C: Sentiment Analysis 2

Zealand — Chair: *Pascale Fung*

A Question Answering Approach for Emotion Cause Extraction
Lin Gui, Jiannan Hu, Yulan He, Ruifeng Xu, Lu Qin, and Jiachen Du 13:40–14:05

Emotion cause extraction aims to identify the reasons behind a certain emotion expressed in text. It is a much more difficult task compared to emotion classification. Inspired by recent advances in using deep memory networks for question answering (QA), we propose a new approach which considers emotion cause identification as a reading comprehension task in QA. Inspired by convolutional neural networks, we propose a new mechanism to store relevant context in different memory slots to model context information. Our proposed approach can extract both word level sequence features and lexical features. Performance evaluation shows that our method achieves the state-of-the-art performance on a recently released emotion cause dataset, outperforming a number of competitive baselines by at least 3.01% in F-measure.

Story Comprehension for Predicting What Happens Next
Snigdha Chaturvedi, Haoruo Peng, and Dan Roth 14:05–14:30

Automatic story comprehension is a fundamental challenge in Natural Language Understanding, and can enable computers to learn about social norms, human behavior and commonsense. In this paper, we present a story comprehension model that explores three distinct semantic aspects: (i) the sequence of events described in the story, (ii) its emotional trajectory, and (iii) its plot consistency. We judge the model's understanding of real-world stories by inquiring if, like humans, it can develop an expectation of what will happen next in a given story. Specifically, we use it to predict the correct ending of a given short story from possible alternatives. The model uses a hidden variable to weigh the semantic aspects in the context of the story. Our experiments demonstrate the potential of our approach to characterize these semantic aspects, and the strength of the hidden variable based approach. The model outperforms the state-of-the-art approaches and achieves best results on a publicly available dataset.

Using millions of emoji occurrences to learn any-domain representations for detecting sentiment, emotion and sarcasm
Bjarke Felbo, Alan Mislove, Anders Søgaard, Iyad Rahwan, and Sune Lehmann 14:30–14:55

NLP tasks are often limited by scarcity of manually annotated data. In social media sentiment analysis and related tasks, researchers have therefore used binarized emoticons and specific hashtags as forms of distant supervision. Our paper shows that by extending the distant supervision to a more diverse set of noisy labels, the models can learn richer representations. Through emoji prediction on a dataset of 1246 million tweets containing one of 64 common emojis we obtain state-of-the-art performance on 8 benchmark datasets within emotion, sentiment and sarcasm detection using a single pretrained model. Our analyses confirm that the diversity of our emotional labels yield a performance improvement over previous distant supervision approaches.

Opinion Recommendation Using A Neural Model
Zhongqing Wang and Yue Zhang 14:55–15:20

We present opinion recommendation, a novel task of jointly generating a review with a rating score that a certain user would give to a certain product which is unreviewed by the user, given existing reviews to the product by other users, and the reviews that the user has given to other products. A characteristic of opinion recommendation is the reliance of multiple data sources for multi-task joint learning. We use a single neural network to model users and products, generating customised product representations using a deep memory network, from which customised ratings and reviews are constructed jointly. Results show that our opinion recommendation system gives ratings that are closer to real user ratings on Yelp.com data compared with Yelp's own ratings. our methods give better results compared to several pipelines baselines.

Session 5D: Poster Session. Syntax 3

13:40–15:20
Aarhus
Chair: *Ryan Cotterell*

CRF Autoencoder for Unsupervised Dependency Parsing
Jiong Cai, Yong Jiang, and Kewei Tu

Unsupervised dependency parsing, which tries to discover linguistic dependency structures from unannotated data, is a very challenging task. Almost all previous work on this task focuses on learning generative models. In this paper, we develop an unsupervised dependency parsing model based on the CRF autoencoder. The encoder part of our model is discriminative and globally normalized which allows us to use rich features as well as universal linguistic priors. We propose an exact algorithm for parsing as well as a tractable learning algorithm. We evaluated the performance of our model on eight multilingual treebanks and found that our model achieved comparable performance with state-of-the-art approaches.

Efficient Discontinuous Phrase-Structure Parsing via the Generalized Maximum Spanning Arborescence
Caio Corro, Joseph Le Roux, and Mathieu Lacroix

We present a new method for the joint task of tagging and non-projective dependency parsing. We demonstrate its usefulness with an application to discontinuous phrase-structure parsing where decoding lexicalized spines and syntactic derivations is performed jointly. The main contributions of this paper are (1) a reduction from joint tagging and non-projective dependency parsing to the Generalized Maximum Spanning Arborescence problem, and (2) a novel decoding algorithm for this problem through Lagrangian relaxation. We evaluate this model and obtain state-of-the-art results despite strong independence assumptions.

Incremental Graph-based Neural Dependency Parsing
Xiaoqing Zheng

Very recently, some studies on neural dependency parsers have shown advantage over the traditional ones on a wide variety of languages. However, for graph-based neural dependency parsing systems, they either count on the long-term memory and attention mechanism to implicitly capture the high-order features or give up the global exhaustive inference algorithms in order to harness the features over a rich history of parsing decisions. The former might miss out the important features for specific headword predictions without the help of the explicit structural information, and the latter may suffer from the error propagation as false early structural constraints are used to create features when making future predictions. We explore the feasibility of explicitly taking high-order features into account while remaining the main advantage of global inference and learning for graph-based parsing. The proposed parser first forms an initial parse tree by head-modifier predictions based on the first-order factorization. High-order features (such as grandparent, sibling, and uncle) then can be defined over the initial tree, and used to refine the parse tree in an iterative fashion. Experimental results showed that our model (called INDP) archived competitive performance to existing benchmark parsers on both English and Chinese datasets.

Neural Discontinuous Constituency Parsing
Miloš Stanojević and Raquel Garrido Alhama

One of the most pressing issues in discontinuous constituency transition-based parsing is that the relevant information for parsing decisions could be located in any part of the stack or the buffer. In this paper, we propose a solution to this problem by replacing the structured perceptron model with a recursive neural model that computes a global representation of the configuration, therefore allowing even the most remote parts of the configuration to influence the parsing decisions. We also provide a detailed analysis of how this representation should be built out of sub-representations of its core elements (words, trees and stack). Additionally, we investigate how different types of swap oracles influence the results. Our model is the first neural discontinuous constituency parser, and it outperforms all the previously published models on three out of four datasets while on the fourth it obtains second place by a tiny difference.

Stack-based Multi-layer Attention for Transition-based Dependency Parsing
Zhirui Zhang, Shujie Liu, Mu Li, Ming Zhou, and Enhong Chen

Although sequence-to-sequence (seq2seq) network has achieved significant success in many NLP tasks such as machine translation and text summarization, simply applying this approach to transition-based dependency parsing cannot yield a comparable performance gain as in other state-of-the-art methods, such as stack-LSTM and head selection. In this paper, we propose a stack-based multi-layer attention

model for seq2seq learning to better leverage structural linguistics information. In our method, two binary vectors are used to track the decoding stack in transition-based parsing, and multi-layer attention is introduced to capture multiple word dependencies in partial trees. We conduct experiments on PTB and CTB datasets, and the results show that our proposed model achieves state-of-the-art accuracy and significant improvement in labeled precision with respect to the baseline seq2seq model.

Dependency Grammar Induction with Neural Lexicalization and Big Training Data
Wenjuan Han, Yong Jiang, and Kewei Tu

We study the impact of big models (in terms of the degree of lexicalization) and big data (in terms of the training corpus size) on dependency grammar induction. We experimented with L-DMV, a lexicalized version of Dependency Model with Valence and L-NDMV, our lexicalized extension of the Neural Dependency Model with Valence. We find that L-DMV only benefits from very small degrees of lexicalization and moderate sizes of training corpora. L-NDMV can benefit from big training data and lexicalization of greater degrees, especially when enhanced with good model initialization, and it achieves a result that is competitive with the current state-of-the-art.

Combining Generative and Discriminative Approaches to Unsupervised Dependency Parsing via Dual Decomposition
Yong Jiang, Wenjuan Han, and Kewei Tu

Unsupervised dependency parsing aims to learn a dependency parser from unannotated sentences. Existing work focuses on either learning generative models using the expectation-maximization algorithm and its variants, or learning discriminative models using the discriminative clustering algorithm. In this paper, we propose a new learning strategy that learns a generative model and a discriminative model jointly based on the dual decomposition method. Our method is simple and general, yet effective to capture the advantages of both models and improve their learning results. We tested our method on the UD treebank and achieved a state-of-the-art performance on thirty languages.

Effective Inference for Generative Neural Parsing
Mitchell Stern, Daniel Fried, and Dan Klein

Generative neural models have recently achieved state-of-the-art results for constituency parsing. However, without a feasible search procedure, their use has so far been limited to reranking the output of external parsers in which decoding is more tractable. We describe an alternative to the conventional action-level beam search used for discriminative neural models that enables us to decode directly in these generative models. We then show that by improving our basic candidate selection strategy and using a coarse pruning function, we can improve accuracy while exploring significantly less of the search space. Applied to the model of Choe and Charniak (2016), our inference procedure obtains 92.56 F1 on section 23 of the Penn Treebank, surpassing prior state-of-the-art results for single-model systems.

Semi-supervised Structured Prediction with Neural CRF Autoencoder
Xiao Zhang, Yong Jiang, Hao Peng, Kewei Tu, and Dan Goldwasser

In this paper we propose an end-to-end neural CRF autoencoder (NCRF-AE) model for semi-supervised learning of sequential structured prediction problems. Our NCRF-AE consists of two parts: an encoder which is a CRF model enhanced by deep neural networks, and a decoder which is a generative model trying to reconstruct the input. Our model has a unified structure with different loss functions for labeled and unlabeled data with shared parameters. We developed a variation of the EM algorithm for optimizing both the encoder and the decoder simultaneously by decoupling their parameters. Our Experimental results over the Part-of-Speech (POS) tagging task on eight different languages, show that our model can outperform competitive systems in both supervised and semi-supervised scenarios.

TAG Parsing with Neural Networks and Vector Representations of Supertags
Jungo Kasai, Bob Frank, Tom McCoy, Owen Rambow, and Alexis Nasr

We present supertagging-based models for Tree Adjoining Grammar parsing that use neural network architectures and dense vector representation of supertags (elementary trees) to achieve state-of-the-art performance in unlabeled and labeled attachment scores. The shift-reduce parsing model eschews lexical information entirely, and uses only the 1-best supertags to parse a sentence, providing further support for the claim that supertagging is "almost parsing." We demonstrate that the embedding vector representations the parser induces for supertags possess linguistically interpretable structure, supporting analogies between grammatical structures like those familiar from recent work in distributional semantics. This

dense representation of supertags overcomes the drawbacks for statistical models of TAG as compared to CCG parsing, raising the possibility that TAG is a viable alternative for NLP tasks that require the assignment of richer structural descriptions to sentences.

Session 5E: Poster Session. Relations

13:40–15:20
Odense
Chair: *Bishan Yang*

Global Normalization of Convolutional Neural Networks for Joint Entity and Relation Classification
Heike Adel and Hinrich Schütze
We introduce globally normalized convolutional neural networks for joint entity classification and relation extraction. In particular, we propose a way to utilize a linear-chain conditional random field output layer for predicting entity types and relations between entities at the same time. Our experiments show that global normalization outperforms a locally normalized softmax layer on a benchmark dataset.

End-to-End Neural Relation Extraction with Global Optimization
Meishan Zhang, Yue Zhang, and Guohong Fu
Neural networks have shown promising results for relation extraction. State-of-the-art models cast the task as an end-to-end problem, solved incrementally using a local classifier. Yet previous work using statistical models have demonstrated that global optimization can achieve better performances compared to local classification. We build a globally optimized neural model for end-to-end relation extraction, proposing novel LSTM features in order to better learn context representations. In addition, we present a novel method to integrate syntactic information to facilitate global learning, yet requiring little background on syntactic grammars thus being easy to extend. Experimental results show that our proposed model is highly effective, achieving the best performances on two standard benchmarks.

KGEval: Accuracy Estimation of Automatically Constructed Knowledge Graphs
Prakhar Ojha and Partha Talukdar
Automatic construction of large knowledge graphs (KG) by mining web-scale text datasets has received considerable attention recently. Estimating accuracy of such automatically constructed KGs is a challenging problem due to their size and diversity. This important problem has largely been ignored in prior research — we fill this gap and propose KGEval. KGEval uses coupling constraints to bind facts and crowdsources those few that can infer large parts of the graph. We demonstrate that the objective optimized by KGEval is submodular and NP-hard, allowing guarantees for our approximation algorithm. Through experiments on real-world datasets, we demonstrate that KGEval best estimates KG accuracy compared to other baselines, while requiring significantly lesser number of human evaluations.

Sparsity and Noise: Where Knowledge Graph Embeddings Fall Short
Jay Pujara, Eriq Augustine, and Lise Getoor
Knowledge graph (KG) embedding techniques use structured relationships between entities to learn low-dimensional representations of entities and relations. One prominent goal of these approaches is to improve the quality of knowledge graphs by removing errors and adding missing facts. Surprisingly, most embedding techniques have been evaluated on benchmark datasets consisting of dense and reliable subsets of human-curated KGs, which tend to be fairly complete and have few errors. In this paper, we consider the problem of applying embedding techniques to KGs extracted from text, which are often incomplete and contain errors. We compare the sparsity and unreliability of different KGs and perform empirical experiments demonstrating how embedding approaches degrade as sparsity and unreliability increase.

Dual Tensor Model for Detecting Asymmetric Lexico-Semantic Relations
Goran Glavaš and Simone Paolo Ponzetto
Detection of lexico-semantic relations is one of the central tasks of computational semantics. Although some fundamental relations (e.g., hypernymy) are asymmetric, most existing models account for asymmetry only implicitly and use the same concept representations to support detection of symmetric and asymmetric relations alike. In this work, we propose the Dual Tensor model, a neural architecture with which we explicitly model the asymmetry and capture the translation between unspecialized and specialized word embeddings via a pair of tensors. Although our Dual Tensor model needs only unspecialized embeddings as input, our experiments on hypernymy and meronymy detection suggest that it can outperform more complex and resource-intensive models. We further demonstrate that the model can account for polysemy and that it exhibits stable performance across languages.

Incorporating Relation Paths in Neural Relation Extraction
Wenyuan Zeng, Yankai Lin, Zhiyuan Liu, and Maosong Sun

Distantly supervised relation extraction has been widely used to find novel relational facts from plain text. To predict the relation between a pair of two target entities, existing methods solely rely on those direct sentences containing both entities. In fact, there are also many sentences containing only one of the target entities, which also provide rich useful information but not yet employed by relation extraction. To address this issue, we build inference chains between two target entities via intermediate entities, and propose a path-based neural relation extraction model to encode the relational semantics from both direct sentences and inference chains. Experimental results on real-world datasets show that, our model can make full use of those sentences containing only one target entity, and achieves significant and consistent improvements on relation extraction as compared with strong baselines. The source code of this paper can be obtained from https:// github.com/thunlp/PathNRE.

Adversarial Training for Relation Extraction
Yi Wu, David Bamman, and Stuart Russell

Adversarial training is a mean of regularizing classification algorithms by generating adversarial noise to the training data. We apply adversarial training in relation extraction within the multi-instance multi-label learning framework. We evaluate various neural network architectures on two different datasets. Experimental results demonstrate that adversarial training is generally effective for both CNN and RNN models and significantly improves the precision of predicted relations.

Context-Aware Representations for Knowledge Base Relation Extraction
Daniil Sorokin and Iryna Gurevych

We demonstrate that for sentence-level relation extraction it is beneficial to consider other relations in the sentential context while predicting the target relation. Our architecture uses an LSTM-based encoder to jointly learn representations for all relations in a single sentence. We combine the context representations with an attention mechanism to make the final prediction. We use the Wikidata knowledge base to construct a dataset of multiple relations per sentence and to evaluate our approach. Compared to a baseline system, our method results in an average error reduction of 24 on a held-out set of relations. The code and the dataset to replicate the experiments are made available at https://github.com/ukplab/.

A Soft-label Method for Noise-tolerant Distantly Supervised Relation Extraction
Tianyu Liu, Kexiang Wang, Baobao Chang, and Zhifang Sui

Distant-supervised relation extraction inevitably suffers from wrong labeling problems because it heuristically labels relational facts with knowledge bases. Previous sentence level denoise models don't achieve satisfying performances because they use hard labels which are determined by distant supervision and immutable during training. To this end, we introduce an entity-pair level denoise method which exploits semantic information from correctly labeled entity pairs to correct wrong labels dynamically during training. We propose a joint score function which combines the relational scores based on the entity-pair representation and the confidence of the hard label to obtain a new label, namely a soft label, for certain entity pair. During training, soft labels instead of hard labels serve as gold labels. Experiments on the benchmark dataset show that our method dramatically reduces noisy instances and outperforms other state-of-the-art systems.

A Sequential Model for Classifying Temporal Relations between Intra-Sentence Events
Prafulla Kumar Choubey and Ruihong Huang

We present a sequential model for temporal relation classification between intra-sentence events. The key observation is that the overall syntactic structure and compositional meanings of the multi-word context between events are important for distinguishing among fine-grained temporal relations. Specifically, our approach first extracts a sequence of context words that indicates the temporal relation between two events, which well align with the dependency path between two event mentions. The context word sequence, together with a parts-of-speech tag sequence and a dependency relation sequence that are generated corresponding to the word sequence, are then provided as input to bidirectional recurrent neural network (LSTM) models. The neural nets learn compositional syntactic and semantic representations of contexts surrounding the two events and predict the temporal relation between them. Evaluation of the proposed approach on TimeBank corpus shows that sequential modeling is capable of accurately recognizing temporal relations between events, which outperforms a neural net model using various discrete features as input that imitates previous feature based models.

Deep Residual Learning for Weakly-Supervised Relation Extraction
YiYao Huang and William Yang Wang

Deep residual learning (ResNet) is a new method for training very deep neural networks using identity mapping for shortcut connections. ResNet has won the ImageNet ILSVRC 2015 classification task, and achieved state-of-the-art performances in many computer vision tasks. However, the effect of residual learning on noisy natural language processing tasks is still not well understood. In this paper, we design a novel convolutional neural network (CNN) with residual learning, and investigate its impacts on the task of distantly supervised noisy relation extraction. In contradictory to popular beliefs that ResNet only works well for very deep networks, we found that even with 9 layers of CNNs, using identity mapping could significantly improve the performance for distantly-supervised relation extraction.

Noise-Clustered Distant Supervision for Relation Extraction: A Nonparametric Bayesian Perspective
Qing Zhang and Houfeng Wang

For the task of relation extraction, distant supervision is an efficient approach to generate labeled data by aligning knowledge base with free texts. The essence of it is a challenging incomplete multi-label classification problem with sparse and noisy features. To address the challenge, this work presents a novel nonparametric Bayesian formulation for the task. Experiment results show substantially higher top precision improvements over the traditional state-of-the-art approaches.

Exploring Vector Spaces for Semantic Relations
Kata Gábor, Haifa Zargayouna, Isabelle Tellier, Davide Buscaldi, and Thierry Charnois

Word embeddings are used with success for a variety of tasks involving lexical semantic similarities between individual words. Using unsupervised methods and just cosine similarity, encouraging results were obtained for analogical similarities. In this paper, we explore the potential of pre-trained word embeddings to identify generic types of semantic relations in an unsupervised experiment. We propose a new relational similarity measure based on the combination of word2vec's CBOW input and output vectors which outperforms concurrent vector representations, when used for unsupervised clustering on SemEval 2010 Relation Classification data.

Temporal dynamics of semantic relations in word embeddings: an application to predicting armed conflict participants
Andrey Kutuzov, Erik Velldal, and Lilja Øvrelid

This paper deals with using word embedding models to trace the temporal dynamics of semantic relations between pairs of words. The set-up is similar to the well-known analogies task, but expanded with a time dimension. To this end, we apply incremental updating of the models with new training texts, including incremental vocabulary expansion, coupled with learned transformation matrices that let us map between members of the relation. The proposed approach is evaluated on the task of predicting insurgent armed groups based on geographical locations. The gold standard data for the time span 1994–2010 is extracted from the UCDP Armed Conflicts dataset. The results show that the method is feasible and outperforms the baselines, but also that important work still remains to be done.

Session 5F: Poster Session. Language Models, Text Mining, and Crowd Sourcing

13:40–15:20
Copenhagen
Chair: *Allen Schmaltz*

Dynamic Entity Representations in Neural Language Models
Yangfeng Ji, Chenhao Tan, Sebastian Martschat, Yejin Choi, and Noah A. Smith

Understanding a long document requires tracking how entities are introduced and evolve over time. We present a new type of language model, EntityNLM, that can explicitly model entities, dynamically update their representations, and contextually generate their mentions. Our model is generative and flexible; it can model an arbitrary number of entities in context while generating each entity mention at an arbitrary length. In addition, it can be used for several different tasks such as language modeling, coreference resolution, and entity prediction. Experimental results with all these tasks demonstrate that our model consistently outperforms strong baselines and prior work.

Towards Quantum Language Models
Ivano Basile and Fabio Tamburini

This paper presents a new approach for building Language Models using the Quantum Probability Theory, a Quantum Language Model (QLM). It mainly shows that relying on this probability calculus it is possible to build stochastic models able to benefit from quantum correlations due to interference and entanglement. We extensively tested our approach showing its superior performances, both in terms of model perplexity and inserting it into an automatic speech recognition evaluation setting, when compared with state-of-the-art language modelling techniques.

Reference-Aware Language Models
Zichao Yang, Phil Blunsom, Chris Dyer, and Wang Ling

We propose a general class of language models that treat reference as discrete stochastic latent variables. This decision allows for the creation of entity mentions by accessing external databases of referents (required by, e.g., dialogue generation) or past internal state (required to explicitly model coreferentiality). Beyond simple copying, our coreference model can additionally refer to a referent using varied mention forms (e.g., a reference to "Jane" can be realized as "she"), a characteristic feature of reference in natural languages. Experiments on three representative applications show our model variants outperform models based on deterministic attention and standard language modeling baselines.

A Simple Language Model based on PMI Matrix Approximations
Oren Melamud, Ido Dagan, and Jacob Goldberger

In this study, we introduce a new approach for learning language models by training them to estimate word-context pointwise mutual information (PMI), and then deriving the desired conditional probabilities from PMI at test time. Specifically, we show that with minor modifications to word2vec's algorithm, we get principled language models that are closely related to the well-established Noise Contrastive Estimation (NCE) based language models. A compelling aspect of our approach is that our models are trained with the same simple negative sampling objective function that is commonly used in word2vec to learn word embeddings.

Syllable-aware Neural Language Models: A Failure to Beat Character-aware Ones
Zhenisbek Assylbekov, Rustem Takhanov, Bagdat Myrzakhmetov, and Jonathan N. Washington

Syllabification does not seem to improve word-level RNN language modeling quality when compared to character-based segmentation. However, our best syllable-aware language model, achieving performance comparable to the competitive character-aware model, has 18%-33% fewer parameters and is trained 1.2-2.2 times faster.

Inducing Semantic Micro-Clusters from Deep Multi-View Representations of Novels
Lea Frermann and György Szarvas

Automatically understanding the plot of novels is important both for informing literary scholarship and applications such as summarization or recommendation. Various models have addressed this task, but their evaluation has remained largely intrinsic and qualitative. Here, we propose a principled and scalable framework leveraging expert-provided semantic tags (e.g., mystery, pirates) to evaluate plot representations in an extrinsic fashion, assessing their ability to produce locally coherent groupings of novels (micro-clusters) in model space. We present a deep recurrent autoencoder model that learns richly struc-

tured multi-view plot representations, and show that they i) yield better micro-clusters than less structured representations; and ii) are interpretable, and thus useful for further literary analysis or labeling of the emerging micro-clusters.

Initializing Convolutional Filters with Semantic Features for Text Classification
Shen Li, Zhe Zhao, Tao Liu, Renfen Hu, and Xiaoyong Du

Convolutional Neural Networks (CNNs) are widely used in NLP tasks. This paper presents a novel weight initialization method to improve the CNNs for text classification. Instead of randomly initializing the convolutional filters, we encode semantic features into them, which helps the model focus on learning useful features at the beginning of the training. Experiments demonstrate the effectiveness of the initialization technique on seven text classification tasks, including sentiment analysis and topic classification.

Shortest-Path Graph Kernels for Document Similarity
Giannis Nikolentzos, Polykarpos Meladianos, Francois Rousseau, Yannis Stavrakas, and Michalis Vazirgiannis

In this paper, we present a novel document similarity measure based on the definition of a graph kernel between pairs of documents. The proposed measure takes into account both the terms contained in the documents and the relationships between them. By representing each document as a graph-of-words, we are able to model these relationships and then determine how similar two documents are by using a modified shortest-path graph kernel. We evaluate our approach on two tasks and compare it against several baseline approaches using various performance metrics such as DET curves and macro-average F1-score. Experimental results on a range of datasets showed that our proposed approach outperforms traditional techniques and is capable of measuring more accurately the similarity between two documents.

A Joint Many-Task Model: Growing a Neural Network for Multiple NLP Tasks
Kazuma Hashimoto, Caiming Xiong, Yoshimasa Tsuruoka, and Richard Socher

Transfer and multi-task learning have traditionally focused on either a single source-target pair or very few, similar tasks. Ideally, the linguistic levels of morphology, syntax and semantics would benefit each other by being trained in a single model. We introduce a joint many-task model together with a strategy for successively growing its depth to solve increasingly complex tasks. Higher layers include shortcut connections to lower-level task predictions to reflect linguistic hierarchies. We use a simple regularization term to allow for optimizing all model weights to improve one task's loss without exhibiting catastrophic interference of the other tasks. Our single end-to-end model obtains state-of-the-art or competitive results on five different tasks from tagging, parsing, relatedness, and entailment tasks.

Adapting Topic Models using Lexical Associations with Tree Priors
Weiwei Yang, Jordan Boyd-Graber, and Philip Resnik

Models work best when they are optimized taking into account the evaluation criteria that people care about. For topic models, people often care about interpretability, which can be approximated using measures of lexical association. We integrate lexical association into topic optimization using tree priors, which provide a flexible framework that can take advantage of both first order word associations and the higher-order associations captured by word embeddings. Tree priors improve topic interpretability without hurting extrinsic performance.

Finding Patterns in Noisy Crowds: Regression-based Annotation Aggregation for Crowdsourced Data
Natalie Parde and Rodney Nielsen

Crowdsourcing offers a convenient means of obtaining labeled data quickly and inexpensively. However, crowdsourced labels are often noisier than expert-annotated data, making it difficult to aggregate them meaningfully. We present an aggregation approach that learns a regression model from crowdsourced annotations to predict aggregated labels for instances that have no expert adjudications. The predicted labels achieve a correlation of 0.594 with expert labels on our data, outperforming the best alternative aggregation method by 11.9%. Our approach also outperforms the alternatives on third-party datasets.

CROWD-IN-THE-LOOP: A Hybrid Approach for Annotating Semantic Roles
Chenguang Wang, Alan Akbik, Laura Chiticariu, Yunyao Li, Fei Xia, and Anbang Xu

Crowdsourcing has proven to be an effective method for generating labeled data for a range of NLP tasks. However, multiple recent attempts of using crowdsourcing to generate gold-labeled training data

for semantic role labeling (SRL) reported only modest results, indicating that SRL is perhaps too difficult a task to be effectively crowdsourced. In this paper, we postulate that while producing SRL annotation does require expert involvement in general, a large subset of SRL labeling tasks is in fact appropriate for the crowd. We present a novel workflow in which we employ a classifier to identify difficult annotation tasks and route each task either to experts or crowd workers according to their difficulties. Our experimental evaluation shows that the proposed approach reduces the workload for experts by over two-thirds, and thus significantly reduces the cost of producing SRL annotation at little loss in quality.

Session 6 Overview – Sunday, September 10, 2017

Oral tracks

	Track A	**Track B**	**Track C**
	Machine Translation 2	*Text Mining and NLP applications*	*Machine Comprehension*
	Jutland	Funen	Zealand
15:50	Earth Mover's Distance Minimization for Unsupervised Bilingual Lexicon Induction *Zhang, Liu, Luan, and Sun*	Satirical News Detection and Analysis using Attention Mechanism and Linguistic Features *Yang, Mukherjee, and Dragut*	Accurate Supervised and Semi-Supervised Machine Reading for Long Documents *Hewlett, Jones, Lacoste, and Gur*
16:15	Unfolding and Shrinking Neural Machine Translation Ensembles *Stahlberg and Byrne*	Fine Grained Citation Span for References in Wikipedia *Fetahu, Markert, and Anand*	Adversarial Examples for Evaluating Reading Comprehension Systems *Jia and Liang*
16:40	Graph Convolutional Encoders for Syntax-aware Neural Machine Translation *Bastings, Titov, Aziz, Marcheggiani, and Simaan*	[TACL] Joint Modeling of Topics, Citations, and Topical Authority in Academic Corpora *Kim, Kim, and Oh*	Reasoning with Heterogeneous Knowledge for Commonsense Machine Comprehension *Lin, Sun, and Han*
17:05	Trainable Greedy Decoding for Neural Machine Translation *Gu, Cho, and Li*	Identifying Semantic Edit Intentions from Revisions in Wikipedia *Yang, Halfaker, Kraut, and Hovy*	Document-Level Multi-Aspect Sentiment Classification as Machine Comprehension *Yin, Song, and Zhang*

Poster tracks

15:50–17:30

Track D: *Poster Session. Summarization, Generation, Dialog, and Discourse 1* — Aarhus

Track E: *Poster Session. Summarization, Generation, Dialog, and Discourse 2* — Odense

Track F: *Poster Session. Computational Social Science 2* — Copenhagen

Parallel Session 6

Session 6A: Machine Translation 2

Jutland — Chair: *Timothy Baldwin*

Earth Mover's Distance Minimization for Unsupervised Bilingual Lexicon Induction
Meng Zhang, Yang Liu, Huanbo Luan, and Maosong Sun 15:50–16:15

Cross-lingual natural language processing hinges on the premise that there exists invariance across languages. At the word level, researchers have identified such invariance in the word embedding semantic spaces of different languages. However, in order to connect the separate spaces, cross-lingual supervision encoded in parallel data is typically required. In this paper, we attempt to establish the cross-lingual connection without relying on any cross-lingual supervision. By viewing word embedding spaces as distributions, we propose to minimize their earth mover's distance, a measure of divergence between distributions. We demonstrate the success on the unsupervised bilingual lexicon induction task. In addition, we reveal an interesting finding that the earth mover's distance shows potential as a measure of language difference.

Unfolding and Shrinking Neural Machine Translation Ensembles
Felix Stahlberg and Bill Byrne 16:15–16:40

Ensembling is a well-known technique in neural machine translation (NMT) to improve system performance. Instead of a single neural net, multiple neural nets with the same topology are trained separately, and the decoder generates predictions by averaging over the individual models. Ensembling often improves the quality of the generated translations drastically. However, it is not suitable for production systems because it is cumbersome and slow. This work aims to reduce the runtime to be on par with a single system without compromising the translation quality. First, we show that the ensemble can be unfolded into a single large neural network which imitates the output of the ensemble system. We show that unfolding can already improve the runtime in practice since more work can be done on the GPU. We proceed by describing a set of techniques to shrink the unfolded network by reducing the dimensionality of layers. On Japanese-English we report that the resulting network has the size and decoding speed of a single NMT network but performs on the level of a 3-ensemble system.

Graph Convolutional Encoders for Syntax-aware Neural Machine Translation
Joost Bastings, Ivan Titov, Wilker Aziz, Diego Marcheggiani, and Khalil Simaan 16:40–17:05

We present a simple and effective approach to incorporating syntactic structure into neural attention-based encoder-decoder models for machine translation. We rely on graph-convolutional networks (GCNs), a recent class of neural networks developed for modeling graph-structured data. Our GCNs use predicted syntactic dependency trees of source sentences to produce representations of words (i.e. hidden states of the encoder) that are sensitive to their syntactic neighborhoods. GCNs take word representations as input and produce word representations as output, so they can easily be incorporated as layers into standard encoders (e.g., on top of bidirectional RNNs or convolutional neural networks). We evaluate their effectiveness with English-German and English-Czech translation experiments for different types of encoders and observe substantial improvements over their syntax-agnostic versions in all the considered setups.

Trainable Greedy Decoding for Neural Machine Translation
Jiatao Gu, Kyunghyun Cho, and Victor O.K. Li 17:05–17:30

Recent research in neural machine translation has largely focused on two aspects; neural network architectures and end-to-end learning algorithms. The problem of decoding, however, has received relatively little attention from the research community. In this paper, we solely focus on the problem of decoding given a trained neural machine translation model. Instead of trying to build a new decoding algorithm for any specific decoding objective, we propose the idea of trainable decoding algorithm in which we train a decoding algorithm to find a translation that maximizes an arbitrary decoding objective. More specifically, we design an actor that observes and manipulates the hidden state of the neural machine translation decoder and propose to train it using a variant of deterministic policy gradient. We extensively evaluate the proposed algorithm using four language pairs and two decoding objectives and show that we can indeed train a trainable greedy decoder that generates a better translation (in terms of a target decoding objective) with minimal computational overhead.

Session 6B: Text Mining and NLP applications

Funen Chair: *Jill Burstein*

Satirical News Detection and Analysis using Attention Mechanism and Linguistic Features
Fan Yang, Arjun Mukherjee, and Eduard Dragut 15:50–16:15

Satirical news is considered to be entertainment, but it is potentially deceptive and harmful. Despite the embedded genre in the article, not everyone can recognize the satirical cues and therefore believe the news as true news. We observe that satirical cues are often reflected in certain paragraphs rather than the whole document. Existing works only consider document-level features to detect the satire, which could be limited. We consider paragraph-level linguistic features to unveil the satire by incorporating neural network and attention mechanism. We investigate the difference between paragraph-level features and document-level features, and analyze them on a large satirical news dataset. The evaluation shows that the proposed model detects satirical news effectively and reveals what features are important at which level.

Fine Grained Citation Span for References in Wikipedia
Besnik Fetahu, Katja Markert, and Avishek Anand 16:15–16:40

Verifiability is one of the core editing principles in Wikipedia, where editors are encouraged to provide citations for the added content. For a Wikipedia article determining what content is covered by a citation or the citation span is not trivial, an important aspect for automated citation finding for uncovered content, or fact assessments. We address the problem of determining the citation span in Wikipedia articles. We approach this problem by classifying which textual fragments in an article are covered or hold true given a citation. We propose a sequence classification approach where for a paragraph and a citation, we determine the citation span at a fine-grained level. We provide a thorough experimental evaluation and compare our approach against baselines adopted from the scientific domain, where we show improvement for all evaluation metrics.

[TACL] Joint Modeling of Topics, Citations, and Topical Authority in Academic Corpora
Jooyeon Kim, Dongwoo Kim, and Alice Oh 16:40–17:05

Much of scientific progress stems from previously published findings, but searching through the vast sea of scientific publications is difficult. We often rely on metrics of scholarly authority to find the prominent authors but these authority indices do not differentiate authority based on research topics. We present Latent Topical-Authority Indexing (LTAI) for jointly modeling the topics, citations, and topical authority in a corpus of academic papers. Compared to previous models, LTAI differs in two main aspects. First, it explicitly models the generative process of the citations, rather than treating the citations as given. Second, it models each author's influence on citations of a paper based on the topics of the cited papers, as well as the citing papers. We fit LTAI to four academic corpora: CORA, Arxiv Physics, PNAS, and Citeseer. We compare the performance of LTAI against various baselines, starting with the latent Dirichlet allocation, to the more advanced models including author-link topic model and dynamic author citation topic model. The results show that LTAI achieves improved accuracy over other similar models when predicting words, citations and authors of publications.

Identifying Semantic Edit Intentions from Revisions in Wikipedia
Diyi Yang, Aaron Halfaker, Robert Kraut, and Eduard Hovy 17:05–17:30

Most studies on human editing focus merely on syntactic revision operations, failing to capture the intentions behind revision changes, which are essential for facilitating the single and collaborative writing process. In this work, we develop in collaboration with Wikipedia editors a 13-category taxonomy of the semantic intention behind edits in Wikipedia articles. Using labeled article edits, we build a computational classifier of intentions that achieved a micro-averaged F1 score of 0.621. We use this model to investigate edit intention effectiveness: how different types of edits predict the retention of newcomers and changes in the quality of articles, two key concerns for Wikipedia today. Our analysis shows that the types of edits that users make in their first session predict their subsequent survival as Wikipedia editors, and articles in different stages need different types of edits.

Session 6C: Machine Comprehension

Zealand Chair: *Ndapa Nakashole*

Accurate Supervised and Semi-Supervised Machine Reading for Long Documents
Daniel Hewlett, Llion Jones, Alexandre Lacoste, and Izzeddin Gur 15:50–16:15

We introduce a hierarchical architecture for machine reading capable of extracting precise information from long documents. The model divides the document into small, overlapping windows and encodes all windows in parallel with an RNN. It then attends over these window encodings, reducing them to a single encoding, which is decoded into an answer using a sequence decoder. This hierarchical approach allows the model to scale to longer documents without increasing the number of sequential steps. In a supervised setting, our model achieves state of the art accuracy of 76.8 on the WikiReading dataset. We also evaluate the model in a semi-supervised setting by downsampling the WikiReading training set to create increasingly smaller amounts of supervision, while leaving the full unlabeled document corpus to train a sequence autoencoder on document windows. We evaluate models that can reuse autoencoder states and outputs without fine-tuning their weights, allowing for more efficient training and inference.

Adversarial Examples for Evaluating Reading Comprehension Systems
Robin Jia and Percy Liang 16:15–16:40

Standard accuracy metrics indicate that reading comprehension systems are making rapid progress, but the extent to which these systems truly understand language remains unclear. To reward systems with real language understanding abilities, we propose an adversarial evaluation scheme for the Stanford Question Answering Dataset (SQuAD). Our method tests whether systems can answer questions about paragraphs that contain adversarially inserted sentences, which are automatically generated to distract computer systems without changing the correct answer or misleading humans. In this adversarial setting, the accuracy of sixteen published models drops from an average of 75% F1 score to 36%; when the adversary is allowed to add ungrammatical sequences of words, average accuracy on four models decreases further to 7%. We hope our insights will motivate the development of new models that understand language more precisely.

Reasoning with Heterogeneous Knowledge for Commonsense Machine Comprehension
Hongyu Lin, Le Sun, and Xianpei Han 16:40–17:05

Reasoning with commonsense knowledge is critical for natural language understanding. Traditional methods for commonsense machine comprehension mostly only focus on one specific kind of knowledge, neglecting the fact that commonsense reasoning requires simultaneously considering different kinds of commonsense knowledge. In this paper, we propose a multi-knowledge reasoning method, which can exploit heterogeneous knowledge for commonsense machine comprehension. Specifically, we first mine different kinds of knowledge (including event narrative knowledge, entity semantic knowledge and sentiment coherent knowledge) and encode them as inference rules with costs. Then we propose a multi-knowledge reasoning model, which selects inference rules for a specific reasoning context using attention mechanism, and reasons by summarizing all valid inference rules. Experiments on RocStories show that our method outperforms traditional models significantly.

Document-Level Multi-Aspect Sentiment Classification as Machine Comprehension
Yichun Yin, Yangqiu Song, and Ming Zhang 17:05–17:30

Document-level multi-aspect sentiment classification is an important task for customer relation management. In this paper, we model the task as a machine comprehension problem where pseudo question-answer pairs are constructed by a small number of aspect-related keywords and aspect ratings. A hierarchical iterative attention model is introduced to build aspectspecific representations by frequent and repeated interactions between documents and aspect questions. We adopt a hierarchical architecture to represent both word level and sentence level information, and use the attention operations for aspect questions and documents alternatively with the multiple hop mechanism. Experimental results on the TripAdvisor and BeerAdvocate datasets show that our model outperforms classical baselines. We will release our code and data for the method replicability.

Session 6D: Poster Session. Summarization, Generation, Dialog, and Discourse 1

15:50–17:30
Aarhus
Chair: *Yangfeng Ji*

What is the Essence of a Claim? Cross-Domain Claim Identification
Johannes Daxenberger, Steffen Eger, Ivan Habernal, Christian Stab, and Iryna Gurevych

Argument mining has become a popular research area in NLP. It typically includes the identification of argumentative components, e.g. claims, as the central component of an argument. We perform a qualitative analysis across six different datasets and show that these appear to conceptualize claims quite differently. To learn about the consequences of such different conceptualizations of claim for practical applications, we carried out extensive experiments using state-of-the-art feature-rich and deep learning systems, to identify claims in a cross-domain fashion. While the divergent conceptualization of claims in different datasets is indeed harmful to cross-domain classification, we show that there are shared properties on the lexical level as well as system configurations that can help to overcome these gaps.

Identifying Where to Focus in Reading Comprehension for Neural Question Generation
Xinya Du and Claire Cardie

A first step in the task of automatically generating questions for testing reading comprehension is to identify *question-worthy* sentences, i.e. sentences in a text passage that humans find it worthwhile to ask questions about. We propose a hierarchical neural sentence-level sequence tagging model for this task, which existing approaches to question generation have ignored. The approach is fully data-driven — with no sophisticated NLP pipelines or any hand-crafted rules/features — and compares favorably to a number of baselines when evaluated on the SQuAD data set. When incorporated into an existing neural question generation system, the resulting end-to-end system achieves state-of-the-art performance for paragraph-level question generation for reading comprehension.

Break it Down for Me: A Study in Automated Lyric Annotation
Lucas Sterckx, Jason Naradowsky, Bill Byrne, Thomas Demeester, and Chris Develder

Comprehending lyrics, as found in songs and poems, can pose a challenge to human and machine readers alike. This motivates the need for systems that can understand the ambiguity and jargon found in such creative texts, and provide commentary to aid readers in reaching the correct interpretation. We introduce the task of automated lyric annotation (ALA). Like text simplification, a goal of ALA is to rephrase the original text in a more easily understandable manner. However, in ALA the system must often include additional information to clarify niche terminology and abstract concepts. To stimulate research on this task, we release a large collection of crowdsourced annotations for song lyrics. We analyze the performance of translation and retrieval models on this task, measuring performance with both automated and human evaluation. We find that each model captures a unique type of information important to the task.

Cascaded Attention based Unsupervised Information Distillation for Compressive Summarization
Piji Li, Wai Lam, Lidong Bing, Weiwei Guo, and Hang Li

When people recall and digest what they have read for writing summaries, the important content is more likely to attract their attention. Inspired by this observation, we propose a cascaded attention based unsupervised model to estimate the salience information from the text for compressive multi-document summarization. The attention weights are learned automatically by an unsupervised data reconstruction framework which can capture the sentence salience. By adding sparsity constraints on the number of output vectors, we can generate condensed information which can be treated as word salience. Fine-grained and coarse-grained sentence compression strategies are incorporated to produce compressive summaries. Experiments on some benchmark data sets show that our framework achieves better results than the state-of-the-art methods.

Deep Recurrent Generative Decoder for Abstractive Text Summarization
Piji Li, Wai Lam, Lidong Bing, and Zihao Wang

We propose a new framework for abstractive text summarization based on a sequence-to-sequence oriented encoder-decoder model equipped with a deep recurrent generative decoder (DRGN). Latent structure information implied in the target summaries is learned based on a recurrent latent random model for improving the summarization quality. Neural variational inference is employed to address the intractable posterior inference for the recurrent latent variables. Abstractive summaries are generated based on both

the generative latent variables and the discriminative deterministic states. Extensive experiments on some benchmark datasets in different languages show that DRGN achieves improvements over the state-of-the-art methods.

Extractive Summarization Using Multi-Task Learning with Document Classification
Masaru Isonuma, Toru Fujino, Junichiro Mori, Yutaka Matsuo, and Ichiro Sakata

The need for automatic document summarization that can be used for practical applications is increasing rapidly. In this paper, we propose a general framework for summarization that extracts sentences from a document using externally related information. Our work is aimed at single document summarization using small amounts of reference summaries. In particular, we address document summarization in the framework of multi-task learning using curriculum learning for sentence extraction and document classification. The proposed framework enables us to obtain better feature representations to extract sentences from documents. We evaluate our proposed summarization method on two datasets: financial report and news corpus. Experimental results demonstrate that our summarizers achieve performance that is comparable to state-of-the-art systems.

Towards Automatic Construction of News Overview Articles by News Synthesis
Jianmin Zhang and Xiaojun Wan

In this paper we investigate a new task of automatically constructing an overview article from a given set of news articles about a news event. We propose a news synthesis approach to address this task based on passage segmentation, ranking, selection and merging. Our proposed approach is compared with several typical multi-document summarization methods on the Wikinews dataset, and achieves the best performance on both automatic evaluation and manual evaluation.

Joint Syntacto-Discourse Parsing and the Syntacto-Discourse Treebank
Kai Zhao and Liang Huang

Discourse parsing has long been treated as a stand-alone problem independent from constituency or dependency parsing. Most attempts at this problem rely on annotated text segmentations (Elementary Discourse Units, EDUs) and sophisticated sparse or continuous features to extract syntactic information. In this paper we propose the first end-to-end discourse parser that jointly parses in both syntax and discourse levels, as well as the first syntactic-discourse treebank by integrating the Penn Treebank and the RST Treebank. Built upon our recent span-based constituency parser, this joint syntactic-discourse parser requires no preprocessing efforts such as segmentation or feature extraction, making discourse parsing more convenient. Empirically, our parser achieves the state-of-the-art end-to-end discourse parsing accuracy.

Event Coreference Resolution by Iteratively Unfolding Inter-dependencies among Events
Prafulla Kumar Choubey and Ruihong Huang

We introduce a novel iterative approach for event coreference resolution that gradually builds event clusters by exploiting inter-dependencies among event mentions within the same chain as well as across event chains. Among event mentions in the same chain, we distinguish within- and cross-document event coreference links by using two distinct pairwise classifiers, trained separately to capture differences in feature distributions of within- and cross-document event clusters. Our event coreference approach alternates between WD and CD clustering and combines arguments from both event clusters after every merge, continuing till no more merge can be made. And then it performs further merging between event chains that are both closely related to a set of other chains of events. Experiments on the ECB+ corpus show that our model outperforms state-of-the-art methods in joint task of WD and CD event coreference resolution.

When to Finish? Optimal Beam Search for Neural Text Generation (modulo beam size)
Liang Huang, Kai Zhao, and Mingbo Ma

In neural text generation such as neural machine translation, summarization, and image captioning, beam search is widely used to improve the output text quality. However, in the neural generation set- ting, hypotheses can finish in different steps, which makes it difficult to decide when to end beam search to ensure op- timality. We propose a provably optimal beam search algorithm that will always re- turn the optimal-score complete hypothe- sis (modulo beam size), and finish as soon as the optimality is established. To counter neural generation's tendency for shorter hypotheses, we also introduce a bounded length reward mechanism which allows a modified version of our beam search al- gorithm to remain optimal. Experiments on neural machine translation demonstrate that our principled beam search algorithm leads to improvement in BLEU score over previously proposed alternatives.

Steering Output Style and Topic in Neural Response Generation
Di Wang, Nebojsa Jojic, Chris Brockett, and Eric Nyberg

We propose simple and flexible training and decoding methods for influencing output style and topic in neural encoder-decoder based language generation. This capability is desirable in a variety of applications, including conversational systems, where successful agents need to produce language in a specific style and generate responses steered by a human puppeteer or external knowledge. We decompose the neural generation process into empirically easier sub-problems: a faithfulness model and a decoding method based on selective-sampling. We also describe training and sampling algorithms that bias the generation process with a specific language style restriction, or a topic restriction. Human evaluation results show that our proposed methods are able to to restrict style and topic without degrading output quality in conversational tasks.

Session 6E: Poster Session. Summarization, Generation, Dialog, and Discourse 2

15:50–17:30
Odense
Chairs: *Elkin Dario Gutierrez , Natalie Schluter*

Preserving Distributional Information in Dialogue Act Classification
Quan Hung Tran, Ingrid Zukerman, and Gholamreza Haffari

This paper introduces a novel training/decoding strategy for sequence labeling. Instead of greedily choosing a label at each time step, and using it for the next prediction, we retain the probability distribution over the current label, and pass this distribution to the next prediction. This approach allows us to avoid the effect of label bias and error propagation in sequence learning/decoding. Our experiments on dialogue act classification demonstrate the effectiveness of this approach. Even though our underlying neural network model is relatively simple, it outperforms more complex neural models, achieving state-of-the-art results on the MapTask and Switchboard corpora.

Adversarial Learning for Neural Dialogue Generation
Jiwei Li, Will Monroe, Tianlin Shi, Sébastien Jean, Alan Ritter, and Dan Jurafsky

We apply adversarial training to open-domain dialogue generation, training a system to produce sequences that are indistinguishable from human-generated dialogue utterances. We cast the task as a reinforcement learning problem where we jointly train two systems: a generative model to produce response sequences, and a discriminator—analagous to the human evaluator in the Turing test— to distinguish between the human-generated dialogues and the machine-generated ones. In this generative adversarial network approach, the outputs from the discriminator are used to encourage the system towards more human-like dialogue. Further, we investigate models for adversarial evaluation that uses success in fooling an adversary as a dialogue evaluation metric, while avoiding a number of potential pitfalls. Experimental results on several metrics, including adversarial evaluation, demonstrate that the adversarially-trained system generates higher-quality responses than previous baselines

Using Context Information for Dialog Act Classification in DNN Framework
Yang Liu, Kun Han, Zhao Tan, and Yun Lei

Previous work on dialog act (DA) classification has investigated different methods, such as hidden Markov models, maximum entropy, conditional random fields, graphical models, and support vector machines. A few recent studies explored using deep learning neural networks for DA classification, however, it is not clear yet what is the best method for using dialog context or DA sequential information, and how much gain it brings. This paper proposes several ways of using context information for DA classification, all in the deep learning framework. The baseline system classifies each utterance using the convolutional neural networks (CNN). Our proposed methods include using hierarchical models (recurrent neural networks (RNN) or CNN) for DA sequence tagging where the bottom layer takes the sentence CNN representation as input, concatenating predictions from the previous utterances with the CNN vector for classification, and performing sequence decoding based on the predictions from the sentence CNN model. We conduct thorough experiments and comparisons on the Switchboard corpus, demonstrate that incorporating context information significantly improves DA classification, and show that we achieve new state-of-the-art performance for this task.

Modeling Dialogue Acts with Content Word Filtering and Speaker Preferences
Yohan Jo, Michael Yoder, Hyeju Jang, and Carolyn Rose

We present an unsupervised model of dialogue act sequences in conversation. By modeling topical themes as transitioning more slowly than dialogue acts in conversation, our model de-emphasizes content-related words in order to focus on conversational function words that signal dialogue acts. We also incorporate speaker tendencies to use some acts more than others as an additional predictor of dialogue act prevalence beyond temporal dependencies. According to the evaluation presented on two dissimilar corpora, the CNET forum and NPS Chat corpus, the effectiveness of each modeling assumption is found to vary depending on characteristics of the data. De-emphasizing content-related words yields improvement on the CNET corpus, while utilizing speaker tendencies is advantageous on the NPS corpus. The components of our model complement one another to achieve robust performance on both corpora and outperform state-of-the-art baseline models.

Towards Implicit Content-Introducing for Generative Short-Text Conversation Systems
Lili Yao, Yaoyuan Zhang, Yansong Feng, Dongyan Zhao, and Rui Yan

The study on human-computer conversation systems is a hot research topic nowadays. One of the prevailing methods to build the system is using the generative Sequence-to-Sequence (Seq2Seq) model through neural networks. However, the standard Seq2Seq model is prone to generate trivial responses. In this paper, we aim to generate a more meaningful and informative reply when answering a given question. We propose an implicit content-introducing method which incorporates additional information into the Seq2Seq model in a flexible way. Specifically, we fuse the general decoding and the auxiliary cue word information through our proposed hierarchical gated fusion unit. Experiments on real-life data demonstrate that our model consistently outperforms a set of competitive baselines in terms of BLEU scores and human evaluation.

Affordable On-line Dialogue Policy Learning
Cheng Chang, Runzhe Yang, Lu Chen, Xiang Zhou, and Kai Yu

The key to building an evolvable dialogue system in real-world scenarios is to ensure an affordable on-line dialogue policy learning, which requires the on-line learning process to be safe, efficient and economical. But in reality, due to the scarcity of real interaction data, the dialogue system usually grows slowly. Besides, the poor initial dialogue policy easily leads to bad user experience and incurs a failure of attracting users to contribute training data, so that the learning process is unsustainable. To accurately depict this, two quantitative metrics are proposed to assess safety and efficiency issues. For solving the unsustainable learning problem, we proposed a complete companion teaching framework incorporating the guidance from the human teacher. Since the human teaching is expensive, we compared various teaching schemes answering the question how and when to teach, to economically utilize teaching budget, so that make the online learning process affordable.

Generating High-Quality and Informative Conversation Responses with Sequence-to-Sequence Models
Yuanlong Shao, Stephan Gouws, Denny Britz, Anna Goldie, Brian Strope, and Ray Kurzweil

Sequence-to-sequence models have been applied to the conversation response generation problem where the source sequence is the conversation history and the target sequence is the response. Unlike translation, conversation responding is inherently creative. The generation of long, informative, coherent, and diverse responses remains a hard task. In this work, we focus on the single turn setting. We add self-attention to the decoder to maintain coherence in longer responses, and we propose a practical approach, called the glimpse-model, for scaling to large datasets. We introduce a stochastic beam-search algorithm with segment-by-segment reranking which lets us inject diversity earlier in the generation process. We trained on a combined data set of over 2.3B conversation messages mined from the web. In human evaluation studies, our method produces longer responses overall, with a higher proportion rated as acceptable and excellent as length increases, compared to baseline sequence-to-sequence models with explicit length-promotion. A back-off strategy produces better responses overall, in the full spectrum of lengths.

Bootstrapping incremental dialogue systems from minimal data: the generalisation power of dialogue grammars
Arash Eshghi, Igor Shalyminov, and Oliver Lemon

We investigate an end-to-end method for automatically inducing task-based dialogue systems from small amounts of unannotated dialogue data. It combines an incremental semantic grammar - Dynamic Syntax and Type Theory with Records (DS-TTR) - with Reinforcement Learning (RL), where language generation and dialogue management are a joint decision problem. The systems thus produced are incremental: dialogues are processed word-by-word, shown previously to be essential in supporting natural, spontaneous dialogue. We hypothesised that the rich linguistic knowledge within the grammar should enable a combinatorially large number of dialogue variations to be processed, even when trained on very few dialogues. Our experiments show that our model can process 74% of the Facebook AI bAbI dataset even when trained on only 0.13% of the data (5 dialogues). It can in addition process 65% of bAbI+, a corpus we created by systematically adding incremental dialogue phenomena such as restarts and self-corrections to bAbI. We compare our model with a state-of-the-art retrieval model, MEMN2N. We find that, in terms of semantic accuracy, the MEMN2N model shows very poor robustness to the bAbI+ transformations even when trained on the full bAbI dataset.

Composite Task-Completion Dialogue Policy Learning via Hierarchical Deep Reinforcement Learning
Baolin Peng, Xiujun Li, Lihong Li, Jianfeng Gao, Asli Celikyilmaz, Sungjin Lee, and Kam-Fai Wong

Building a dialogue agent to fulfill complex tasks, such as travel planning, is challenging because the agent has to learn to collectively complete multiple subtasks. For example, the agent needs to reserve a hotel and book a flight so that there leaves enough time for commute between arrival and hotel check-in. This paper addresses this challenge by formulating the task in the mathematical framework of options over Markov Decision Processes (MDPs), and proposing a hierarchical deep reinforcement learning approach to learning a dialogue manager that operates at different temporal scales. The dialogue manager consists of: (1) a top-level dialogue policy that selects among subtasks or options, (2) a low-level dialogue policy that selects primitive actions to complete the subtask given by the top-level policy, and (3) a global state tracker that helps ensure all cross-subtask constraints be satisfied. Experiments on a travel planning task with simulated and real users show that our approach leads to significant improvements over three baselines, two based on handcrafted rules and the other based on flat deep reinforcement learning.

Why We Need New Evaluation Metrics for NLG
Jekaterina Novikova, Ondřej Dušek, Amanda Cercas Curry, and Verena Rieser

The majority of NLG evaluation relies on automatic metrics, such as BLEU . In this paper, we motivate the need for novel, system- and data-independent automatic evaluation methods: We investigate a wide range of metrics, including state-of-the-art word-based and novel grammar-based ones, and demonstrate that they only weakly reflect human judgements of system outputs as generated by data-driven, end-to-end NLG. We also show that metric performance is data- and system-specific. Nevertheless, our results also suggest that automatic metrics perform reliably at system-level and can support system development by finding cases where a system performs poorly.

Challenges in Data-to-Document Generation
Sam Wiseman, Stuart Shieber, and Alexander Rush

Recent neural models have shown significant progress on the problem of generating short descriptive texts conditioned on a small number of database records. In this work, we suggest a slightly more difficult data-to-text generation task, and investigate how effective current approaches are on this task. In particular, we introduce a new, large-scale corpus of data records paired with descriptive documents, propose a series of extractive evaluation methods for analyzing performance, and obtain baseline results using current neural generation methods. Experiments show that these models produce fluent text, but fail to convincingly approximate human-generated documents. Moreover, even templated baselines exceed the performance of these neural models on some metrics, though copy- and reconstruction-based extensions lead to noticeable improvements.

Session 6F: Poster Session. Computational Social Science 2

15:50–17:30
Copenhagen
Chair: *Afshin Rahimi*

All that is English may be Hindi: Enhancing language identification through automatic ranking of the likeliness of word borrowing in social media
Jasabanta Patro, Bidisha Samanta, Saurabh Singh, Abhipsa Basu, Prithwish Mukherjee, Monojit Choudhury, and Animesh Mukherjee

n this paper, we present a set of computational methods to identify the likeliness of a word being borrowed, based on the signals from social media. In terms of Spearman's correlation values, our methods perform more than two times better (0.62) in predicting the borrowing likeliness compared to the best performing baseline (0.26) reported in literature. Based on this likeliness estimate we asked annotators to re-annotate the language tags of foreign words in predominantly native contexts. In 88% of cases the annotators felt that the foreign language tag should be replaced by native language tag, thus indicating a huge scope for improvement of automatic language identification systems.

Multi-View Unsupervised User Feature Embedding for Social Media-based Substance Use Prediction
Tao Ding, Warren K. Bickel, and Shimei Pan

In this paper, we demonstrate how the state-of-the-art machine learning and text mining techniques can be used to build effective social media-based substance use detection systems. Since a substance use ground truth is difficult to obtain on a large scale, to maximize system performance, we explore different unsupervised feature learning methods to take advantage of a large amount of unsupervised social media data. We also demonstrate the benefit of using multi-view unsupervised feature learning to combine heterogeneous user information such as Facebook "likes" and "status updates" to enhance system performance. Based on our evaluation, our best models achieved 86% AUC for predicting tobacco use, 81% for alcohol use and 84% for illicit drug use, all of which significantly outperformed existing methods. Our investigation has also uncovered interesting relations between a user's social media behavior (e.g., word usage) and substance use.

Demographic-aware word associations
Aparna Garimella, Carmen Banea, and Rada Mihalcea

Variations of word associations across different groups of people can provide insights into people's psychologies and their world views. To capture these variations, we introduce the task of demographic-aware word associations. We build a new gold standard dataset consisting of word association responses for approximately 300 stimulus words, collected from more than 800 respondents of different gender (male/female) and from different locations (India/United States), and show that there are significant variations in the word associations made by these groups. We also introduce a new demographic-aware word association model based on a neural net skip-gram architecture, and show how computational methods for measuring word associations that specifically account for writer demographics can outperform generic methods that are agnostic to such information.

A Factored Neural Network Model for Characterizing Online Discussions in Vector Space
Hao Cheng, Hao Fang, and Mari Ostendorf

We develop a novel factored neural model that learns comment embeddings in an unsupervised way leveraging the structure of distributional context in online discussion forums. The model links different context with related language factors in the embedding space, providing a way to interpret the factored embeddings. Evaluated on a community endorsement prediction task using a large collection of topic-varying Reddit discussions, the factored embeddings consistently achieve improvement over other text representations. Qualitative analysis shows that the model captures community style and topic, as well as response trigger patterns.

Dimensions of Interpersonal Relationships: Corpus and Experiments
Farzana Rashid and Eduardo Blanco

This paper presents a corpus and experiments to determine dimensions of interpersonal relationships. We define a set of dimensions heavily inspired by work in social science. We create a corpus by retrieving pairs of people, and then annotating dimensions for their relationships. A corpus analysis shows that dimensions can be annotated reliably. Experimental results show that given a pair of people, values to dimensions can be assigned automatically.

Argument Mining on Twitter: Arguments, Facts and Sources
Mihai Dusmanu, Elena Cabrio, and Serena Villata

Social media collect and spread on the Web personal opinions, facts, fake news and all kind of information users may be interested in. Applying argument mining methods to such heterogeneous data sources is a challenging open research issue, in particular considering the peculiarities of the language used to write textual messages on social media. In addition, new issues emerge when dealing with arguments posted on such platforms, such as the need to make a distinction between personal opinions and actual facts, and to detect the source disseminating information about such facts to allow for provenance verification. In this paper, we apply supervised classification to identify arguments on Twitter, and we present two new tasks for argument mining, namely facts recognition and source identification. We study the feasibility of the approaches proposed to address these tasks on a set of tweets related to the Grexit and Brexit news topics.

Distinguishing Japanese Non-standard Usages from Standard Ones
Tatsuya Aoki, Ryohei Sasano, Hiroya Takamura, and Manabu Okumura

We focus on non-standard usages of common words on social media. In the context of social media, words sometimes have other usages that are totally different from their original. In this study, we attempt to distinguish non-standard usages on social media from standard ones in an unsupervised manner. Our basic idea is that non-standardness can be measured by the inconsistency between the expected meaning of the target word and the given context. For this purpose, we use context embeddings derived from word embeddings. Our experimental results show that the model leveraging the context embedding outperforms other methods and provide us with findings, for example, on how to construct context embeddings and which corpus to use.

Connotation Frames of Power and Agency in Modern Films
Maarten Sap, Marcella Cindy Prasettio, Ari Holtzman, Hannah Rashkin, and Yejin Choi

The framing of an action influences how we perceive its actor. We introduce connotation frames of power and agency, a pragmatic formalism organized using frame semantic representations, to model how different levels of power and agency are implicitly projected on actors through their actions. We use the new power and agency frames to measure the subtle, but prevalent, gender bias in the portrayal of modern film characters and provide insights that deviate from the well-known Bechdel test. Our contributions include an extended lexicon of connotation frames along with a web interface that provides a comprehensive analysis through the lens of connotation frames.

Controlling Human Perception of Basic User Traits
Daniel Preoţiuc-Pietro, Sharath Chandra Guntuku, and Lyle Ungar

Much of our online communication is text-mediated and, lately, more common with automated agents. Unlike interacting with humans, these agents currently do not tailor their language to the type of person they are communicating to. In this pilot study, we measure the extent to which human perception of basic user trait information – gender and age – is controllable through text. Using automatic models of gender and age prediction, we estimate which tweets posted by a user are more likely to mis-characterize his traits. We perform multiple controlled crowdsourcing experiments in which we show that we can reduce the human prediction accuracy of gender to almost random – an over 20% drop in accuracy. Our experiments show that it is practically feasible for multiple applications such as text generation, text summarization or machine translation to be tailored to specific traits and perceived as such.

Topic Signatures in Political Campaign Speeches
Clément Gautrais, Peggy Cellier, René Quiniou, and Alexandre Termier

Highlighting the recurrence of topics usage in candidates speeches is a key feature to identify the main ideas of each candidate during a political campaign. In this paper, we present a method combining standard topic modeling with signature mining for analyzing topic recurrence in speeches of Clinton and Trump during the 2016 American presidential campaign. The results show that the method extracts automatically the main ideas of each candidate and, in addition, provides information about the evolution of these topics during the campaign.

Assessing Objective Recommendation Quality through Political Forecasting
H. Andrew Schwartz, Masoud Rouhizadeh, Michael Bishop, Philip Tetlock, Barbara Mellers, and Lyle Ungar

Recommendations are often rated for their subjective quality, but few researchers have studied comment quality in terms of objective utility. We explore recommendation quality assessment with respect to both subjective (i.e. users' ratings) and objective (i.e., did it influence? did it improve decisions?) metrics in a massive online geopolitical forecasting system, ultimately comparing linguistic characteristics of each quality metric. Using a variety of features, we predict all types of quality with better accuracy than the simple yet strong baseline of comment length. Looking at the most predictive content illustrates rater biases; for example, forecasters are subjectively biased in favor of comments mentioning business transactions or dealings as well as material things, even though such comments do not indeed prove any more useful objectively. Additionally, more complex sentence constructions, as evidenced by subordinate conjunctions, are characteristic of comments leading to objective improvements in forecasting.

Never Abandon Minorities: Exhaustive Extraction of Bursty Phrases on Microblogs Using Set Cover Problem
Masumi Shirakawa, Takahiro Hara, and Takuya Maekawa

We propose a language-independent data-driven method to exhaustively extract bursty phrases of arbitrary forms (e.g., phrases other than simple noun phrases) from microblogs. The burst (i.e., the rapid increase of the occurrence) of a phrase causes the burst of overlapping N-grams including incomplete ones. In other words, bursty incomplete N-grams inevitably overlap bursty phrases. Thus, the proposed method performs the extraction of bursty phrases as the set cover problem in which all bursty N-grams are covered by a minimum set of bursty phrases. Experimental results using Japanese Twitter data showed that the proposed method outperformed word-based, noun phrase-based, and segmentation-based methods both in terms of accuracy and coverage.

SIGDAT Business Meeting

Date: Sunday, September 10, 2017
Time: 12:40–13:40
Venue: Jutland

Chair: *Pascale Fung*

All attendees are encouraged to participate in the business meeting.

Social Event

Sunday, September 10, 2017, 18:00–22:00

Øksnehallen Courtyard

The EMNLP 2017 social event will take place in the court yard just outside the venue, directly after the last session of Sunday. Some social events are designed to give attendees a chance to see a popular tourist attraction, but this year, the social event is pure fun and relaxation. All attendees are invited to grab free food at our food trucks, listen to great music, or just chill in the lounge areas. Of course the social event will include surprises, also, but - hey, they're surprises.

7

Main Conference: Monday, September 11

Overview

07:30 – 17:30 **Registration Day 3** *Foyer*

08:00 – 09:00 **Morning Coffee** *Foyer*

09:00 – 10:00 **Invited Talk.** *Jutland*

"Does This Vehicle Belong to You"? Processing the Language of Policing for Improving Police-Community Relations (Dan Jurafsky)

10:00 – 10:30 **Coffee Break** *Foyer*

Session 7

10:30 – 12:10	Machine Learning 3 *Jutland*	Syntax 4 *Funen*	Dialogue *Zealand*
	Poster Session. Machine Translation and Multilingual NLP 2 *Aarhus*	Poster Session. Information Extraction 2 *Odense*	Poster Session. NLP Applications *Copenhagen*

12:10 – 13:40 **Lunch**

Session 8

13:40 – 15:25	Machine Translation and Multilingual/Multimodal NLP (Short) *Jutland*	Machine Learning (Short) *Funen*	NLP Applications (Short) *Zealand*

15:25 – 15:50 **Coffee Break** *Foyer*

15:50 – 17:25 **Plenary Session. Best Paper Awards** *Jutland*

17:25 – 17:45 **Plenary Session. Closing Remarks** *Jutland*

17:25 – 17:45 **Closing Remarks (General Chair)** *Jutland*

Invited Talk: Dan Jurafsky

"Does This Vehicle Belong to You"? Processing the Language of Policing for Improving Police-Community Relations

Monday, September 11, 2017, 9:00–10:00

Jutland

Abstract: Police body-cameras have the potential to play an important role in understanding and improving police-community relations. In this talk I describe a series of studies conducted by our large interdisciplinary team at Stanford that use speech and natural language processing on body-camera recordings to model the interactions between police officers and community members in traffic stops. We use text and speech features to automatically measure linguistic aspects of the interaction, from discourse factors like conversational structure to social factors like respect. I describe the differences we find in the language directed toward black versus white community members, and offer suggestions for how these findings can be used to help improve the fraught relations between police officers and the communities they serve.

Biography: Dan Jurafsky is Professor and Chair of Linguistics and Professor of Computer Science, at Stanford University. His research has focused on the extraction of meaning, intention, and affect from text and speech, on the processing of Chinese, and on applying natural language processing to the cognitive and social sciences. Dan's deep interest in NLP education led him to co-write with Jim Martin the widely-used textbook "Speech and Language Processing" (whose 3rd edition is in (slow) progress) and co-teach with Chris Manning the first massive open online class on natural language processing. Dan was the recipient of the 2002 MacArthur Fellowship and is a 2015 James Beard Award Nominee for his book, "The Language of Food: A Linguist Reads the Menu".

Session 7 Overview – Monday, September 11, 2017

Oral tracks

Track A *Machine Learning 3* Jutland	**Track B** *Syntax 4* Funen	**Track C** *Dialogue* Zealand	
Maximum Margin Reward Networks for Learning from Explicit and Implicit Supervision *Peng, Chang, and Yih*	Part-of-Speech Tagging for Twitter with Adversarial Neural Networks *Gui, Zhang, Huang, Peng, and Huang*	Deal or No Deal? End-to-End Learning of Negotiation Dialogues *Lewis, Yarats, Dauphin, Parikh, and Batra*	10:30
The Impact of Modeling Overall Argumentation with Tree Kernels *Wachsmuth, Da San Martino, Kiesel, and Stein*	Investigating Different Syntactic Context Types and Context Representations for Learning Word Embeddings *Li, Liu, Zhao, Tang, Drozd, Rogers, and Du*	Agent-Aware Dropout DQN for Safe and Efficient Online Dialogue Policy Learning *Chen, Zhou, Chang, Yang, and Yu*	10:55
Learning Generic Sentence Representations Using Convolutional Neural Networks *Gan, Pu, Henao, Li, He, and Carin*	Does syntax help discourse segmentation? Not so much *Braud, Lacroix, and Søgaard*	Towards Debate Automation: a Recurrent Model for Predicting Debate Winners *Potash and Rumshisky*	11:20
Repeat before Forgetting: Spaced Repetition for Efficient and Effective Training of Neural Networks *Amiri, Miller, and Savova*	[TACL] Nonparametric Bayesian Semi-supervised Word Segmentation *Mochihashi, Fujii, and Domoto*	[TACL] Conversation Modeling on Reddit Using a Graph-Structured LSTM *Zayats and Ostendorf*	11:45

Poster tracks

10:30–12:10

Track D: *Poster Session. Machine Translation and Multilingual NLP 2* — Aarhus

Track E: *Poster Session. Information Extraction 2* — Odense

Track F: *Poster Session. NLP Applications* — Copenhagen

Parallel Session 7

Session 7A: Machine Learning 3

Jutland Chair: *Barbara Plank*

Maximum Margin Reward Networks for Learning from Explicit and Implicit Supervision
Haoruo Peng, Ming-Wei Chang, and Wen-tau Yih 10:30–10:55

Neural networks have achieved state-of-the-art performance on several structured-output prediction tasks, trained in a fully supervised fashion. However, annotated examples in structured domains are often costly to obtain, which thus limits the applications of neural networks. In this work, we propose Maximum Margin Reward Networks, a neural network-based framework that aims to learn from both explicit (full structures) and implicit supervision signals (delayed feedback on the correctness of the predicted structure). On named entity recognition and semantic parsing, our model outperforms previous systems on the benchmark datasets, CoNLL-2003 and WebQuestionsSP.

The Impact of Modeling Overall Argumentation with Tree Kernels
Henning Wachsmuth, Giovanni Da San Martino, Dora Kiesel, and Benno Stein 10:55–11:20

Several approaches have been proposed to model either the explicit sequential structure of an argumentative text or its implicit hierarchical structure. So far, the adequacy of these models of overall argumentation remains unclear. This paper asks what type of structure is actually important to tackle downstream tasks in computational argumentation. We analyze patterns in the overall argumentation of texts from three corpora. Then, we adapt the idea of positional tree kernels in order to capture sequential and hierarchical argumentative structure together for the first time. In systematic experiments for three text classification tasks, we find strong evidence for the impact of both types of structure. Our results suggest that either of them is necessary while their combination may be beneficial.

Learning Generic Sentence Representations Using Convolutional Neural Networks
Zhe Gan, Yunchen Pu, Ricardo Henao, Chunyuan Li, Xiaodong He, and Lawrence Carin 11:20–11:45

We propose a new encoder-decoder approach to learn distributed sentence representations that are applicable to multiple purposes. The model is learned by using a convolutional neural network as an encoder to map an input sentence into a continuous vector, and using a long short-term memory recurrent neural network as a decoder. Several tasks are considered, including sentence reconstruction and future sentence prediction. Further, a hierarchical encoder-decoder model is proposed to encode a sentence to predict multiple future sentences. By training our models on a large collection of novels, we obtain a highly generic convolutional sentence encoder that performs well in practice. Experimental results on several benchmark datasets, and across a broad range of applications, demonstrate the superiority of the proposed model over competing methods.

Repeat before Forgetting: Spaced Repetition for Efficient and Effective Training of Neural Networks
Hadi Amiri, Timothy Miller, and Guergana Savova 11:45–12:10

We present a novel approach for training artificial neural networks. Our approach is inspired by broad evidence in psychology that shows human learners can learn efficiently and effectively by increasing intervals of time between subsequent reviews of previously learned materials (spaced repetition). We investigate the analogy between training neural models and findings in psychology about human memory model and develop an efficient and effective algorithm to train neural models. The core part of our algorithm is a cognitively-motivated scheduler according to which training instances and their "reviews" are spaced over time. Our algorithm uses only 34-50% of data per epoch, is 2.9-4.8 times faster than standard training, and outperforms competing state-of-the-art baselines. Our code is available at scholar.harvard.edu/hadi/RbF/.

Session 7B: Syntax 4

Funen — Chair: *Zeljko Agic*

Part-of-Speech Tagging for Twitter with Adversarial Neural Networks
Tao Gui, Qi Zhang, Haoran Huang, Minlong Peng, and Xuanjing Huang 10:30–10:55

In this work, we study the problem of part-of-speech tagging for Tweets. In contrast to newswire articles, Tweets are usually informal and contain numerous out-of-vocabulary words. Moreover, there is a lack of large scale labeled datasets for this domain. To tackle these challenges, we propose a novel neural network to make use of out-of-domain labeled data, unlabeled in-domain data, and labeled in-domain data. Inspired by adversarial neural networks, the proposed method tries to learn common features through adversarial discriminator. In addition, we hypothesize that domain-specific features of target domain should be preserved in some degree. Hence, the proposed method adopts a sequence-to-sequence autoencoder to perform this task. Experimental results on three different datasets show that our method achieves better performance than state-of-the-art methods.

Investigating Different Syntactic Context Types and Context Representations for Learning Word Embeddings
Bofang Li, Tao Liu, Zhe Zhao, Buzhou Tang, Aleksandr Drozd, Anna Rogers, and Xiaoyong Du 10:55–11:20

The number of word embedding models is growing every year. Most of them are based on the co-occurrence information of words and their contexts. However, it is still an open question what is the best definition of context. We provide a systematical investigation of 4 different syntactic context types and context representations for learning word embeddings. Comprehensive experiments are conducted to evaluate their effectiveness on 6 extrinsic and intrinsic tasks. We hope that this paper, along with the published code, would be helpful for choosing the best context type and representation for a given task.

Does syntax help discourse segmentation? Not so much
Chloé Braud, Ophélie Lacroix, and Anders Søgaard 11:20–11:45

Discourse segmentation is the first step in building discourse parsers. Most work on discourse segmentation does not scale to real-world discourse parsing across languages, for two reasons: (i) models rely on constituent trees, and (ii) experiments have relied on gold standard identification of sentence and token boundaries. We therefore investigate to what extent constituents can be replaced with universal dependencies, or left out completely, as well as how state-of-the-art segmenters fare in the absence of sentence boundaries. Our results show that dependency information is less useful than expected, but we provide a fully scalable, robust model that only relies on part-of-speech information, and show that it performs well across languages in the absence of any gold-standard annotation.

[TACL] Nonparametric Bayesian Semi-supervised Word Segmentation
Daichi Mochihashi, Ryo Fujii, and Ryo Domoto 11:45–12:10

This paper presents a novel hybrid generative/discriminative model of word segmentation based on nonparametric Bayesian methods. Unlike ordinary discriminative word segmentation which relies only on labeled data, our semi-supervised model also leverages a huge amount of unlabeled text to automatically learn new "words", and further constrains them by using a labeled data to segment non-standard texts such as those found in social networking services. Specifically, our hybrid model combines a discriminative classifier (CRF; Lafferty et al. (2001)) and unsupervised word segmentation (NPYLM; Mochihashi et al. (2009)) with a transparent exchange of information between these two model structures within the semi-supervised framework (JESS-CM; Suzuki et al. (2008)). We confirmed that it can appropriately segment non-standard texts like those in Twitter and Weibo and has nearly state-of-the-art accuracy on standard datasets in Japanese, Chinese, and Thai.

Session 7C: Dialogue

Zealand — Chair: *Amanda Stent*

Deal or No Deal? End-to-End Learning of Negotiation Dialogues
Mike Lewis, Denis Yarats, Yann Dauphin, Devi Parikh, and Dhruv Batra 10:30–10:55

Much of human dialogue occurs in semi-cooperative settings, where agents with different goals attempt to agree on common decisions. Negotiations require complex communication and reasoning skills, but success is easy to measure, making this an interesting task for AI. We gather a large dataset of human-human negotiations on a multi-issue bargaining task, where agents who cannot observe each other's reward functions must reach an agreement (or a deal) via natural language dialogue. For the first time, we show it is possible to train end-to-end models for negotiation, which must learn both linguistic and reasoning skills with no annotated dialogue states. We also introduce dialogue rollouts, in which the model plans ahead by simulating possible complete continuations of the conversation, and find that this technique dramatically improves performance. Our code and dataset are publicly available.

Agent-Aware Dropout DQN for Safe and Efficient On-line Dialogue Policy Learning
Lu Chen, Xiang Zhou, Cheng Chang, Runzhe Yang, and Kai Yu 10:55–11:20

Hand-crafted rules and reinforcement learning (RL) are two popular choices to obtain dialogue policy. The rule-based policy is often reliable within predefined scope but not self-adaptable, whereas RL is evolvable with data but often suffers from a bad initial performance. We employ a *companion learning* framework to integrate the two approaches for *on-line* dialogue policy learning, in which a pre-defined rule-based policy acts as a "teacher" and guides a data-driven RL system by giving example actions as well as additional rewards. A novel *agent-aware dropout* Deep Q-Network (AAD-DQN) is proposed to address the problem of when to consult the teacher and how to learn from the teacher's experiences. AAD-DQN, as a data-driven student policy, provides (1) two separate experience memories for student and teacher, (2) an uncertainty estimated by dropout to control the timing of consultation and learning. Simulation experiments showed that the proposed approach can significantly improve both *safety* and *efficiency* of on-line policy optimization compared to other companion learning approaches as well as supervised pre-training using static dialogue corpus.

Towards Debate Automation: a Recurrent Model for Predicting Debate Winners
Peter Potash and Anna Rumshisky 11:20–11:45

In this paper we introduce a practical first step towards the creation of an automated debate agent: a state-of-the-art recurrent predictive model for predicting debate winners. By having an accurate predictive model, we are able to objectively rate the quality of a statement made at a specific turn in a debate. The model is based on a recurrent neural network architecture with attention, which allows the model to effectively account for the entire debate when making its prediction. Our model achieves state-of-the-art accuracy on a dataset of debate transcripts annotated with audience favorability of the debate teams. Finally, we discuss how future work can leverage our proposed model for the creation of an automated debate agent. We accomplish this by determining the model input that will maximize audience favorability toward a given side of a debate at an arbitrary turn.

[TACL] Conversation Modeling on Reddit Using a Graph-Structured LSTM
Victoria Zayats and Mari Ostendorf 11:45–12:10

This paper presents a novel approach for modeling threaded discussions on social media using a graph-structured bidirectional LSTM which represents both hierarchical and temporal conversation structure. In experiments with a task of predicting popularity of comments in Reddit discussions, the proposed model outperforms a node-independent architecture for different sets of input features. Analyses show a benefit to the model over the full course of the discussion, improving detection in both early and late stages. Further, the use of language cues with the bidirectional tree state updates helps with identifying controversial comments.

Session 7D: Poster Session. Machine Translation and Multilingual NLP 2

10:30–12:10
Aarhus
Chair: *Marianna Apidianaki*

[TACL] Joint Prediction of Word Alignment with Alignment Types

Anahita Mansouri Bigvand, Te Bu, and Anoop Sarkar

Current word alignment models do not distinguish between different types of alignment links. In this paper, we provide a new probabilistic model for word alignment where word alignments are associated with linguistically motivated alignment types. We propose a novel task of joint prediction of word alignment and alignment types and propose novel semi-supervised learning algorithms for this task. We also solve a sub-task of predicting the alignment type given an aligned word pair. In our experimental results, the generative models we introduce to model alignment types significantly outperform the models without alignment types.

Further Investigation into Reference Bias in Monolingual Evaluation of Machine Translation

Qingsong Ma, Yvette Graham, Timothy Baldwin, and Qun Liu

Monolingual evaluation of Machine Translation (MT) aims to simplify human assessment by requiring assessors to compare the meaning of the MT output with a reference translation, opening up the task to a much larger pool of genuinely qualified evaluators. Monolingual evaluation runs the risk, however, of bias in favour of MT systems that happen to produce translations superficially similar to the reference and, consistent with this intuition, previous investigations have concluded monolingual assessment to be strongly biased in this respect. On re-examination of past analyses, we identify a series of potential analytical errors that force some important questions to be raised about the reliability of past conclusions, however. We subsequently carry out further investigation into reference bias via direct human assessment of MT adequacy via quality controlled crowd-sourcing. Contrary to both intuition and past conclusions, results for show no significant evidence of reference bias in monolingual evaluation of MT.

A Challenge Set Approach to Evaluating Machine Translation

Pierre Isabelle, Colin Cherry, and George Foster

Neural machine translation represents an exciting leap forward in translation quality. But what long-standing weaknesses does it resolve, and which remain? We address these questions with a challenge set approach to translation evaluation and error analysis. A challenge set consists of a small set of sentences, each hand-designed to probe a system's capacity to bridge a particular structural divergence between languages. To exemplify this approach, we present an English-French challenge set, and use it to analyze phrase-based and neural systems. The resulting analysis provides not only a more fine-grained picture of the strengths of neural systems, but also insight into which linguistic phenomena remain out of reach.

Knowledge Distillation for Bilingual Dictionary Induction

Ndapandula Nakashole and Raphael Flauger

Leveraging zero-shot learning to learn mapping functions between vector spaces of different languages is a promising approach to bilingual dictionary induction. However, methods using this approach have not yet achieved high accuracy on the task. In this paper, we propose a bridging approach, where our main contribution is a knowledge distillation training objective. As teachers, rich resource translation paths are exploited in this role. And as learners, translation paths involving low resource languages learn from the teachers. Our training objective allows seamless addition of teacher translation paths for any given low resource pair. Since our approach relies on the quality of monolingual word embeddings, we also propose to enhance vector representations of both the source and target language with linguistic information. Our experiments on various languages show large performance gains from our distillation training objective, obtaining as high as 17% accuracy improvements.

Machine Translation, it's a question of style, innit? The case of English tag questions

Rachel Bawden

In this paper, we address the problem of generating English tag questions (TQs) (e.g. it is, isn't it?) in Machine Translation (MT). We propose a post-edition solution, formulating the problem as a multi-class classification task. We present (i) the automatic annotation of English TQs in a parallel corpus of subtitles and (ii) an approach using a series of classifiers to predict TQ forms, which we use to post-edit state-of-the-art MT outputs. Our method provides significant improvements in English TQ translation when

translating from Czech, French and German, in turn improving the fluidity, naturalness, grammatical correctness and pragmatic coherence of MT output.

Deciphering Related Languages
Nima Pourdamghani and Kevin Knight

We present a method for translating texts between close language pairs. The method does not require parallel data, and it does not require the languages to be written in the same script. We show results for six language pairs: Afrikaans/Dutch, Bosnian/Serbian, Danish/Swedish, Macedonian/Bulgarian, Malaysian/Indonesian, and Polish/Belorussian. We report BLEU scores showing our method to outperform others that do not use parallel data.

Identifying Cognate Sets Across Dictionaries of Related Languages
Adam St Arnaud, David Beck, and Grzegorz Kondrak

We present a system for identifying cognate sets across dictionaries of related languages. The likelihood of a cognate relationship is calculated on the basis of a rich set of features that capture both phonetic and semantic similarity, as well as the presence of regular sound correspondences. The similarity scores are used to cluster words from different languages that may originate from a common proto-word. When tested on the Algonquian language family, our system detects 63% of cognate sets while maintaining cluster purity of 70%.

Learning Language Representations for Typology Prediction
Chaitanya Malaviya, Graham Neubig, and Patrick Littell

One central mystery of neural NLP is what neural models "know" about their subject matter. When a neural machine translation system learns to translate from one language to another, does it learn the syntax or semantics of the languages? Can this knowledge be extracted from the system to fill holes in human scientific knowledge? Existing typological databases contain relatively full feature specifications for only a few hundred languages. Exploiting the existence of parallel texts in more than a thousand languages, we build a massive many-to-one NMT system from 1017 languages into English, and use this to predict information missing from typological databases. Experiments show that the proposed method is able to infer not only syntactic, but also phonological and phonetic inventory features, and improves over a baseline that has access to information about the languages geographic and phylogenetic neighbors.

Cheap Translation for Cross-Lingual Named Entity Recognition
Stephen Mayhew, Chen-Tse Tsai, and Dan Roth

Recent work in NLP has attempted to deal with low-resource languages but still assumed a resource level that is not present for most languages, e.g., the availability of Wikipedia in the target language. We propose a simple method for cross-lingual named entity recognition (NER) that works well in settings with *very* minimal resources. Our approach makes use of a lexicon to "translate" annotated data available in one or several high resource language(s) into the target language, and learns a standard monolingual NER model there. Further, when Wikipedia is available in the target language, our method can enhance Wikipedia based methods to yield state-of-the-art NER results; we evaluate on 7 diverse languages, improving the state-of-the-art by an average of 5.5% F1 points. With the minimal resources required, this is an extremely portable cross-lingual NER approach, as illustrated using a truly low-resource language, Uyghur.

Cross-Lingual Induction and Transfer of Verb Classes Based on Word Vector Space Specialisation
Ivan Vulić, Nikola Mrkšić, and Anna Korhonen

Existing approaches to automatic VerbNet-style verb classification are heavily dependent on feature engineering and therefore limited to languages with mature NLP pipelines. In this work, we propose a novel cross-lingual transfer method for inducing VerbNets for multiple languages. To the best of our knowledge, this is the first study which demonstrates how the architectures for learning word embeddings can be applied to this challenging syntactic-semantic task. Our method uses cross-lingual translation pairs to tie each of the six target languages into a bilingual vector space with English, jointly specialising the representations to encode the relational information from English VerbNet. A standard clustering algorithm is then run on top of the VerbNet-specialised representations, using vector dimensions as features for learning verb classes. Our results show that the proposed cross-lingual transfer approach sets new state-of-the-art verb classification performance across all six target languages explored in this work.

Classification of telicity using cross-linguistic annotation projection
Annemarie Friedrich and Damyana Gateva

This paper addresses the automatic recognition of telicity, an aspectual notion. A telic event includes a natural endpoint ("she walked home"), while an atelic event does not ("she walked around"). Recognizing this difference is a prerequisite for temporal natural language understanding. In English, this classification task is difficult, as telicity is a covert linguistic category. In contrast, in Slavic languages, aspect is part of a verb's meaning and even available in machine-readable dictionaries. Our contributions are as follows. We successfully leverage additional silver standard training data in the form of projected annotations from parallel English-Czech data as well as context information, improving automatic telicity classification for English significantly compared to previous work. We also create a new data set of English texts manually annotated with telicity.

[TACL] Semantic Specialisation of Distributional Word Vector Spaces using Monolingual and Cross-Lingual Constraints
Nikola Mrkšić, Ivan Vulić, Diarmuid Ó Séaghdha, Roi Reichart, Ira Leviant, Milica Gašić, Anna Korhonen, and Steve Young

We present Attract-Repel, an algorithm for improving the semantic quality of word vectors by injecting constraints extracted from lexical resources. Attract-Repel facilitates the use of constraints from mono- and cross-lingual resources, yielding semantically specialised cross-lingual vector spaces. Our evaluation shows that the method can make use of existing cross-lingual lexicons to construct high-quality vector spaces for a plethora of different languages, facilitating semantic transfer from high- to lower-resource ones. The effectiveness of our approach is demonstrated with state-of-the-art results on semantic similarity datasets in six languages. We next show that Attract-Repel-specialised vectors boost performance in the downstream task of dialogue state tracking (DST) across multiple languages. Finally, we show that cross-lingual vector spaces produced by our algorithm facilitate the training of multilingual DST models, which brings further performance improvements.

Counterfactual Learning from Bandit Feedback under Deterministic Logging : A Case Study in Statistical Machine Translation
Carolin Lawrence, Artem Sokolov, and Stefan Riezler

The goal of counterfactual learning for statistical machine translation (SMT) is to optimize a target SMT system from logged data that consist of user feedback to translations that were predicted by another, historic SMT system. A challenge arises by the fact that risk-averse commercial SMT systems deterministically log the most probable translation. The lack of sufficient exploration of the SMT output space seemingly contradicts the theoretical requirements for counterfactual learning. We show that counterfactual learning from deterministic bandit logs is possible nevertheless by smoothing out deterministic components in learning. This can be achieved by additive and multiplicative control variates that avoid degenerate behavior in empirical risk minimization. Our simulation experiments show improvements of up to 2 BLEU points by counterfactual learning from deterministic bandit feedback.

Session 7E: Poster Session. Information Extraction 2

10:30–12:10
Odense
Chair: *Isabelle Augenstein*

Learning Fine-grained Relations from Chinese User Generated Categories
Chengyu Wang, Yan Fan, Xiaofeng He, and Aoying Zhou

User generated categories (UGCs) are short texts that reflect how people describe and organize entities, expressing rich semantic relations implicitly. While most methods on UGC relation extraction are based on pattern matching in English circumstances, learning relations from Chinese UGCs poses different challenges due to the flexibility of expressions. In this paper, we present a weakly supervised learning framework to harvest relations from Chinese UGCs. We identify is-a relations via word embedding based projection and inference, extract non-taxonomic relations and their category patterns by graph mining. We conduct experiments on Chinese Wikipedia and achieve high accuracy, outperforming state-of-the-art methods.

Improving Slot Filling Performance with Attentive Neural Networks on Dependency Structures
Lifu Huang, Avirup Sil, Heng Ji, and Radu Florian

Slot Filling (SF) aims to extract the values of certain types of attributes (or slots, such as person:cities_of_residence) for a given entity from a large collection of source documents. In this paper we propose an effective DNN architecture for SF with the following new strategies: (1). Take a regularized dependency graph instead of a raw sentence as input to DNN, to compress the wide contexts between query and candidate filler; (2). Incorporate two attention mechanisms: local attention learned from query and candidate filler, and global attention learned from external knowledge bases, to guide the model to better select indicative contexts to determine slot type. Experiments show that this framework outperforms state-of-the-art on both relation extraction (16% absolute F-score gain) and slot filling validation for each individual system (up to 8.5% absolute F-score gain).

Identifying Products in Online Cybercrime Marketplaces: A Dataset for Fine-grained Domain Adaptation
Greg Durrett, Jonathan K. Kummerfeld, Taylor Berg-Kirkpatrick, Rebecca Portnoff, Sadia Afroz, Damon McCoy, Kirill Levchenko, and Vern Paxson

One weakness of machine-learned NLP models is that they typically perform poorly on out-of-domain data. In this work, we study the task of identifying products being bought and sold in online cybercrime forums, which exhibits particularly challenging cross-domain effects. We formulate a task that represents a hybrid of slot-filling information extraction and named entity recognition and annotate data from four different forums. Each of these forums constitutes its own "fine-grained domain" in that the forums cover different market sectors with different properties, even though all forums are in the broad domain of cybercrime. We characterize these domain differences in the context of a learning-based system: supervised models see decreased accuracy when applied to new forums, and standard techniques for semi-supervised learning and domain adaptation have limited effectiveness on this data, which suggests the need to improve these techniques. We release a dataset of 1,938 annotated posts from across the four forums.

Labeling Gaps Between Words: Recognizing Overlapping Mentions with Mention Separators
Aldrian Obaja Muis and Wei Lu

In this paper, we propose a new model that is capable of recognizing overlapping mentions. We introduce a novel notion of mention separators that can be effectively used to capture how mentions overlap with one another. On top of a novel multigraph representation that we introduce, we show that efficient and exact inference can still be performed. We present some theoretical analysis on the differences between our model and a recently proposed model for recognizing overlapping mentions, and discuss the possible implications of the differences. Through extensive empirical analysis on standard datasets, we demonstrate the effectiveness of our approach.

Deep Joint Entity Disambiguation with Local Neural Attention
Octavian-Eugen Ganea and Thomas Hofmann

We propose a novel deep learning model for joint document-level entity disambiguation, which leverages learned neural representations. Key components are entity embeddings, a neural attention mechanism over local context windows, and a differentiable joint inference stage for disambiguation. Our approach thereby combines benefits of deep learning with more traditional approaches such as graphical models

and probabilistic mention-entity maps. Extensive experiments show that we are able to obtain competitive or state-of-the-art accuracy at moderate computational costs.

MinIE: Minimizing Facts in Open Information Extraction
Kiril Gashteovski, Rainer Gemulla, and Luciano Del Corro

The goal of Open Information Extraction (OIE) is to extract surface relations and their arguments from natural-language text in an unsupervised, domain-independent manner. In this paper, we propose MinIE, an OIE system that aims to provide useful, compact extractions with high precision and recall. MinIE approaches these goals by (1) representing information about polarity, modality, attribution, and quantities with semantic annotations instead of in the actual extraction, and (2) identifying and removing parts that are considered overly specific. We conducted an experimental study with several real-world datasets and found that MinIE achieves competitive or higher precision and recall than most prior systems, while at the same time producing shorter, semantically enriched extractions.

Scientific Information Extraction with Semi-supervised Neural Tagging
Yi Luan, Mari Ostendorf, and Hannaneh Hajishirzi

This paper addresses the problem of extracting keyphrases from scientific articles and categorizing them as corresponding to a task, process, or material. We cast the problem as sequence tagging and introduce semi-supervised methods to a neural tagging model, which builds on recent advances in named entity recognition. Since annotated training data is scarce in this domain, we introduce a graph-based semi-supervised algorithm together with a data selection scheme to leverage unannotated articles. Both inductive and transductive semi-supervised learning strategies outperform state-of-the-art information extraction performance on the 2017 SemEval Task 10 ScienceIE task.

NITE: A Neural Inductive Teaching Framework for Domain Specific NER
Siliang Tang, Ning Zhang, Jinjiang Zhang, Fei Wu, and Yueting Zhuang

In domain-specific NER, due to insufficient labeled training data, deep models usually fail to behave normally. In this paper, we proposed a novel Neural Inductive TEaching framework (NITE) to transfer knowledge from existing domain-specific NER models into an arbitrary deep neural network in a teacher-student training manner. NITE is a general framework that builds upon transfer learning and multiple instance learning, which collaboratively not only transfers knowledge to a deep student network but also reduces the noise from teachers. NITE can help deep learning methods to effectively utilize existing resources (i.e., models, labeled and unlabeled data) in a small domain. The experiment resulted on Disease NER proved that without using any labeled data, NITE can significantly boost the performance of a CNN-bidirectional LSTM-CRF NER neural network nearly over 30% in terms of F1-score.

Speeding up Reinforcement Learning-based Information Extraction Training using Asynchronous Methods
Aditya Sharma, Zarana Parekh, and Partha Talukdar

RLIE-DQN is a recently proposed Reinforcement Learning-based Information Extraction (IE) technique which is able to incorporate external evidence during the extraction process. RLIE-DQN trains a single agent sequentially, training on one instance at a time. This results in significant training slowdown which is undesirable. We leverage recent advances in parallel RL training using asynchronous methods and propose RLIE-A3C. RLIE-A3C trains multiple agents in parallel and is able to achieve upto 6x training speedup over RLIE-DQN, while suffering no loss in average accuracy.

Leveraging Linguistic Structures for Named Entity Recognition with Bidirectional Recursive Neural Networks
Peng-Hsuan Li, Ruo-Ping Dong, Yu-Siang Wang, Ju-Chieh Chou, and Wei-Yun Ma

In this paper, we utilize the linguistic structures of texts to improve named entity recognition by BRNN-CNN, a special bidirectional recursive network attached with a convolutional network. Motivated by the observation that named entities are highly related to linguistic constituents, we propose a constituent-based BRNN-CNN for named entity recognition. In contrast to classical sequential labeling methods, the system first identifies which text chunks are possible named entities by whether they are linguistic constituents. Then it classifies these chunks with a constituency tree structure by recursively propagating syntactic and semantic information to each constituent node. This method surpasses current state-of-the-art on OntoNotes 5.0 with automatically generated parses.

Fast and Accurate Entity Recognition with Iterated Dilated Convolutions
Emma Strubell, Patrick Verga, David Belanger, and Andrew McCallum

Today when many practitioners run basic NLP on the entire web and large-volume traffic, faster methods are paramount to saving time and energy costs. Recent advances in GPU hardware have led to the emergence of bi-directional LSTMs as a standard method for obtaining per-token vector representations serving as input to labeling tasks such as NER (often followed by prediction in a linear-chain CRF). Though expressive and accurate, these models fail to fully exploit GPU parallelism, limiting their computational efficiency. This paper proposes a faster alternative to Bi-LSTMs for NER: Iterated Dilated Convolutional Neural Networks (ID-CNNs), which have better capacity than traditional CNNs for large context and structured prediction. Unlike LSTMs whose sequential processing on sentences of length N requires O(N) time even in the face of parallelism, ID-CNNs permit fixed-depth convolutions to run in parallel across entire documents. We describe a distinct combination of network structure, parameter sharing and training procedures that enable dramatic 14-20x test-time speedups while retaining accuracy comparable to the Bi-LSTM-CRF. Moreover, ID-CNNs trained to aggregate context from the entire document are more accurate than Bi-LSTM-CRFs while attaining 8x faster test time speeds.

Entity Linking via Joint Encoding of Types, Descriptions, and Context
Nitish Gupta, Sameer Singh, and Dan Roth

For accurate entity linking, we need to capture various information aspects of an entity, such as its description in a KB, contexts in which it is mentioned, and structured knowledge. Additionally, a linking system should work on texts from different domains without requiring domain-specific training data or hand-engineered features. In this work we present a neural, modular entity linking system that learns a unified dense representation for each entity using multiple sources of information, such as its description, contexts around its mentions, and its fine-grained types. We show that the resulting entity linking system is effective at combining these sources, and performs competitively, sometimes out-performing current state-of-the-art systems across datasets, without requiring any domain-specific training data or hand-engineered features. We also show that our model can effectively "embed" entities that are new to the KB, and is able to link its mentions accurately.

An Insight Extraction System on BioMedical Literature with Deep Neural Networks
Hua He, Kris Ganjam, Navendu Jain, Jessica Lundin, Ryen White, and Jimmy Lin

Mining biomedical text offers an opportunity to automatically discover important facts and infer associations among them. As new scientific findings appear across a large collection of biomedical publications, our aim is to tap into this literature to automate biomedical knowledge extraction and identify important insights from them. Towards that goal, we develop a system with novel deep neural networks to extract insights on biomedical literature. Evaluation shows our system is able to provide insights with competitive accuracy of human acceptance and its relation extraction component outperforms previous work.

Session 7F: Poster Session. NLP Applications

10:30–12:10
Copenhagen
Chair: *Courtney Napoles*

Word Etymology as Native Language Interference
Vivi Nastase and Carlo Strapparava

We present experiments that show the influence of native language on lexical choice when producing text in another language – in this particular case English. We start from the premise that non-native English speakers will choose lexical items that are close to words in their native language. This leads us to an etymology-based representation of documents written by people whose mother tongue is an Indo-European language. Based on this representation we grow a language family tree, that matches closely the Indo-European language tree.

A Simpler and More Generalizable Story Detector using Verb and Character Features
Joshua Eisenberg and Mark Finlayson

Story detection is the task of determining whether or not a unit of text contains a story. Prior approaches achieved a maximum performance of 0.66 F1, and did not generalize well across different corpora. We present a new state-of-the-art detector that achieves a maximum performance of 0.75 F1 (a 14% improvement), with significantly greater generalizability than previous work. In particular, our detector achieves performance above 0.70 F1 across a variety of combinations of lexically different corpora for training and testing, as well as dramatic improvements (up to 4,000%) in performance when trained on a small, disfluent data set. The new detector uses two basic types of features—ones related to events, and ones related to characters—totaling 283 specific features overall; previous detectors used tens of thousands of features, and so this detector represents a significant simplification along with increased performance.

Multi-modular domain-tailored OCR post-correction
Sarah Schulz and Jonas Kuhn

One of the main obstacles for many Digital Humanities projects is the low data availability. Texts have to be digitized in an expensive and time consuming process whereas Optical Character Recognition (OCR) post-correction is one of the time-critical factors. At the example of OCR post-correction, we show the adaptation of a generic system to solve a specific problem with little data. The system accounts for a diversity of errors encountered in OCRed texts coming from different time periods in the domain of literature. We show that the combination of different approaches, such as e.g. Statistical Machine Translation and spell checking, with the help of a ranking mechanism tremendously improves over single-handed approaches. Since we consider the accessibility of the resulting tool as a crucial part of Digital Humanities collaborations, we describe the workflow we suggest for efficient text recognition and subsequent automatic and manual post-correction

Learning to Predict Charges for Criminal Cases with Legal Basis
Bingfeng Luo, Yansong Feng, Jianbo Xu, Xiang Zhang, and Dongyan Zhao

The charge prediction task is to determine appropriate charges for a given case, which is helpful for legal assistant systems where the user input is fact description. We argue that relevant law articles play an important role in this task, and therefore propose an attention-based neural network method to jointly model the charge prediction task and the relevant article extraction task in a unified framework. The experimental results show that, besides providing legal basis, the relevant articles can also clearly improve the charge prediction results, and our full model can effectively predict appropriate charges for cases with different expression styles.

Quantifying the Effects of Text Duplication on Semantic Models
Alexandra Schofield, Laure Thompson, and David Mimno

Duplicate documents are a pervasive problem in text datasets and can have a strong effect on unsupervised models. Methods to remove duplicate texts are typically heuristic or very expensive, so it is vital to know when and why they are needed. We measure the sensitivity of two latent semantic methods to the presence of different levels of document repetition. By artificially creating different forms of duplicate text we confirm several hypotheses about how repeated text impacts models. While a small amount of duplication is tolerable, substantial over-representation of subsets of the text may overwhelm meaningful topical patterns.

Identifying Semantically Deviating Outlier Documents
Honglei Zhuang, Chi Wang, Fangbo Tao, Lance Kaplan, and Jiawei Han

A document outlier is a document that substantially deviates in semantics from the majority ones in a corpus. Automatic identification of document outliers can be valuable in many applications, such as screening health records for medical mistakes. In this paper, we study the problem of mining semantically deviating document outliers in a given corpus. We develop a generative model to identify frequent and characteristic semantic regions in the word embedding space to represent the given corpus, and a robust outlierness measure which is resistant to noisy content in documents. Experiments conducted on two real-world textual data sets show that our method can achieve an up to 135% improvement over baselines in terms of recall at top-1% of the outlier ranking.

Detecting and Explaining Causes From Text For a Time Series Event
Dongyeop Kang, Varun Gangal, Ang Lu, Zheng Chen, and Eduard Hovy

Explaining underlying causes or effects about events is a challenging but valuable task. We define a novel problem of generating explanations of a time series event by (1) searching cause and effect relationships of the time series with textual data and (2) constructing a connecting chain between them to generate an explanation. To detect causal features from text, we propose a novel method based on the Granger causality of time series between features extracted from text such as N-grams, topics, sentiments, and their composition. The generation of the sequence of causal entities requires a commonsense causative knowledge base with efficient reasoning. To ensure good interpretability and appropriate lexical usage we combine symbolic and neural representations, using a neural reasoning algorithm trained on commonsense causal tuples to predict the next cause step. Our quantitative and human analysis show empirical evidence that our method successfully extracts meaningful causality relationships between time series with textual features and generates appropriate explanation between them.

A Novel Cascade Model for Learning Latent Similarity from Heterogeneous Sequential Data of MOOC
Zhuoxuan Jiang, Shanshan Feng, Gao Cong, Chunyan Miao, and Xiaoming Li

Recent years have witnessed the proliferation of Massive Open Online Courses (MOOCs). With massive learners being offered MOOCs, there is a demand that the forum contents within MOOCs need to be classified in order to facilitate both learners and instructors. Therefore we investigate a significant application, which is to associate forum threads to subtitles of video clips. This task can be regarded as a document ranking problem, and the key is how to learn a distinguishable text representation from word sequences and learners' behavior sequences. In this paper, we propose a novel cascade model, which can capture both the latent semantics and latent similarity by modeling MOOC data. Experimental results on two real-world datasets demonstrate that our textual representation outperforms state-of-the-art unsupervised counterparts for the application.

Identifying the Provision of Choices in Privacy Policy Text
Kanthashree Mysore Sathyendra, Shomir Wilson, Florian Schaub, Sebastian Zimmeck, and Norman Sadeh

Websites' and mobile apps' privacy policies, written in natural language, tend to be long and difficult to understand. Information privacy revolves around the fundamental principle of Notice and choice, namely the idea that users should be able to make informed decisions about what information about them can be collected and how it can be used. Internet users want control over their privacy, but their choices are often hidden in long and convoluted privacy policy texts. Moreover, little (if any) prior work has been done to detect the provision of choices in text. We address this challenge of enabling user choice by automatically identifying and extracting pertinent choice language in privacy policies. In particular, we present a two-stage architecture of classification models to identify opt-out choices in privacy policy text, labelling common varieties of choices with a mean F1 score of 0.735. Our techniques enable the creation of systems to help Internet users to learn about their choices, thereby effectuating notice and choice and improving Internet privacy.

An Empirical Analysis of Edit Importance between Document Versions
Tanya Goyal, Sachin Kelkar, Manas Agarwal, and Jeenu Grover

In this paper, we present a novel approach to infer significance of various textual edits to documents. An author may make several edits to a document; each edit varies in its impact to the content of the document. While some edits are surface changes and introduce negligible change, other edits may change the content/tone of the document significantly. In this paper, we perform an analysis on the human perceptions of edit importance while reviewing documents from one version to the next. We identify

linguistic features that influence edit importance and model it in a regression based setting. We show that the predicted importance by our approach is highly correlated with the human perceived importance, established by a Mechanical Turk study.

Transition-Based Disfluency Detection using LSTMs
Shaolei Wang, Wanxiang Che, Yue Zhang, Meishan Zhang, and Ting Liu

In this paper, we model the problem of disfluency detection using a transition-based framework, which incrementally constructs and labels the disfluency chunk of input sentences using a new transition system without syntax information. Compared with sequence labeling methods, it can capture non-local chunk-level features; compared with joint parsing and disfluency detection methods, it is free for noise in syntax. Experiments show that our model achieves state-of-the-art f-score of 87.5% on the commonly used English Switchboard test set, and a set of in-house annotated Chinese data.

Neural Sequence-Labelling Models for Grammatical Error Correction
Helen Yannakoudakis, Marek Rei, Øistein E. Andersen, and Zheng Yuan

We propose an approach to N -best list re- ranking using neural sequence-labelling models. We train a compositional model for error detection that calculates the prob- ability of each token in a sentence being correct or incorrect, utilising the full sen- tence as context. Using the error detec- tion model, we then re-rank the N best hypotheses generated by statistical ma- chine translation systems. Our approach achieves state-of-the-art results on error correction for three different datasets, and it has the additional advantage of only us- ing a small set of easily computed features that require no linguistic input.

Adapting Sequence Models for Sentence Correction
Allen Schmaltz, Yoon Kim, Alexander Rush, and Stuart Shieber

In a controlled experiment of sequence-to-sequence approaches for the task of sentence correction, we find that character-based models are generally more effective than word-based models and models that encode subword information via convolutions, and that modeling the output data as a series of diffs improves effectiveness over standard approaches. Our strongest sequence-to-sequence model improves over our strongest phrase-based statistical machine translation model, with access to the same data, by 6 M2 (0.5 GLEU) points. Additionally, in the data environment of the standard CoNLL-2014 setup, we demonstrate that modeling (and tuning against) diffs yields similar or better M2 scores with simpler models and/or significantly less data than previous sequence-to-sequence approaches.

Session 8 Overview – Monday, September 11, 2017

Oral tracks

	Track A *Machine Translation and Multilingual/Multimodal NLP (Short)* Jutland	**Track B** *Machine Learning (Short)* Funen	**Track C** *NLP Applications (Short)* Zealand
13:40	A Study of Style in Machine Translation: Controlling the Formality of Machine Translation Output *Niu, Martindale, and Carpuat*	Sequence Effects in Crowdsourced Annotations *Mathur, Baldwin, and Cohn*	Learning what to read: Focused machine reading *Noriega-Atala, Valenzuela-Escárcega, Morrison, and Surdeanu*
13:55	Sharp Models on Dull Hardware: Fast and Accurate Neural Machine Translation Decoding on the CPU *Devlin*	No Need to Pay Attention: Simple Recurrent Neural Networks Work! *Ture and Jojic*	DOC: Deep Open Classification of Text Documents *Shu, Xu, and Liu*
14:10	Exploiting Cross-Sentence Context for Neural Machine Translation *Wang, Tu, Way, and Liu*	The strange geometry of skip-gram with negative sampling *Mimno and Thompson*	Charmanteau: Character Embedding Models For Portmanteau Creation *Gangal, Jhamtani, Neubig, Hovy, and Nyberg*
14:25	Cross-Lingual Transfer Learning for POS Tagging without Cross-Lingual Resources *Kim, Kim, Sarikaya, and Fosler-Lussier*	Natural Language Processing with Small Feed-Forward Networks *Botha, Pitler, Ma, Bakalov, Salcianu, Weiss, McDonald, and Petrov*	Using Automated Metaphor Identification to Aid in Detection and Prediction of First-Episode Schizophrenia *Gutierrez, Cecchi, Corcoran, and Corlett*
14:40	Image Pivoting for Learning Multilingual Multimodal Representations *Gella, Sennrich, Keller, and Lapata*	Deep Multi-Task Learning for Aspect Term Extraction with Memory Interaction *Li and Lam*	Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking *Rashkin, Choi, Jang, Volkova, and Choi*
14:55	Neural Machine Translation with Source Dependency Representation *Chen, Wang, Utiyama, Liu, Tamura, Sumita, and Zhao*	Analogs of Linguistic Structure in Deep Representations *Andreas and Klein*	Topic-Based Agreement and Disagreement in US Electoral Manifestos *Menini, Nanni, Ponzetto, and Tonelli*
15:10	Visual Denotations for Recognizing Textual Entailment *Han, Martínez-Gómez, and Mineshima*	A Simple Regularization-based Algorithm for Learning Cross-Domain Word Embeddings *Yang, Lu, and Zheng*	Zipporah: a Fast and Scalable Data Cleaning System for Noisy Web-Crawled Parallel Corpora *Xu and Koehn*

Parallel Session 8

Session 8A: Machine Translation and Multilingual/Multimodal NLP (Short)

Jutland Chair: *Yulia Tsvetkov*

A Study of Style in Machine Translation: Controlling the Formality of Machine Translation Output
Xing Niu, Marianna Martindale, and Marine Carpuat 13:40–13:55

Stylistic variations of language, such as formality, carry speakers' intention beyond literal meaning and should be conveyed adequately in translation. We propose to use lexical formality models to control the formality level of machine translation output. We demonstrate the effectiveness of our approach in empirical evaluations, as measured by automatic metrics and human assessments.

Sharp Models on Dull Hardware: Fast and Accurate Neural Machine Translation Decoding on the CPU
Jacob Devlin 13:55–14:10

Attentional sequence-to-sequence models have become the new standard for machine translation, but one challenge of such models is a significant increase in training and decoding cost compared to phrase-based systems. In this work we focus on efficient decoding, with a goal of achieving accuracy close the state-of-the-art in neural machine translation (NMT), while achieving CPU decoding speed/throughput close to that of a phrasal decoder. We approach this problem from two angles: First, we describe several techniques for speeding up an NMT beam search decoder, which obtain a 4.4x speedup over a very efficient baseline decoder without changing the decoder output. Second, we propose a simple but powerful network architecture which uses an RNN (GRU/LSTM) layer at bottom, followed by a series of stacked fully-connected layers applied at every timestep. This architecture achieves similar accuracy to a deep recurrent model, at a small fraction of the training and decoding cost. By combining these techniques, our best system achieves a very competitive accuracy of 38.3 BLEU on WMT English-French NewsTest2014, while decoding at 100 words/sec on single-threaded CPU. We believe this is the best published accuracy/speed trade-off of an NMT system.

Exploiting Cross-Sentence Context for Neural Machine Translation
Longyue Wang, Zhaopeng Tu, Andy Way, and Qun Liu 14:10–14:25

In translation, considering the document as a whole can help to resolve ambiguities and inconsistencies. In this paper, we propose a cross-sentence context-aware approach and investigate the influence of historical contextual information on the performance of neural machine translation (NMT). First, this history is summarized in a hierarchical way. We then integrate the historical representation into NMT in two strategies: 1) a warm-start of encoder and decoder states, and 2) an auxiliary context source for updating decoder states. Experimental results on a large Chinese-English translation task show that our approach significantly improves upon a strong attention-based NMT system by up to +2.1 BLEU points.

Cross-Lingual Transfer Learning for POS Tagging without Cross-Lingual Resources
Joo-Kyung Kim, Young-Bum Kim, Ruhi Sarikaya, and Eric Fosler-Lussier 14:25–14:40

Training a POS tagging model with crosslingual transfer learning usually requires linguistic knowledge and resources about the relation between the source language and the target language. In this paper, we introduce a cross-lingual transfer learning model for POS tagging without ancillary resources such as parallel corpora. The proposed cross-lingual model utilizes a common BLSTM that enables knowledge transfer from other languages, and private BLSTMs for language-specific representations. The cross-lingual model is trained with language-adversarial training and bidirectional language modeling as auxiliary objectives to better represent language-general information while not losing the information about a specific target language. Evaluating on POS datasets from 14 languages in the Universal Dependencies corpus, we show that the proposed transfer learning model improves the POS tagging performance of the target languages without exploiting any linguistic knowledge between the source language and the target language.

Image Pivoting for Learning Multilingual Multimodal Representations
Spandana Gella, Rico Sennrich, Frank Keller, and Mirella Lapata 14:40–14:55

In this paper we propose a model to learn multimodal multilingual representations for matching images and sentences in different languages, with the aim of advancing multilingual versions of image search and image understanding. Our model learns a common representation for images and their descriptions

in two different languages (which need not be parallel) by considering the image as a pivot between two languages. We introduce a new pairwise ranking loss function which can handle both symmetric and asymmetric similarity between the two modalities. We evaluate our models on image-description ranking for German and English, and on semantic textual similarity of image descriptions in English. In both cases we achieve state-of-the-art performance.

Neural Machine Translation with Source Dependency Representation
Kehai Chen, Rui Wang, Masao Utiyama, Lemao Liu, Akihiro Tamura, Eiichiro Sumita, and Tiejun Zhao 14:55–15:10

Source dependency information has been successfully introduced into statistical machine translation. However, there are only a few preliminary attempts for Neural Machine Translation (NMT), such as concatenating representations of source word and its dependency label together. In this paper, we propose a novel NMT with source dependency representation to improve translation performance of NMT, especially long sentences. Empirical results on NIST Chinese-to-English translation task show that our method achieves 1.6 BLEU improvements on average over a strong NMT system.

Visual Denotations for Recognizing Textual Entailment
Dan Han, Pascual Martínez-Gómez, and Koji Mineshima 15:10–15:25

In the logic approach to Recognizing Textual Entailment, identifying phrase-to-phrase semantic relations is still an unsolved problem. Resources such as the Paraphrase Database offer limited coverage despite their large size whereas unsupervised distributional models of meaning often fail to recognize phrasal entailments. We propose to map phrases to their visual denotations and compare their meaning in terms of their images. We show that our approach is effective in the task of Recognizing Textual Entailment when combined with specific linguistic and logic features.

Session 8B: Machine Learning (Short)

Funen Chair: *Samuel R. Bowman*

Sequence Effects in Crowdsourced Annotations
Nitika Mathur, Timothy Baldwin, and Trevor Cohn 13:40–13:55

Manual data annotation is a vital component of NLP research. When designing annotation tasks, properties of the annotation interface can unintentionally lead to artefacts in the resulting dataset, biasing the evaluation. In this paper, we explore sequence effects where annotations of an item are affected by the preceding items. Having assigned one label to an instance, the annotator may be less (or more) likely to assign the same label to the next. During rating tasks, seeing a low quality item may affect the score given to the next item either positively or negatively. We see clear evidence of both types of effects using auto-correlation studies over three different crowdsourced datasets. We then recommend a simple way to minimise sequence effects.

No Need to Pay Attention: Simple Recurrent Neural Networks Work!
Ferhan Ture and Oliver Jojic 13:55–14:10

First-order factoid question answering assumes that the question can be answered by a single fact in a knowledge base (KB). While this does not seem like a challenging task, many recent attempts that apply either complex linguistic reasoning or deep neural networks achieve 65%–76% accuracy on benchmark sets. Our approach formulates the task as two machine learning problems: detecting the entities in the question, and classifying the question as one of the relation types in the KB. We train a recurrent neural network to solve each problem. On the SimpleQuestions dataset, our approach yields substantial improvements over previously published results — even neural networks based on much more complex architectures. The simplicity of our approach also has practical advantages, such as efficiency and modularity, that are valuable especially in an industry setting. In fact, we present a preliminary analysis of the performance of our model on real queries from Comcast's X1 entertainment platform with millions of users every day.

The strange geometry of skip-gram with negative sampling
David Mimno and Laure Thompson 14:10–14:25

Despite their ubiquity, word embeddings trained with skip-gram negative sampling (SGNS) remain poorly understood. We find that vector positions are not simply determined by semantic similarity, but rather occupy a narrow cone, diametrically opposed to the context vectors. We show that this geometric concentration depends on the ratio of positive to negative examples, and that it is neither theoretically nor empirically inherent in related embedding algorithms.

Natural Language Processing with Small Feed-Forward Networks
Jan A. Botha, Emily Pitler, Ji Ma, Anton Bakalov, Alex Salcianu, David Weiss, Ryan McDonald, and Slav Petrov 14:25–14:40

We show that small and shallow feed-forward neural networks can achieve near state-of-the-art results on a range of unstructured and structured language processing tasks while being considerably cheaper in memory and computational requirements than deep recurrent models. Motivated by resource-constrained environments like mobile phones, we showcase simple techniques for obtaining such small neural network models, and investigate different tradeoffs when deciding how to allocate a small memory budget.

Deep Multi-Task Learning for Aspect Term Extraction with Memory Interaction
Xin Li and Wai Lam 14:40–14:55

We propose a novel LSTM-based deep multi-task learning framework for aspect term extraction from user review sentences. Two LSTMs equipped with extended memories and neural memory operations are designed for jointly handling the extraction tasks of aspects and opinions via memory interactions. Sentimental sentence constraint is also added for more accurate prediction via another LSTM. Experiment results over two benchmark datasets demonstrate the effectiveness of our framework.

Analogs of Linguistic Structure in Deep Representations
Jacob Andreas and Dan Klein 14:55–15:10

We investigate the compositional structure of message vectors computed by a deep network trained on a communication game. By comparing truth-conditional representations of encoder-produced message vectors to human-produced referring expressions, we are able to identify aligned (vector, utterance) pairs with the same meaning. We then search for structured relationships among these aligned pairs to discover simple vector space transformations corresponding to negation, conjunction, and disjunction. Our results

suggest that neural representations are capable of spontaneously developing a "syntax" with functional analogues to qualitative properties of natural language.

A Simple Regularization-based Algorithm for Learning Cross-Domain Word Embeddings
Wei Yang, Wei Lu, and Vincent Zheng 15:10–15:25

Learning word embeddings has received a significant amount of attention recently. Often, word embeddings are learned in an unsupervised manner from a large collection of text. The genre of the text typically plays an important role in the effectiveness of the resulting embeddings. How to effectively train word embedding models using data from different domains remains a problem that is less explored. In this paper, we present a simple yet effective method for learning word embeddings based on text from different domains. We demonstrate the effectiveness of our approach through extensive experiments on various down-stream NLP tasks.

Session 8C: NLP Applications (Short)

Zealand Chair: *Joel Tetreault*

Learning what to read: Focused machine reading
Enrique Noriega-Atala, Marco A. Valenzuela-Escárcega, Clayton Morrison, and Mihai Surdeanu
13:40–13:55

Recent efforts in bioinformatics have achieved tremendous progress in the ma- chine reading of biomedical literature, and the assembly of the extracted biochem- ical interactions into large-scale models such as protein signaling pathways. How- ever, batch machine reading of literature at today's scale (PubMed alone indexes over 1 million papers per year) is unfea- sible due to both cost and processing over- head. In this work, we introduce a focused reading approach to guide the machine reading of biomedical literature towards what literature should be read to answer a biomedical query as efficiently as pos- sible. We introduce a family of algorithms for focused reading, including an intuitive, strong baseline, and a second approach which uses a reinforcement learning (RL) framework that learns when to explore (widen the search) or exploit (narrow it). We demonstrate that the RL approach is capable of answering more queries than the baseline, while being more efficient, i.e., reading fewer documents.

DOC: Deep Open Classification of Text Documents
Lei Shu, Hu Xu, and Bing Liu 13:55–14:10

Traditional supervised learning makes the closed-world assumption that the classes appeared in the test data must have appeared in training. This also applies to text learning or text classification. As learning is used increasingly in dynamic open environments where some new/test documents may not belong to any of the training classes, identifying these novel documents during classification presents an important problem. This problem is called open-world classification or open classification. This paper proposes a novel deep learning based approach. It outperforms existing state-of-the-art techniques dramatically.

Charmanteau: Character Embedding Models For Portmanteau Creation
Varun Gangal, Harsh Jhamtani, Graham Neubig, Eduard Hovy, and Eric Nyberg 14:10–14:25

Portmanteaus are a word formation phenomenon where two words combine into a new word. We propose character-level neural sequence-to-sequence (S2S) methods for the task of portmanteau generation that are end-to-end-trainable, language independent, and do not explicitly use additional phonetic information. We propose a noisy-channel-style model, which allows for the incorporation of unsupervised word lists, improving performance over a standard source-to-target model. This model is made possible by an exhaustive candidate generation strategy specifically enabled by the features of the portmanteau task. Experiments find our approach superior to a state-of-the-art FST-based baseline with respect to ground truth accuracy and human evaluation.

Using Automated Metaphor Identification to Aid in Detection and Prediction of First-Episode Schizophrenia
E. Dario Gutierrez, Guillermo Cecchi, Cheryl Corcoran, and Philip Corlett 14:25–14:40

The diagnosis of serious mental health conditions such as schizophrenia is based on the judgment of clinicians whose training takes several years, and cannot be easily formalized into objective measures. However, previous research suggests there are disturbances in aspects of the language use of patients with schizophrenia. Using metaphor-identification and sentiment-analysis algorithms to automatically generate features, we create a classifier, that, with high accuracy, can predict which patients will develop (or currently suffer from) schizophrenia. To our knowledge, this study is the first to demonstrate the utility of automated metaphor identification algorithms for detection or prediction of disease.

Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking
Hannah Rashkin, Eunsol Choi, Jin Yea Jang, Svitlana Volkova, and Yejin Choi 14:40–14:55

We present an analytic study on the language of news media in the context of political fact-checking and fake news detection. We compare the language of real news with that of satire, hoaxes, and propaganda to find linguistic characteristics of untrustworthy text. To probe the feasibility of automatic political fact-checking, we also present a case study based on PolitiFact.com using their factuality judgments on a 6-point scale. Experiments show that while media fact-checking remains to be an open research question, stylistic cues can help determine the truthfulness of text.

Topic-Based Agreement and Disagreement in US Electoral Manifestos
Stefano Menini, Federico Nanni, Simone Paolo Ponzetto, and Sara Tonelli 14:55–15:10

We present a topic-based analysis of agreement and disagreement in political manifestos, which relies on a new method for topic detection based on key concept clustering. Our approach outperforms both standard techniques like LDA and a state-of-the-art graph-based method, and provides promising initial results for this new task in computational social science.

Zipporah: a Fast and Scalable Data Cleaning System for Noisy Web-Crawled Parallel Corpora
Hainan Xu and Philipp Koehn 15:10–15:25

We introduce Zipporah, a fast and scalable data cleaning system. We propose a novel type of bag-of-words translation feature, and train logistic regression models to classify good data and synthetic noisy data in the proposed feature space. The trained model is used to score parallel sentences in the data pool for selection. As shown in experiments, Zipporah selects a high-quality parallel corpus from a large, mixed quality data pool. In particular, for one noisy dataset, Zipporah achieves a 2.1 BLEU score improvement with using 1/5 of the data over using the entire corpus.

Plenary Session. Best Paper

15:50–17:25
Aarhus
Chairs: *Rebecca Hwa , Sebastian Riedel*

Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints
Jieyu Zhao, Tianlu Wang, Mark Yatskar, Vicente Ordonez, and Kai-Wei Chang

Language is increasingly being used to de-fine rich visual recognition problems with supporting image collections sourced from the web. Structured prediction models are used in these tasks to take advantage of correlations between co-occurring labels and visual input but risk inadvertently encoding social biases found in web corpora. In this work, we study data and models associated with multilabel object classification and visual semantic role labeling. We find that (a) datasets for these tasks contain significant gender bias and (b) models trained on these datasets further amplify existing bias. For example, the activity cooking is over 33% more likely to involve females than males in a training set, and a trained model further amplifies the disparity to 68% at test time. We propose to inject corpus-level constraints for calibrating existing structured prediction models and design an algorithm based on Lagrangian relaxation for collective inference. Our method results in almost no performance loss for the underlying recognition task but decreases the magnitude of bias amplification by 47.5% and 40.5% for multilabel classification and visual semantic role labeling, respectively

Natural Language Does Not Emerge 'Naturally' in Multi-Agent Dialog
Satwik Kottur, José Moura, Stefan Lee, and Dhruv Batra

A number of recent works have proposed techniques for end-to-end learning of communication protocols among cooperative multi-agent populations, and have simultaneously found the emergence of grounded human-interpretable language in the protocols developed by the agents, learned without any human supervision! In this paper, using a Task & Talk reference game between two agents as a testbed, we present a sequence of 'negative' results culminating in a 'positive' one – showing that while most agent-invented languages are effective (i.e. achieve near-perfect task rewards), they are decidedly not interpretable or compositional. In essence, we find that natural language does not emerge 'naturally',despite the semblance of ease of natural-language-emergence that one may gather from recent literature. We discuss how it is possible to coax the invented languages to become more and more human-like and compositional by increasing restrictions on how two agents may communicate.

Depression and Self-Harm Risk Assessment in Online Forums
Andrew Yates, Arman Cohan, and Nazli Goharian

Users suffering from mental health conditions often turn to online resources for support, including specialized online support communities or general communities such as Twitter and Reddit. In this work, we present a framework for supporting and studying users in both types of communities. We propose methods for identifying posts in support communities that may indicate a risk of self-harm, and demonstrate that our approach outperforms strong previously proposed methods for identifying such posts. Self-harm is closely related to depression, which makes identifying depressed users on general forums a crucial related task. We introduce a large-scale general forum dataset consisting of users with self-reported depression diagnoses matched with control users. We show how our method can be applied to effectively identify depressed users from their use of language alone. We demonstrate that our method outperforms strong baselines on this general forum dataset.

Bringing Structure into Summaries: Crowdsourcing a Benchmark Corpus of Concept Maps
Tobias Falke and Iryna Gurevych

Concept maps can be used to concisely represent important information and bring structure into large document collections. Therefore, we study a variant of multi-document summarization that produces summaries in the form of concept maps. However, suitable evaluation datasets for this task are currently missing. To close this gap, we present a newly created corpus of concept maps that summarize heterogeneous collections of web documents on educational topics. It was created using a novel crowdsourcing approach that allows us to efficiently determine important elements in large document collections. We release the corpus along with a baseline system and proposed evaluation protocol to enable further research on this variant of summarization.

Honorable Mentions

EMNLP 2017 has seen many strong submissions. In addition to the four papers that received the Best Paper Awards, the award committee would also like to recognize the following six papers with an "Outstanding Paper" distinction.

- Haim Dubossarsky, Daphna Weinshall and Eitan Grossman. *Outta control: Laws of semantic change and inherent biases in word representation models*
- Alexis Conneau, Douwe Kiela, Holger Schwenk, Loïc Barrault and Antoine Bordes. *Supervised Learning of Universal Sentence Representations from Natural Language Inference Data*
- David Mimno. *The strange geometry of skip-gram with negative sampling*
- Stefano Menini, Federico Nanni, Simone Paolo Ponzetto and Sara Tonelli. *Topic-Based Agreement and Disagreement in US Electoral Manifestos*
- Robin Jia and Percy Liang. *Adversarial Examples for Evaluating Reading Comprehension Systems*
- Yuhao Zhang, Victor Zhong, Danqi Chen, Gabor Angeli and Christopher D. Manning. *Position-aware Attention and Supervised Data Improve Slot Filling*

8

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 the ACL. 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 all the members, as well as participants at our events and in our programs.

Harassment and hostile behavior are unwelcome at any ACL conference, associated event, or in ACL-affiliated on-line discussions. This includes: speech or behavior that intimidates, creates discomfort, or interferes with a person's participation or opportunity for participation in a conference or an event. We aim for ACL-related activities 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, appearance, 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. The policy is not intended to inhibit challenging scientific debate, but rather to promote it through ensuring that all are welcome to participate in shared spirit of scientific inquiry.

It is the responsibility of the community as a whole to promote an inclusive and positive environment for our scholarly activities. In addition, anyone who experiences harassment or hostile behavior may contact any current member of the ACL Executive Committee or contact Priscilla Rasmussen, who is usually available at the registration desk during ACL conferences. Members of the executive committee will be instructed to keep any such contact in strict confidence, and those who approach the committee will be consulted before any actions are taken.

Approved by ACL Executive Committee in 2016.

The policy is also available from ACL's main page.

Index

9 Local Guide

This guide was written by Maria Barrett, Joachim Bingel, Mareike Hartmann, Dirk Hovy.
For the most up-to-date version, please visit `http://www.emnlp2017.net`

General

Velkommen til København!

For getting a sense of the **city's geography**, it is helpful to realize that there are three major boroughs surrounding the Inner City in the West, North, and East, which are aptly called *Vesterbro* (where the venue is located), *Nørrebro*, and *Østerbro*. Crammed between Vesterbro and Nørrebro lies *Frederiksberg*. In the north, the Inner City is walled off by a stretch of artificial lakes, *Søerne*, and in the south, on the other side of the canal, there's the island of *Amager*.

To **get around** the city, it's best to buy a multi-day pass or rent bikes at a local bike shop. There are also **city bikes** (*bycykler*) at many locations throughout the city. You can get access at the bike stations or online at `https://my.bycyklen.dk/en/account/register`. When biking, clearly indicate where you are going. Do not stop without indicating so (by raising a hand as if to greet someone). Do not swerve and change lanes. Any of these things will bring out the inner Viking in otherwise mild-mannered Danish cyclists.

Public transport is excellent and gets you everywhere. You certainly don't need a taxi between almost anywhere in the city and the airport. But note that there are three zones to get from the airport to central Copenhagen (and regular ticket controls!). Otherwise two zones will get you around central Copenhagen. Find connections between addresses in Denmark at `http://journeyplanner.dk`. If you stay longer and travel more than 10 times, you might want to get a **Rejsekort** (reloadable travel card valid throughout all of Denmark on all modes of transportation), which you can buy at vending machines placed at every metro station. Note that the card itself costs 80 DKK (non-refundable), and you'll need to top it up before your first ride.

You can pay with **credit or debit card** pretty much anywhere, although some places (usually smaller kiosks) require a Danish card or cash.

The country code is +45.

Sightseeing

Kødbyen (literally, the "meat town") is a lively place, and you can spend the entire conference within the walls of the old meatpacking district, trying new restaurants, listening to upcoming bands, and drinking hipster coffee with the locals. However, *DGI Byen* – the part of Kødbyen in which Øksnehallen and Cph Conference are located – contains a number of other facilities. This includes indoor swimming and gym, that are free for participants. Check out `http://www.dgi-byen.com/` for more information. The following sightseeing suggestions all take you out of Kødbyen.

Carlsberg Glyptotek

Nice collection of statues and paintings. The courtyard has a cafe where you can sit among palm trees. Free entrance on Tuesdays.

`http://www.glyptoteket.com`

Dantes Plads 7	Tue, Wed, Fri, Sat, Sun	11:00–18:00
1556 Copenhagen K	Thu	11:00–22:00

Christiania

Mostly known for weed and dead hippie dreams. Actually has nice jazz concerts and other events (often in 'Operaen', not to be confused with the big opera house). The old parts along the ramparts are full of nice, self-made houses. Nice vegetarian restaurant, Morgenstedet.

`http://www.visitcopenhagen.com/copenhagen/culture/alternative-christiania`

Main entrance is on Prinsessegade, but there are several other ways into the area.

Danish Design Museum

Next to Kastellet in an old hospital. A series of several rooms arranged around a courtyard, each presenting one decade or artist of Danish design (lamps, chairs, etc.). Nice café, and a museum shop with many of the things you just saw.

`http://www.designmuseum.dk/en`

Bredgade 68	Tue, Fri, Thu, Sat, Sun	10:00–18:00
1260 Copenhagen K	Wed	10:00–21:00

Den Sorte Diamant (*The Black Diamond*)

The Royal Library. Modern addition to the old library. Great architecture. Changing exhibitions in the basement, a free exhibition of rare books and old manuscripts in a pop-art decorated room upstairs. Also features changing photography exhibitions.

`http://www.kb.dk/en/`

Søren Kierkegaards Pl. 1	Mon–Fri	08:00–21:00
1221 Copenhagen K	Sat	09:00–19:00

Get a GoBoat

Get an electric-powered boat for up to eight people and set to sea (well, don't actually set to sea: the canal). You can explore Copenhagen's waterways on your own while enjoying some drinks and snacks that you bring yourself or pick up, e.g. at Papirøen (see 'Culinary Highlights in the City').

`http://goboat.dk/en/`

Islands Brygge 10
2300 Copenhagen S

Harbour bus

Cheapskate boat sightseeing. Pay a regular bus ticket or show your multi-day pass and get a sea view of Copenhagen. Stops e.g. by Den Sorte Diamant, Nyhavn, The Opera House and Langelinje.

`https://www.google.dk/maps/place/Havnebussen/`

Kastellet

The old castle at the northern end of the city. Shaped like a star, and fun to walk on. Next to the Little Mermaid, which you can happily skip, but if you must, you can see it from here. Was built to defend the city, but captured before it was finished.

National Museum

One of the best Viking collections in the world.

`http://en.natmus.dk/museums/the-national-museum-of-denmark/`

Ny Vestergade 10 — Tue–Sun 10:00–17:00
1471 Copenhagen K

Rundetårn

Round tower with a spiraling pathway inside, built so that the king could ride up on a horse to the observatory. Nice view over the city, and you can run down the spiral and shout "wheeee!"

`http://www.rundetaarn.dk/en/`

Købmagergade 52A — All days 10:00–20:00
1150 Copenhagen K

Torvehallerne

Two modern market halls, filled with food vendors and food-related items. Some overpriced stuff, but also a lot of specialized shops with things that are otherwise hard to find. The smørrebrød shop is becoming a new in-spot.

`http://torvehallernekbh.dk`

Frederiksborggade 21 — Mon–Wed 10:00–19:00 Sat 10:00–18:00
1360 Copenhagen K — Thu–Fri 10:00–20:00 Sun 11:00–17:00

The Little Mermaid

Yeah… Don't be too disappointed.

Langelinie
2100 Copenhagen Ø

Statens Museum for Kunst

Decent art museum. Has a bar on the first Friday of the month.

`http://www.smk.dk/en/`

Sølvgade 48-50	Tue–Sun	11:00–17:00
1307 Copenhagen K	Wed	11:00–20:00

Ørestad

If you like new architecture, take the metro and visit Ørestad. It is a new area of Copenhagen with award-winning houses. There is a mall directly in front of Ørestad station (Field's), or take the metro out to Vestamager station and walk up the 8tallet house.

`http://www.visitcopenhagen.com/copenhagen/architecture/architectural-orestad`

Jægersborggade

This street in Nørrebro is a hipster's paradise. Very close to Assistens Kirkegård, it is home to a number of cafés/bars, galleries, craft and (second-hand) clothing shops, as well as combinations thereof (Sneakers&Coffee, Beer&Vinyl, ...). Great for finding gifts to take home.

`http://jaegersborggade.com/wpAB/en/shops/`

Shops

B&W Hallerne. *Second hand furniture in a giant, factory hall, by the water, in industrial Copenhagen. Only open every other weekend. Bring cash.* Refshalevej 171, 1432 Copenhagen. `https://www.facebook.com/BW-LOPPEMARKED-171235476283625/`

Faraos Cigarer. *Cartoon shop – or three, actually, but all next-door – close to Rundetårn.* Skindergade 27, 1159 Copenhagen K. `https://www.faraos.dk`

Sort kaffe & Vinyl. *Hipster record shop with good coffee.* Skydebanegade 4 1709 Copenhagen V. `https://www.facebook.com/sortkaffeogvinyl/`

København K. *Second-hand clothing in a street packed with fashion stores.* Studiestræde 30, 1455 Copenhagen K. `http://koebenhavnk.com`

Dance and Music Venues

Global CPH. *World music.* Nørre Allé 7, 2200 Copenhagen N. `http://globalcph.dk/english/`

Danshallerne. *Dance and theatre.* Bohrsgade 19, 1799 Copenhagen V. `http://www.dansehallerne.dk/en/`

La Fontaine. *Intimate, 100-capacity veteran of the Scandinavian jazz scene staging nightly jam sessions.* Kompagnistræde 11, 1208 Copenhagen K. `http://lafontaine.dk`

Montmartre. *Classic jazz joint in central Copenhagen.* Store Regnegade 19A, 1110 Copenhagen K. http://www.jazzhusmontmartre.dk

Mojo. *Blues, jazz & folk from Scandinavian & international artists in a compact club open until 5am.* Løngangstræde 21C, 1468 Copenhagen K. https://mojo.dk

Pumpehuset. *Rock, pop, hip hop, world music – lots of different music styles, but generally nice, national and international bookings.* Studiestræde 52, 1554 Copenhagen V. http://pumpehuset.dk/

Royal Theatre. *See ballet at the old stage, a play in the new theatre house by the sea or opera in the Opera House.* https://kglteater.dk/en/

Vega. *Rock and pop scene not too far from the venue.* Enghavevej 40, 1674 Copenhagen V. http://vega.dk

Events

Moreover, if your conference schedule permits, here's a list of events happening in Copenhagen and the wider region, during EMNLP:

Cirque du Soleil. Malmö Arena, September 6–10. Hyllie Stationstorg 2, 215 32 Malmö, Sweden. https://www.cirquedusoleil.com/sweden/malmo/shows

Copenhagen World Music. Runs September 6–10 at different venues in Copenhagen. http://cphworld.dk

Luisi leads Danish composer Carl Nielsen's 5th Symphony in Koncerthuset. September 7. Koncerthuset, Emil Holms Kanal 20, 2300 Copenhagen S. http://drkoncerthuset.dk/

Day-Trips

Arken

Modern art museum south of the city, in an unassuming suburb. Nice changing exhibitions and a modern collection.

http://uk.arken.dk

Skovvej 100
2635 Ishøj

Louisiana

Modern art museum on the coast north of the city. The buildings are integrated into a park overlooking the cliffs towards Sweden. Great exhibits and a world-renowned collection. Sneak in picnic stuff and eat on the lawns.

https://en.louisiana.dk

Gl Strandvej 13
3050 Humlebæk

Roskilde Viking Ship Museum

30 min by train from the main station, in a lovely small town lies this museum, which contains five well-preserved Viking ships that were sunk there to form a barrier. See the 1:1 viking ship reconstructions and try to sail one.

`http://www.vikingeskibsmuseet.dk/en/`

Vindeboder 12
4000 Roskilde

Other Things to Do

If you're less into classical sightseeing but want to live a day as a Copenhagener would (big claim, we know…), the following suggestions might be interesting. Obviously, the best way to get around is by bike.

Dance the Night Away

If you're into electronic music, you'll probably enjoy Copenhagen's old Meat Packing District (*Kødbyen*), where all the following bars are located. *Jolene* and *Bakken* are small and sweaty, and true to their origin and location. They look a bit run-down (the area used to be all slaughterhouses and meat auction halls). There's also *KB18* with more deep house/techno, but don't go there before 1 am. *Mesteren & Lærlingen* plays hip hop/soul/reggae.

Parks

Copenhagen has a number of very nice parks that invite you to hang out or do sports. The biggest, *Fælledparken*, is right next to the Department of Computer Science and has a little lake, otherwise it's a big green meadow with lots of runners and football players. A hidden gem with a super beautiful (albeit artificially created) lake, or rather trench, is *Østre Anlæg*, right behind the National Art Gallery. You can have a barbeque there if you bring your own coal, stationary grills are provided. *Assistens Kirkegård* in Nørrebro is a very beautiful graveyard. People from other places might find this strange, but it's a favourite pastime among Copenhageners to hang out here. *Nørrebroparken* is wonderful in springtime, go there to join young Copenhageners for a beer or a frisbee game (disclaimer: don't literally try to join them, they are Danes. They will panic). Just outside of Copenhagen near Gentofte station, is *Bernstorff's Park* which – aside from usual park stuff and a castle – holds Pometet which is a very broad range of Danish fruit trees as well as a smaller selection of exotic fruit trees. Originally for the royal family, but nowadays free for visitors to sample. During weekends in spring, summer and early autumn you can have afternoon tea in Queen Louise' teahouse or in the her rose garden.

Eating & Drinking

Unless you come from Norway, everything will be more expensive than in your home country. It's best not to convert the prices and just take them as-is. A coffee/latte costs 20–50 DKK, a beer almost everywhere 50 DKK. A main course in a restaurant is typically 150–250 DKK.

The following sections list a selection of restaurants along with the price range and their distance from the conference venue.

Price range indicator:

$ Main course < 150 DKK
$$ Main course 150–250 DKK
$$$ Main course > 250 DKK

Restaurants Close to the Conference Venue

The following restaurants are in very close proximity to the conference venue, most of them in *Kødbyen*, the old meat packing district, which is now a center of Copenhagen nightlife and home to numerous restaurants offering foods from around the globe. The following are all great picks, but there are plenty more for you to discover.

Fiskebaren **0.4 km** **$$$**

Allegedly one of the 10 best fish restaurants in Europe. Very nice dishes, but expensive. Go for the starters and medium dishes and share. Make sure to check out the column-shaped aquarium.

`http://fiskebaren.dk`

Flæsketorvet 100	Mon–Thu	17:30–00:00	Sat	11:30–02:00
1711 Copenhagen V	Fri	17:30–02:00	Sun	11:30–00:00

Fleisch **0.3 km** **$$**

Restaurant in Kødbyen with its own butcher counter. Have some meat-based open-face sandwiches there, or grab a few of their beer sausages to go and enjoy them outside.

`http://www.fleisch.dk`

Slagterboderne 7	Tue–Thu	11:30–00:00	Sun	11:30–00:00
1716 Copenhagen V	Fri–Sat	11:30–01:00		

Bio Mio Organic Bistro **0.2 km** **$$**

Organic hearty bistro. Vegan and vegetarian options.

	Mon–Wed	12:00–16:00 & 17:00–22:00
Halmtorvet 19	Thu	12:00–16:00 & 17:00–22:30
1700 Copenhagen V	Fri–Sat	12:00–16:00 & 17:30–22:30
	Sun	12:00–16:00 & 17:00–21:00

Hija de Sanchez **0.3 km** **$**

Ex-Noma chef decided to leave the big business and make everyday tacos. Get 3 for 100DKK.

`http://www.hijadesanchez.dk`

Slagterboderne 8	Mon–Fri	17:30–00:00
1716 Copenhagen V	Sun	17:30–22:00

Isted Grill **0.7 km** **$**

Not a restaurant, but a hole-in-the-wall late-night snack food place. Get the flæskesteg sandwich (grilled pork roast with pickles and red cabbage) after a night of drinking. Cheap and tasty.

Istedgade 92	Sun–Thu	12:00–00:00
1650 Copenhagen V	Fri–Sat	12:00–02.00

Madklubben **0.7 km** **$**

Several locations, closest on Vesterbrogade. Nice "home-made" food, reasonable prices. Very good for groups (with a reservation). Vegetarian options.

`http://madklubben.dk/en/`

Vesterbrogade 62	Mon–Thu	17:30–00:00
1620 Copenhagen V	Fri–Sat	17:00–00:00

Magasasa Dim Sum & Cocktails **0.5 km** **$**

The name says it all. Raw interior. Reasonably priced.

`http://magasasa.dk/dim-sum-cocktails/?lang=en`

Flæsketorvet 54–56	Mon–Thu	11:00–23:00
1711 Copenhagen V	Fri–Sat	11:00–00:00

Mother **0.4 km** **$**

Great pizza!

`http://mother.dk`

Høkerboderne 9	All days	11:00–01:00
1712 Copenhagen V		

Nose2Tail **0.6 km** **$$**

As cozy as a former slaughterhouse basement can get (the one in the Kødbyen basement is the best of their locations). Serves daily changing meat, fish, and innard dishes. Freshly made cracklings (Danish delicacy made from pork skin), served with bacon mayonnaise – because fat! Chase with a Fernet Branca. Very good beer and friendly staff with awesome leather aprons.

`http://nose2tail.dk/`

Kødboderne 9	Tue–Thu	18:00–00:00
1711 Copenhagen	Fri–Sat	18:00–01:00

PatePate **0.2 km** **$$**

Small plates, international cuisine. Great for sharing. Medium price range. Also nice to start your evening in Kødbyen. Vegetarian options.

`http://www.patepate.dk/`

Slagterboderne 1	Mon–Wed	09:00–00:00	Fri	09:00–01:00
1716 Copenhagen V	Thu	09:00–01:00	Sat	11:30–01:00

Restaurant Cofoco **0.3 km** **$$**

Part of the Cooking for Copenhagen group. Small delicious plates. Slightly upscale, but reasonable.

`http://cofoco.dk/en/`

Abel Cathrines Gade 7 1654 Copenhagen V	Mon–Sat 18:00–00:00

Tommi's Burger Joint **0.4 km** **$**

Tasty minimalistic burgers! Part of a small international chain originating from Iceland.

`https://www.burgerjoint.dk/`

Høkerboderne 21-23 1712 København V	Thu–Sun 11:00–22:00 Mon–Wed 11:00–21:00

WEDOFOOD **0.2 km** **$**

Homemade salads. Vegetarian options.

`http://wedofood.dk`

Halmtorvet 21 1700 Copenhagen V	Mon-Sat 10:00–21:00 Sun 10:00–20:00

Warpigs **0.3 km** **$$**

One word: pork. Great craft beer.

`http://warpigs.dk`

Flæsketorvet 25 1711 Copenhagen V	Mon–Thu 11:30–00:00 Fri–Sat 11:00–02:00

Culinary Highlights in the City

Bror **1.3 km** **$$$**

High-end dining, comes with reasonably priced wine pairings. More expensive, but worth it.

`http://www.restaurantbror.dk`

Sankt Peders Stræde 24A 1453 Copenhagen K	Wed–Sun 17:30–00:00

Geist **2.3 km** **$$**

Super-minimalist nordic cooking. The menu describes exactly what you get ("Carrots, braised in orange juice, with ginger"), but whatever it is, it's cooked to perfection! You order several small plates, each reasonably priced, but it adds up. You can watch the busy, but eerily quiet kitchen, lit only by candles. Fancy cocktail bar, too, albeit a bit short on classics.

`http://restaurantgeist.dk`

Kongens Nytorv 8 1050 Copenhagen K	All days	12:00–15:00 and 17:30–01:00

Gran Torino **2.3 km** **$**

Part of the Madklubben group. Set meals (from 200 DKK) includes pasta, pizza and extra nice Tiramisu.

`http://madklubben.dk/gran-torino/`

Sortedam Dossering 5 2200 Copenhagen N	Mon–Fri Sun	17:30–00:00 17:30–22:00

Höst **1.6 km** **$$$**

Get a fixed price menu with lots of interesting culinary and visual effects (clams on burning juniper bushes). Pricey, but worth the money. Throw in an extra few hundred for the wine pairing, and you won't be disappointed.

`http://hostvakst.dk/host/restaurant/?lang=en`

Nørre Farimagsgade 41 1364 Copenhagen K	All days	17:30–00:00

Morgenstedet **3.3 km** **$**

Vegetarian Restaurant in the heart of Christiania. Has an improvised canteen feel to it, but very tasty food.

Fabriksområdet 134 1440 Copenhagen K	Tue–Sun	12:00–21:00

Papirøen **3.3 km** **$-$$**

A collection of street food vendors in one old factory building, ranging from duck-fat fries to Asian noodle salads. And plenty of drinks. If the weather is nice, you can sit in deck chairs and watch the boats on the canal. Cheap to mid-range depending on the vendor.

`http://copenhagenstreetfood.dk/en/`

Trangravsvej 14, hal 7/8 1436 Copenhagen K	Mon–Thu Fri–Sun	12:00–21:00 12:00–21:00

Ramen to Biiru **1.4 km** **$**

Japanese ramen meets Danish craft beer.

`http://ramentobiiru.dk/vesterbro/`

Enghavevej 58 1674 Copenhagen V	Mon–Thu Fri-Sat	12:00–22:00 12:00–23:00	Sun	12:00–21:00

SimpleRAW **1.7 km** **$**

Rawfood. Vegan and vegetarian.

https://www.simpleraw.dk

Gråbrødre Torv 9	Mon-Sat	10:00–22:00
1154 Copenhagen K	Sun	10:00–20:00

Cafés and Bars

Bang og Jensen **0.9 km**

At the far end of Istedgade, both a cafe and a bar, and extremely hyggelig.

http://www.bangogjensen.dk

Istedgade 130	Mon-Fri	07:30-02:00 Sun 10:00-00:00
1650 Copenhagen V	Sat	10:00-02:00

Bastard **1.3 km**

Board game café.

http://bastardcafe.dk

Rådhusstræde 13	Mon–Thu, Sun	12:00–00:00
1466 Copenhagen K	Fri–Sat	12:00–02:00

Kaffe **0.7 km**

On Istedgade, at the intersection with Skydebanegade, and easy to miss. Full of wooden trinkets and good coffee. Very small!

http://kaffeistedgade.dk

Istedgade 90	Mon–Fri	08:00–22:00
1650 Copenhagen V	Sat–Sun	09:00–22:00

Library Bar **0.6 km**

Next to the train station, the bar of the Plaza Hotel. Plush leather sofas, dark wood panels, and on some nights: live piano and songs. Mixed crowd, go in a suit or with hiking clothes.

https://ligula.se/en/the-library-bar/

Bernstorffsgade 4	Mon–Thu	16:00–00:00
1577 Copenhagen V	Fri–Sat	16:00–01:00

Mikkeller **0.4 km**

Slightly overhyped micro-brewery. Some good beers, some misses. If you ever wondered what chocolate-liquorice blueberry porter tastes like: here you might find out. Excellent sausages to go with the beer.

http://mikkeller.dk/location/mikkeller-bar-viktoriagade-copenhagen/

Viktoriagade 8	Sun–Wed	13:00–01:00 Sat 12:00–02:00
1655 Copenhagen V	Thu–Fri	13:00–02:00

Paludan **2.3 km**

Sit in an antique book shop and sip a beer. Or a coffee. Very decent snack food (try the charcuterie board). Also a good place to work.

`https://www.paludan-cafe.dk/home-eng`

Fiolstræde 10	Mon–Fri	09:00–22:00	Sun	10:00–22:00
1171 Copenhagen K	Sat	10:00–22:00		

Risteriet **0.2 km**

Coffee place close to Kødbyen.

`http://www.risteriet.dk/risteriet-halmtorvet/`

Helgolandsgade 21	Mon–Fri	07:30–18:00	Sun	09:00–18:00
1700 Copenhagen V	Sat	09:00–18:00		

Sort kaffe og vinyl **0.6 km**

Another small coffee place, around the corner from Kaffe. You can also buy old vinyl disks while sipping your latte.

`https://facebook.com/sortkaffeogvinyl/`

Skydebanegade 4	Mon–Fri	08:00–19:00	Sun	09:00–18:00
1709 Copenhagen	Sat	09:00–19:00		

Health

Pharmacies are few and far between. If you need one, there is a 24h pharmacy next to the train station on Vesterbrogade 6, 1620 Copenhagen V, 0.6 km from the venue.

For **24h medical advice**, call 1813. They can also help you see a doctor outside regular opening hours.

In **emergency situations**, call 112.

Siteimprove

One of Denmark's Top 10 Best IT Workplaces*

In one of the world's happiest countries

With some of the world's best restaurants

**And as Carlsberg would say –
probably the best colleagues in the world!**

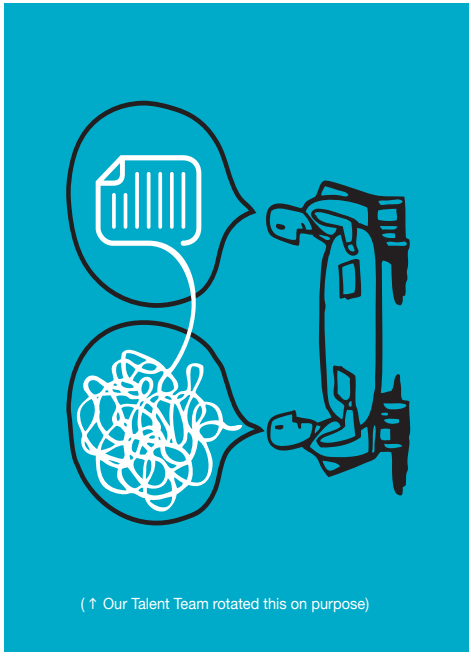
We have a small but dedicated department committed to being the best in Natural Language Processing – are you our next team member?

*2015, 2016 Great Place to Work® Institute

Stop by our booth to find out more!

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.



(↑ Our Talent Team rotated this on purpose)

textkernel

Innovation, Artificial Intelligence and Machine Learning for matching jobs and people

- Information extraction and retrieval
- Recommendation systems
- Deep Learning and Big Data
- Week long hackathon
- Influence on products

We're hiring!

Visit textkernel.careers or email us at jobs@textkernel.com

Maluuba

A Microsoft company

Building Literate Machines

Maluuba is a Microsoft company teaching machines to think, reason and communicate.

We demonstrate progress by focusing on challenging problems, driving new techniques in reinforcement learning, with a focus on machine comprehension and conversational interfaces. Our vision is to solve 'Artificial General Intelligence' by creating literate machines that can think, reason and communicate like humans.

At Maluuba Microsoft Canada, people are the source of our energy. The people at Maluuba are creative, from all types of backgrounds, bringing passion and new ideas, meeting challenges, and realizing their potential.

We're hiring talented and motivated people to join our team and help us achieve our vision.

- Research Scientists
- Applied Researchers
- Research Engineers
- Research Internships

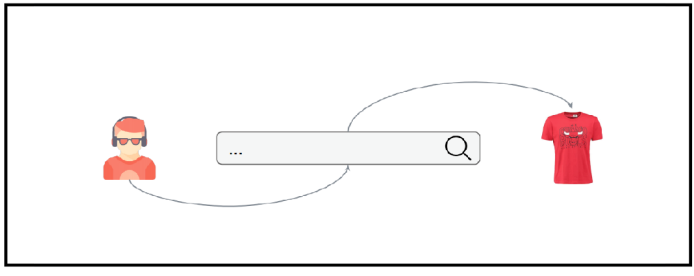
@MaluubaInc

Maluuba.com

contact@maluuba.com

DATA IS ONE OF ZALANDO'S BIGGEST AND MOST IMPORTANT ASSETS.

Our teams work to reshape the way people think about online fashion.



@ZalandoTech
research@zalando.de
research.zalando.com

zalando research

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

The EMNLP organizers gratefully acknowledge the support from the following sponsors.

Platinum







Gold









Silver










Bronze


Snap Inc.