



# CONFERENCE HANDBOOK

SEPTEMBER 17-21

**EMNLP** 2015  
CONFERENCE ON EMPIRICAL METHODS  
IN NATURAL LANGUAGE PROCESSING

[contact@emnlp2015.org](mailto:contact@emnlp2015.org)

[WWW.EMNLP2015.ORG](http://WWW.EMNLP2015.ORG)

PLATINUM SPONSORS

Bloomberg  Baidu

GOLD SPONSORS

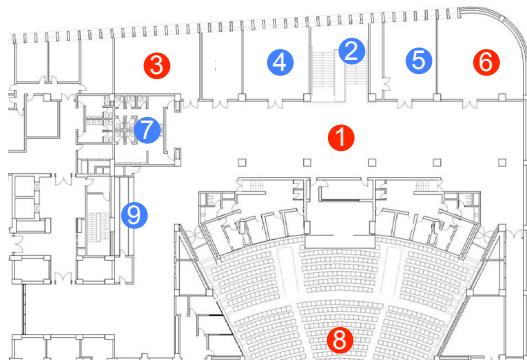
 Google  

SILVER SPONSORS

 RECRUIT  
Recruit Technologies Co.,Ltd.  NUANCE  amazon  AltaISI  VoiceBox  DigitalGenius

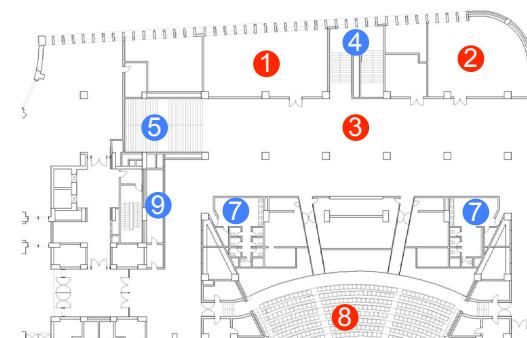
## Level 1

1. Higher level foyer
2. Lower level foyer access
3. Room 3
4. Room 4
5. Room 5
6. Room 6
7. WC
8. Main Auditorium
9. Bar



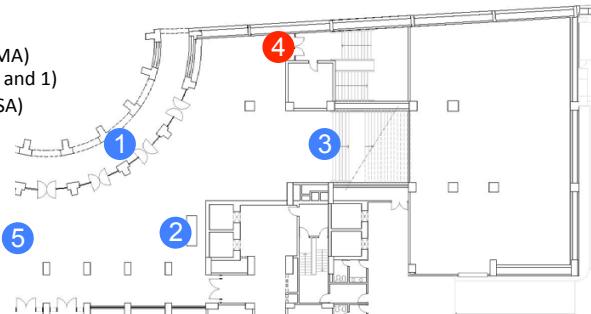
## Level -1

1. Room 1
2. Room 2
3. Lower level foyer
4. Higher level foyer access
5. Main entrance
6. Cloakroom
7. WC
8. Main Auditorium
9. Bar



## Level -2: Main Entrance

1. Main entrance
2. Reception desk
3. Access to Main Auditorium (MA) and Meeting rooms (levels -1 and 1)
4. Access to Small Auditorium (SA) (level -3)
5. Cafeteria



Cover design by Sarah Almeida

Thanks to Matt Post for his helpful notes about creating a handbook

Handbook assembled by Fernando Batista

## Schedule at a Glance

		Sep. 18 (Friday)	Sep. 19 (Saturday)	Sep. 20 (Sunday)	Sep. 21 (Monday)
Tutorials 9:00 - 12:30	Workshops 9:00 - 12:30			Registration 07:30 - 18:00	
			Opening & Invited Talk 08:40 - 10:00	Invited Talk 9:00 - 10:00	Morning Coffee 9:00 - 10:00
				Coffee break 10:00 - 10:30	Best Paper Awards 9:00 - 10:00
			Parallel Session 1 10:00 - 12:10	Parallel Session 4 10:00 - 12:10	Parallel Session 7 10:00 - 12:10
			Lunch break 12:10 - 13:30	Lunch break 12:50 - 13:30	Lunch break 12:10 - 13:30
				Parallel Session 2 13:30 - 15:10	Parallel Session 5 13:30 - 15:10
				Coffee break 15:10 - 15:40	Parallel Session 8 13:30 - 15:10
					Closing Remarks 17:30 - 17:50
					Conference Dinner 19:00 - 23:00
					Farewell drink 18:30 - 20:00
					Welcome Reception 18:30 - 20:00

## Schedule Overview

---

### **Thursday, September 17**

---

09:00–17:30 **Tutorials**

09:00–19:30 **Workshops**

### **Friday, September 18**

---

09:00–17:30 **Tutorials**

09:00–18:00 **Workshops**

18:30–20:00 **Welcome Reception**

Culturgest

### **Saturday, September 19**

---

07:30–18:00 **Registration**

Culturgest

08:00–08:40 **Morning Coffee**

Foyers

08:40–10:00 **Session P1: Plenary Session**

Main Auditorium

08:40–09:00 *Opening Remarks and Introductory Speeches (General Chair, Program Co-Chairs and Local Co-Chairs)*

09:00–10:00 *Invited Talk: Deep Learning of Semantic Representations (Yoshua Bengio)*

10:00–10:30 **Coffee break**

Foyers

10:30–12:10 **Session 1**

Semantics (Long + TACL Papers)  Main Auditorium	Machine Translation (Long + TACL Papers)  Small Auditorium	NLP for the Web and Social Media, including Computational Social Science (Long Papers)  Room 2	Long Paper Posters  Short Paper Posters  Lower Level Foyer
---	--	--	--

12:10–13:30 **Lunch**

13:30–15:10 **Session 2**

Statistical Models and Machine Learning Methods (Long + TACL Papers)  Main Auditorium	Tagging, Chunking and Parsing (Long +TACL Papers)  Small Auditorium	Summarization (Long Papers)  Room 2	Long Paper Posters  Short Paper Posters  Lower Level Foyer
---	---	---	--

15:10–15:40 **Coffee break**

Foyers

15:40–17:20 **Session 3**

Sentiment Analysis and Opinion Mining (Long Papers)  Main Auditorium	Semantics (Long +TACL Papers)  Small Auditorium	Information Retrieval and Question Answering (Long Papers)  Room 2	Long + TACL Paper Posters  Short Paper Posters  Lower Level Foyer
--	---	--	---

### **Sunday, September 20**

---

07:30–18:00 **Registration**

Culturgest

08:00–09:00 **Morning Coffee**

Foyers

09:00–10:00 **Session P2: Plenary Session**

Main Auditorium

*Invited Talk: Measuring How Elected Officials and Constituents Communicate (Justin Grimmer)*

10:00–10:30 **Coffee break**

Foyers

10:30–12:10 **Session 4**

Information Extraction (Long Papers)  Main Auditorium	Statistical Models and Machine Learning Methods (Long Papers)  Small Auditorium	Discourse (Long +TACL Papers)  Room 2	Long Paper Posters  Short Paper Posters  Lower Level Foyer
---	---	---	--

12:10–12:50 **Lunch**

12:50 – 13:30 <b>Session P3: SIGDAT business meeting</b>	Main Auditorium		
13:30 – 15:10 <b>Session 5</b>			
Text Mining and NLP Applications (Long + TACL Papers) <i>Main Auditorium</i>	Semantics (Long +TACL Papers) <i>Small Auditorium</i>	Phonology and Word Segmentation (Long Papers) <i>Room 2</i>	Long Paper Posters Short Paper Posters <i>Lower Level Foyer</i>
15:10 – 15:40 <b>Coffee break</b>	Foyers		
15:40 – 17:20 <b>Session 6</b>			
Machine Translation (Long Papers) <i>Main Auditorium</i>	Sentiment Analysis and Opinion Mining / Tagging, Chunking and Parsing (Long Papers) <i>Small Auditorium</i>	Language and Vision / Information Extraction (Long Papers) <i>Room 2</i>	Long Paper Posters Short Paper Posters <i>Lower Level Foyer</i>
19:00 – 23:00 <b>Conference Dinner</b>	<i>Pateo Alfacinha</i>		
<b>Monday, September 21</b>			
07:30 – 18:00 <b>Registration</b>	Culturgest		
08:00 – 09:00 <b>Morning Coffee</b>	Foyers		
09:00 – 10:00 <b>Session P4: Plenary Session</b>	Main Auditorium		
09:00 – 09:05 Best Paper Awards ( <i>Chris Callison-Burch and Jian Su</i> )			
09:05 – 09:30 Broad-coverage CCG Semantic Parsing with AMR <i>Yoav Artzi, Kenton Lee, and Luke Zettlemoyer</i>	(p. 119)		
09:30 – 09:55 Semantically Conditioned LSTM-based Natural Language Generation for Spoken Dialogue Systems <i>Tsung-Hsien Wen, Milica Gasic, Nikola Mrkšić, Pei-Hao Su, David Vandyke, and Steve Young</i>	(p. 119)		
09:55 – 10:05 A large annotated corpus for learning natural language inference <i>Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning</i>	(p. 119)		
10:05 – 10:30 <b>Coffee break</b>	Foyers		
10:30 – 12:10 <b>Session 7</b>			
Semantics (Long +TACL Papers) <i>Main Auditorium</i>	Information Extraction (Long Papers) <i>Small Auditorium</i>	Computational Psycholinguistics / Machine Translation (Long Papers) <i>Room 2</i>	Long +TACL Paper Posters Short Paper Posters <i>Lower Level Foyer</i>
12:10 – 13:30 <b>Lunch</b>			
13:30 – 15:15 <b>Session 8</b>			
Fun and Quirky Topics (Short Papers) <i>Main Auditorium</i>	Semantics (Short Papers) <i>Small Auditorium</i>	Statistical Models and Machine Learning Methods / Machine Translation (Short Papers) <i>Room 2</i>	Long Paper Posters Short Paper Posters <i>Lower Level Foyer</i>
15:15 – 15:40 <b>Coffee break</b>	Foyers		
15:40 – 17:20 <b>Session 9</b>			
Statistical Models and Machine Learning Methods (Long + TACL Papers) <i>Main Auditorium</i>	Text Mining and NLP Applications (Long Papers) <i>Small Auditorium</i>	Spoken Language Processing and Language Modeling (Long Papers) <i>Room 2</i>	Long Paper Posters Short Paper Posters <i>Lower Level Foyer</i>
17:30 – 17:50 <b>Session P5: Closing Remarks</b>	Main Auditorium		
18:30 – 20:00 <b>Farewell Drink</b>	Culturgest		

## **WIFI access**

---

**Username:** EMNLP2015

**Password:** Lisboa2015

## **Links**

---

Web site: [www.emnlp2015.org](http://www.emnlp2015.org)

Contact email: [contact@emnlp2015.org](mailto:contact@emnlp2015.org)

Find us on facebook: <https://www.facebook.com/emnlp2015>

Follow us on twitter: <https://twitter.com/emnlp2015>

Discussion group: <https://groups.google.com/forum/?hl=en#!forum/emnlp-2015>

Conference4me app: <http://conference4me.psnc.pl/download/>

## Contents

<b>1 Conference Information</b>	<b>1</b>
Preface by the General Chair . . . . .	1
Preface by the Program Committee Co-Chairs . . . . .	4
Conference Committee . . . . .	6
Social events . . . . .	9
<b>2 Tutorials: Thursday – Friday, September, 17–18</b>	<b>11</b>
<b>T1:</b> Semantic Similarity Frontiers: From Concepts to Documents . . . . .	12
<b>T2:</b> Personality Research for NLP . . . . .	14
<b>T3:</b> Transparent Machine Learning for Information Extraction: State-of-the-Art and the Future . . . . .	15
<b>T4:</b> Knowledge Acquisition for Web Search . . . . .	16
<b>T5:</b> Learning Semantic Relations from Text . . . . .	17
<b>T6:</b> Applications of Social Media Text Analysis . . . . .	19
<b>T7:</b> Robust Semantic Analysis of Multiword Expressions with FrameNet . . . . .	20
<b>T8:</b> Computational Analysis of Affect and Emotion in Language . . . . .	21
<b>3 Workshops</b>	<b>23</b>
<b>W1:</b> WMT . . . . .	24
<b>W2:</b> WASSA . . . . .	29
<b>W3:</b> DiscoMT . . . . .	31
<b>W4:</b> LOUHI . . . . .	33
<b>W5:</b> LSDSem . . . . .	35
<b>W6:</b> CogACLL . . . . .	37
<b>W7:</b> VL . . . . .	39
<b>4 Main Conference: Saturday, September 19</b>	<b>41</b>
Invited Speaker: Yoshua Bengio . . . . .	42
Session 1 . . . . .	43
Session 2 . . . . .	54
Session 3 . . . . .	67

<b>5 Main Conference: Sunday, September 20</b>	<b>79</b>
Invited Speaker: Justin Grimmer . . . . .	80
Session 4 . . . . .	81
Session 5 . . . . .	93
Session 6 . . . . .	104
<b>6 Main Conference: Monday, September 21</b>	<b>117</b>
Session 7 . . . . .	120
Session 8 . . . . .	133
Session 9 . . . . .	147
<b>Author index</b>	<b>159</b>
<b>Sponsor Ads</b>	<b>174</b>
<b>Food Quick Reference</b>	<b>180</b>

## Conference Information

### Preface by the General Chair

August 25, 2015

Welcome to the 2015 Conference on Empirical Methods in Natural Language Processing. EMNLP is annually organized by SIGDAT, the Association for Computational Linguistics' special interest group on linguistic data and corpus-based approaches to NLP. This year the conference will be held on September 17–21 in the enchanting city of Lisbon, Portugal.<sup>1</sup>

EMNLP has continued to increase in prominence as one of the most important conferences in Natural Language Processing (NLP). This year the conference has experienced an unprecedented boost in submitted papers. I believe that this reflects both the growth of the NLP field and also the health and strength of the conference itself, with a history of many years of solid work. With this level of interest at submission time, we are also expecting a record attendance. The conference will span a five-day period this year, and it requires a growing organization structure.

Some of the features introduced in EMNLP 2014 will continue this year (e.g., tutorials, new chairs, posters as parallel sessions, flat rates and flexibility for tutorials and workshops, etc.). We also introduce some innovations, like a revised selection process for which talks are presented as talks versus posters.

This year I had the privilege of coordinating the conference from my General Chair position. This has been a very instructive and enriching exercise which showed me the conference as a whole, from many different angles. Prefaces in the proceedings invariably praise the team of organizers. This one will not be an exception. Organizing a large conference as EMNLP requires excellent people working as a team in multiple interrelated tasks. I have been lucky to work with an outstanding team of people, from whom I learnt a lot. These aren't empty words. I would like to thank each and every chair for the hard work that made the conference a reality.

The Program Chairs, Jian Su and Chris Callison-Burch, did an excellent job at putting together a very interesting program with over 300 papers. They had to deal with a very large number of submissions, which exceeded even our most optimistic expectations. As a consequence, they were forced to be creative and to find solutions on the fly to adapt to the situation. They recruited the largest ever program committee and successfully managed a huge reviewing and decision making process under a very tight schedule. A real gift for the general chair. They complemented the program with very interesting keynote speakers, Yoshua Bengio and Justin Grimmer who will present exciting research topics for our community.

<sup>1</sup>Conference website: <http://www.emnlp2015.org>

The EMNLP 2015 main conference is accompanied by 7 workshops and 8 tutorials during the first two days. The Workshops Chairs, Zornitsa Kozareva and Jörg Tiedemann, and the Tutorials Chairs, Maggie Li and Khalil Sima'an, conducted the selection processes in a joint effort with the other ACL conferences in 2015 (NAACL and ACL-IJCNLP). This has been the standard procedure from last years. It has the advantage of starting early, avoiding duplicated reviewing and allowing a more balanced selection among conferences. EMNLP attracted a varied and interesting set of workshops and tutorials, which gives more value to the conference.

Daniele Pighin and Yuval Marton were responsible for the always difficult and sometimes thankless task of putting together the conference publications. This is a very complex effort which involves coordination with almost everyone in the team under the pressure of hard publication deadlines. Yuval is serving in this position for a second year. Staggered two year terms for publication chairs is a new addition for EMNLP starting this year, and we hope that it will be a permanent feature. In the first year, publication chairs will learn and do the bulk of proceedings compilation. During the second year their role will be more advisory, instructing and helping the first-year chair. This procedure will help the transmission of the necessary know-how from year to year. Thanks to Yuval and Daniele for accepting the challenge and making it work wonderfully. Finally, this is the second year that EMNLP uses a mobile app for the conference program (Conference4me). The publication chairs also coordinated the integration of the app with SoftConf, which is now smoother and more seamless.

The local organization team was led by André Martins and João Graça. They did an amazing job, working hard and with all the complexities and subtleties of local arrangements. One of the keys for the success was the creation of a large team of local organizers with clearly defined roles and responsibilities. They appointed very committed people: Isabel Trancoso (Local Publicity Chair), Fernando Batista (Handbook Chair), Bruno Martins (Website and App Chair), Luísa Coheur (Student Volunteer Coordinator), and Helena Moniz (Local Arrangements Chair). Thanks to all. I am especially pleased about the new website, which was revamped and looks more professional everyday. This is certainly a good investment for the future.

A large conference as EMNLP needs to focus on dissemination activities too. Barbara Plank acted as the international Publicity Chair. She did a fantastic job and coordinated very well with the local publicity and the website chairs. The calls for papers, calls for participation, and main news of the conference were timely distributed through ACL, the usual distribution lists, and also through the conference website and two Facebook and Twitter accounts. The EMNLP2015 Twitter account garnered more followers than in previous years.

I am really grateful to SIGDAT. Its secretary, Noah Smith, acted as the liaison between SIGDAT and the conference organizers. He was always available and ready to help with very good advice. SIGDAT also provided the funds for the student scholarship program. These grants help covering traveling expenses to a dozen of students. The committee appointed for collecting the applications and making the decisions was formed by Francisco Guzmán and Lluís Padró, who had to analyze all the information and decide the awardees in only a few days.

Another sign of the health of EMNLP and the field in general is the interest showed by sponsors. Thanks to the work of our sponsorship team, formed by João Graça and Hang Li, in coordination with the ACL International Sponsorship Committee, we got a record number of 13 sponsors for EMNLP 2015 (2 platinum, 3 gold, 6 silver and 2 bronze). In addition to these direct sponsors, we also have several smaller supporters, exhibitors, and institutional

partners. We are extremely grateful to all these companies and institutions, which make a better conference possible at a more affordable registration fee.

Additionally, we counted on the invaluable help of Priscilla Rasmussen, supporting the local organization in all fronts with her broad experience. She took care of the registration process too. We also got very good advice, know-how, and helpful software and forms from last year general chair and local organizers, Alessandro Moschitti and Kareem Darwish. Thank you.

Finally, I would like to thank the authors of submitted and accepted papers, and all the attendees to the conference, who will be the main actors from September 17 to September 21, 2015. I am convinced that we will experience a fantastic conference, scientifically exciting and full of fond memories, in the unique environment of Lisbon.

*Lluís Márquez*  
EMNLP 2015 General Chair

## Preface by the Program Committee Co-Chairs

August 25, 2015

Welcome to the 2015 Conference on Empirical Methods in Natural Language Processing! This year we received a record number of submissions. There were 1300 valid submissions. The 600 long papers and 700 short papers were allocated to one of 15 areas. The most popular areas this year were Semantics, Statistical Models and Machine Learning Methods, Text Mining and NLP applications, and Machine Translation.

Reviewing for a conference this size involves an enormous volunteer effort from many individuals. We are very grateful to our 30 area chairs and to the more than 900 researchers who reviewed the submissions. We accepted 312 papers (157 long and 155 short papers), representing a global acceptance rate of 24%. An additional 17 papers accepted by the TACL journal were presented at the conference as well.

To decide whether the accepted papers should be presented as talks or posters, we asked the area chairs, the reviewers, and the authors of accepted papers to vote on which papers they would like to attend. We showed the title of each paper and its abstract, but not its authors. 400 people provided their input. We selected talks based on popularity, while ensuring that each area was represented by at least one session. Our rationale for taking a vote was that papers that many people wanted to attend would be better served by presenting a talk in a large room, while papers with more specialized interest would benefit from the one-on-one interactions facilitated by posters. Rather than doing large plenary poster sessions, we have scheduled two parallel poster sessions with small batches of thematically similar papers that will be run simultaneously with the talks.

We selected best papers from a shortlist of 20 papers that were nominated by the area chairs. The best paper committee ranked the nominees, and based on their rankings we selected the following papers for the best paper awards:

- Best paper - *Broad-coverage CCG Semantic Parsing with AMR* by Yoav Artzi, Kenton Lee and Luke Zettlemoyer.
- Best paper - *Semantically Conditioned LSTM-based Natural Language Generation for Spoken Dialogue Systems* by Tsung-Hsien Wen, Milica Gasic, Nikola Mrkšić, Pei-Hao Su, David Vandyke and Steve Young.

IBM has provided a cash scholarship for us to award to the best student paper. This will go to Tsung-Hsien Wen, since he is currently a student. The following papers received an honorable mention for the best paper award:

- Honorable mention for best paper - *Traversing Knowledge Graphs in Vector Space* by Kelvin Guu, John Miller and Percy Liang.
- Honorable mention for best paper - *Building a shared world: mapping distributional to model-theoretic semantic spaces* by Aurélie Herbelot and Eva Maria Vecchi.
- Honorable mention for best paper - *Language Understanding for Text-based Games using Deep Reinforcement Learning* by Karthik Narasimhan, Tejas Kulkarni and Regina Barzilay.
- Honorable mention for best short paper - *Joint Lemmatization and Morphological Tagging with Lemming* by Thomas Müller, Ryan Cotterell, Alexander Fraser and Hinrich Schütze.

- Honorable mention for best short paper - *Semi-Supervised Bootstrapping of Relationship Extractors with Distributional Semantics* by David S. Batista, Bruno Martins and Mário J. Silva.

This year we created a new “Best data set or resource” award, since so much work in our community is driven by data. The paper that receiving this inaugural distinction is:

- Best data set or resource - *A large annotated corpus for learning natural language inference* by Samuel R. Bowman, Gabor Angeli, Christopher Potts and Christopher D. Manning.

With two honorable mentions:

- Notable data set or resource - *That's So Annoying!!!: A Lexical and Frame-Semantic Embedding Based Data Augmentation Approach to Automatic Categorization of Annoying Behaviors using #petpeeve Tweets* by William Yang Wang and Diyi Yang.
- Notable data set or resource - *Modeling Reportable Events as Turning Points in Narrative* by Jessica Ouyang and Kathy McKeown.

We decided to give more awards than in past years by recognizing papers with honorable mentions and by creating the new best data or resource award. Our goal was to recognize roughly the top 1% of all of the submissions to the conference with awards (recognizing approximately the top 5% of accepted papers). We are very grateful to our invited speakers Yoshua Bengio and Justin Grimmer.

Yoshua Bengio is professor of Computer Science and Operations Research at the Université de Montréal. He is the author of two books and more than 200 publications, the most cited being in the areas of deep learning, recurrent neural networks, probabilistic learning algorithms, natural language processing and manifold learning. He co-directs the Canadian Institute for Advanced Research’s program on deep learning. He is on the board of NIPS. Professor Bengio’s research into deep learning has had a dramatic impact on the field of NLP in the past few years, and has invigorated interest in AI through machine learning.

Justin Grimmer is an associate professor of Political Science at Stanford University. His research uses statistical methods to examine American politics. He is the author of two books on the topic “Representational Style in Congress: What Legislators Say and Why It Matters” and “The Impression of Influence: How Legislator Communication and Government Spending Cultivate a Personal Vote.” His work has appeared in the American Political Science Review, American Journal of Political Science, Journal of Politics, Political Analysis, Proceedings of the National Academy of Sciences, Regulation and Governance, and Poetics. Professor Grimmer’s research points to exciting new directions for computational social science and how the field of NLP can facilitate research in many areas.

We thank them in advance for coming to the conference and sharing their insights.

We would also like to thank our general chair Lluís Màrquez, André Martins and João Graça and colleagues for their excellent work with the local organization, and Yuval Marton and Daniele Pighin for doing an excellent job assembling these proceedings.

We thank SIGDAT for inviting us to serve as Program Co-Chairs of EMNLP 2015. We hope that the conference is an excellent one. Enjoy your stay in Lisbon!

Chris Callison-Burch and Jian Su  
EMNLP 2015 Program Committee Co-Chairs

## Conference Committee

---

### General Chair

Lluís Márquez, Qatar Computing Research Institute

### Program co-Chairs

Chris Callison-Burch, University of Pennsylvania  
Jian Su, Institute for Infocomm Research (I2R)

### Workshops co-Chairs

Zornitsa Kozareva, Yahoo! Labs  
Jörg Tiedemann, University of Helsinki

### Tutorial co-Chairs

Maggie Li, Hong Kong Polytechnic University  
Khalil Sima'an, University of Amsterdam

### Publication co-Chairs

Daniele Pighin, Google Inc.  
Yuval Marton, Microsoft Corp.

### Publicity Chair

Barbara Plank, University of Copenhagen

### Sponsorship Team

João Graça, Unbabel Inc.  
Hang Li, Huawei Technologies (ISC Representative for EMNLP)

### Student Scholarship co-Chairs

Fancisco Guzmán, Qatar Computing Research Institute  
Lluís Padró, Technical University of Catalonia

### SIGDAT Liaison

Noah Smith, University of Washington

### Local co-Chairs

André Martins, Priberam  
João Graça, Unbabel Inc.

### Program co-Chairs

Chris Callison-Burch, University of Pennsylvania  
Jian Su, Institute for Infocomm Research (I2R)

**Area Chairs***Computational Psycholinguistics*

Vera Demberg, Saarland University

*Discourse, Dialogue, and Pragmatics*

Kentaro Inui, Tohoku University

Oliver Lemon, Heriot Watt University

*Information Extraction*

Heng Ji, Rensselaer Polytechnic Institute

Sebastian Riedel, University College London

Jing Jiang, Singapore Management University

*Information Retrieval and Question Answering*

Zhu Xiaoyan, Tsinghua University

Liu Ting, Harbin Institute of Technology

*Language and Vision*

Julia Hockenmaier, University of Illinois Urbana Champaign

*Machine Translation and Multilinguality*

Zhang Min, SooChow University

Lucia Specia, University of Sheffield

Marine Carput, University of Maryland

*NLP for the Web and Social Media, including Computational Social Science*

Alice Oh, KAIST

Brendan O'Connor, University of Massachusetts Amherst

*Phonology, Morphology, and Word Segmentation*

Greg Kondrak, University of Alberta

Xu Sun, Peking University

*Semantics*

Marco Baroni, University of Trento

Shiqi Zhao, Baidu

Percy Liang, Stanford University

*Sentiment Analysis and Opinion Mining*

Ku Lun-Wei, Institute of Information Science, Academia Sinica

He Yulan, Aston University

*Spoken Language Processing*

Mari Ostendorf, University of Washington

*Statistical Models and Machine Learning Methods*

Trevor Cohn, University of Melbourne

Jordan Boyd-Graber, University of Colorado

*Summarization and Generation*

Meg Mitchell, Microsoft Research  
Furu Wei, Microsoft Research Asia

*Tagging, Chunking, Syntax and Parsing*

Shay Cohen, University of Edinburgh  
Yue Zhang, Singapore University of Technology and Design  
Yusuke Miyao, National Institute of Informatics, Japan

*Text Mining and NLP applications*

Sophia Ananiadou, University of Manchester

**Local Organization Committee**

*Local co-Chairs*

André Martins, Priberam  
João Graça, Unbabel Inc.

*Local Arrangements Chair*

Helena Moniz, University of Lisbon & INESC-ID

*Local Publicity Chair*

Isabel Trancoso, University of Lisbon

*Conference Handbook Chair*

Fernando Batista, University Institute of Lisbon (ISCTE-IUL)

*Website and App Chair*

Bruno Martins, University of Lisbon

*Student Volunteer Coordinator*

Luísa Coheur, University of Lisbon

## Social events

EMNLP 2015 will feature welcome and farewell receptions, respectively on the 18th of September (second day of the conference) and on the 21st of September (the last day of the conference). These receptions will take place at Culturgest, the main conference venue. EMNLP 2015 will also feature a conference dinner, on Sunday the 20th of September, which will take place at Páteo Alfacinha, in the Ajuda area of Lisbon. Transportation to Páteo Alfacinha will be arranged.



Páteo Alfacinha was created in 1981, featuring several spaces typical from the beginning of XX century Lisbon: a chapel, a barber shop, a tavern, a bakery, a pub, an antiquary, and houses for the “alfacinhas” (inhabitants of Lisbon). Nowadays, Páteo Alfacinha remains a multifaceted place, with restaurants and facilities for events such as EMNLP’s conference dinner.

### Welcome Reception

Friday, September 18th at 18:30 at Culturgest

### Conference Dinner

Sunday, September 20th at 19:00 at Páteo Alfacinha  
First buses depart from Culturgest at 18:15.  
Last returning bus depart from Páteo Alfacinha at 24:00.

### Farewell Drink

Monday, September 21th at 18:30 at Culturgest



**Tutorials: Thursday – Friday, September, 17–18****September 17****09:00 – 12:30 Morning Tutorials**

Semantic Similarity Frontiers: From Concepts to Documents

*David Jurgens and Mohammad Taher Pilehvar* (Small Auditorium)

Personality Research for NLP

*Yair Neuman* (Room 2)

**12:30 – 14:00 Lunch break****14:00 – 17:30 Afternoon Tutorials**

Transparent Machine Learning for Information Extraction: State-of-the-Art and the Future

*Laura Chiticariu, Yunyao Li, and Frederick Reiss* (Small Auditorium)

Knowledge Acquisition for Web Search

*Marius Pasca* (Room 2)

**September 18****09:00 – 12:30 Morning Tutorials**

Learning Semantic Relations from Text

*Preslav Nakov, Vivi Nastase, Diarmuid Ó Séaghdha, and Stan Szpakowicz* (Small Auditorium)

Applications of Social Media Text Analysis

*Atefeh Farzindar and Diana Inkpen* (Room 2)

**12:30 – 14:00 Lunch break****14:00 – 17:30 Afternoon Tutorials**

Robust Semantic Analysis of Multiword Expressions with FrameNet

*Miriam R. L. Petrucci and Valia Kordoni* (Room 2)

Computational Analysis of Affect and Emotion in Language

*Saif Mohammad and Cecilia Ovesdotter Alm* (Small Auditorium)

## Semantic Similarity Frontiers: From Concepts to Documents

*David Jurgens and Mohammad Taher Pilehvar*

September 17, 2015 – 09:00–12:30

Small Auditorium

Semantic similarity forms a central component in many NLP systems, from lexical semantics, to part of speech tagging, to social media analysis. Recent years have seen a renewed interest in developing new similarity techniques, buoyed in part by work on embeddings and by SemEval tasks in Semantic Textual Similarity and Cross-Level Semantic Similarity. The increased interest has led to hundreds of techniques for measuring semantic similarity, which makes it difficult for practitioners to identify which state-of-the-art techniques are applicable and easily integrated into projects and for researchers to identify which aspects of the problem require future research.

This tutorial synthesizes the current state of the art for measuring semantic similarity for all types of conceptual or textual pairs and presents a broad overview of current techniques, what resources they use, and the particular inputs or domains to which the methods are most applicable. We survey methods ranging from corpus-based approaches operating on massive or domains-specific corpora to those leveraging structural information from expert-based or collaboratively-constructed lexical resources. Furthermore, we review work on multiple similarity tasks from sense-based comparisons to word, sentence, and document-sized comparisons and highlight general-purpose methods capable of comparing multiple types of inputs. Where possible, we also identify techniques that have been demonstrated to successfully operate in multilingual or cross-lingual settings.

Our tutorial provides a clear overview of currently-available tools and their strengths for practitioners who need out of the box solutions and provides researchers with an understanding of the limitations of current state of the art and what open problems remain in the field. Given the breadth of available approaches, participants will also receive a detailed

---

**David Jurgens** is a postdoctoral scholar at McGill University. He received his PhD from the University of California, Los Angeles. His research interests include lexical semantics, word sense disambiguation, latent attribute inference, and the relationship between language and location. He is currently co-chairing the 2015 and 2016 International Workshops on Semantic Evaluation (SemEval). His research has been featured in the MIT Technology Review, Forbes, Business Insider, and Schneier on Security.

**Mohammad Taher Pilehvar** is a postdoctoral scholar at Sapienza University of Rome. He received his PhD from the same university under the supervision of Roberto Navigli. He does research in multiple areas of Lexical Semantics such as semantic similarity and Word Sense Disambiguation (WSD). His main focus is on unified graph-based semantic similarity measures and large-scale frameworks for the evaluation of WSD systems. He has co-organized a task on Cross Level Semantic Similarity in SemEval-2014 (Jurgens et al., 2014) and is the first author of a paper on semantic similarity that was nominated for the best paper award at ACL 2013 (Pilehvar et al., 2013).

bibliography of approaches (including those not directly covered in the tutorial), annotated according to the approaches abilities, and pointers to when open-source implementations of the algorithms may be obtained.

## **Personality Research for NLP**

*Yair Neuman*

September 17, 2015 – 09:00–12:30

Room 2

“Personality” is a psychological concept describing the individual’s characteristic patterns of thought, emotion, and behavior. In the context of Big Data and granular analytics, it is highly important to measure the individual’s personality dimensions as these may be used for various practical applications. However, personality has been traditionally studied by questionnaires and other forms of low tech methodologies. The availability of textual data and the development of powerful NLP technologies, invite the challenge of automatically measuring personality dimensions for various applications from granular analytics of customers to the forensic identification of potential offenders. While there are emerging attempts to address this challenge, these attempts almost exclusively focus on one theoretical model of personality and on classification tasks limited when tagged data are not available.

The major aim of the tutorial is to provide NLP researchers with an introduction to personality theories that may empower their scope of research. In addition, two secondary aims are to survey some recent directions in computational personality and to point to future directions in which the field may be developed (e.g. Textual Entailment for Personality Analytics).

---

**Yair Neuman** (Ben-Gurion Univ. of the Negev) is the co-director of the Behavioral Insights Research Lab at the University of Toronto and a senior fellow at the Brain Sciences Foundation. Among his fields of interest are the interface of NLP and psychology and the development of novel cognitive-psychological technologies.

## **Transparent Machine Learning for Information Extraction: State-of-the-Art and the Future**

*Laura Chiticariu, Yunyao Li, and Frederick Reiss*

September 17, 2015 – 14:00–17:30

Small Auditorium

The rise of Big Data analytics over unstructured text has led to renewed interest in information extraction (IE). These applications need effective IE as a first step towards solving end-to-end real world problems (e.g. biology, medicine, finance, media and entertainment, etc). Much recent NLP research has focused on addressing specific IE problems using a pipeline of multiple machine learning techniques. This approach requires an analyst with the expertise to answer questions such as: “What ML techniques should I combine to solve this problem?”; “What features will be useful for the composite pipeline?”; and “Why is my model giving the wrong answer on this document?”. The need for this expertise creates problems in real world applications. It is very difficult in practice to find an analyst who both understands the real world problem and has deep knowledge of applied machine learning. As a result, the real impact by current IE research does not match up to the abundant opportunities available.

In this tutorial, we introduce the concept of transparent machine learning. A transparent ML technique is one that:

- produces models that a typical real world user can read and understand;
- uses algorithms that a typical real world user can understand; and
- allows a real world user to adapt models to new domains.

The tutorial is aimed at IE researchers in both the academic and industry communities who are interested in developing and applying transparent ML.

---

**Laura Chiticariu** is a Research Staff Member at IBM Research – Almaden. She received her Ph.D from U.C. Santa Cruz in 2008. Her current research focuses on improving developmental support in information extraction systems.

**Yunyao Li** is a Research Staff Member and Research Manager at IBM Research – Almaden. She received her Ph.D from the University of Michigan, Ann Arbor in 2007. She is particularly interested in designing, developing and analyzing large scale systems that are usable by a wide spectrum of users. Towards this direction, her current research focuses on enterprise-scale natural language processing.

**Frederick Reiss** is a Research Staff Member at IBM Research – Almaden. He received his Ph.D. from U.C. Berkeley in 2006. His research focuses on improving the scalability of text analytics in enterprise applications.

## **Knowledge Acquisition for Web Search**

*Marius Pasca*

September 17, 2015 – 14:00–17:30

Room 2

The identification of textual items, or documents, that best match a user's information need, as expressed in search queries, forms the core functionality of information retrieval systems. Well-known challenges are associated with understanding the intent behind user queries; and, more importantly, with matching inherently-ambiguous queries to documents that may employ lexically different phrases to convey the same meaning. The conversion of semi-structured content from Wikipedia and other resources into structured data produces knowledge potentially more suitable to database-style queries and, ideally, to use in information retrieval. In parallel, the availability of textual documents on the Web enables an aggressive push towards the automatic acquisition of various types of knowledge from text. Methods developed under the umbrella of open-domain information extraction acquire open-domain classes of instances and relations from Web text. The methods operate over unstructured or semi-structured text available within collections of Web documents, or over relatively more intriguing streams of anonymized search queries. Some of the methods import the automatically-extracted data into human-generated resources, or otherwise exploit existing human-generated resources. In both cases, the goal is to expand the coverage of the initial resources, thus providing information about more of the topics that people in general, and Web search users in particular, may be interested in.

---

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

## Learning Semantic Relations from Text

*Preslav Nakov, Vivi Nastase, Diarmuid Ó Séaghdha, and Stan Szpakowicz*

September 18, 2015 – 09:00–12:30

Small Auditorium

Every non-trivial text describes interactions and relations between people, institutions, activities, events and so on. What we know about the world consists in large part of such relations, and that knowledge contributes to the understanding of what texts refer to. Newly found relations can in turn become part of this knowledge that is stored for future use.

To grasp a text’s semantic content, an automatic system must be able to recognize relations in texts and reason about them. This may be done by applying and updating previously acquired knowledge. We focus here in particular on semantic relations which describe the interactions among nouns and compact noun phrases, and we present such relations from both a theoretical and a practical perspective. The theoretical exploration sketches the historical path which has brought us to the contemporary view and interpretation of semantic relations. We discuss a wide range of relation inventories proposed by linguists and by language processing people. Such inventories vary by domain, granularity and suitability for downstream applications.

On the practical side, we investigate the recognition and acquisition of relations from texts. In a look at supervised learning methods, we present available datasets, the variety of features which can describe relation instances, and learning algorithms found appropriate for the task.

Next, we present weakly supervised and unsupervised learning methods of acquiring relations from large corpora with little or no previously annotated data. We show how enduring the bootstrapping algorithm based on seed examples or patterns has proved to be, and how it has been adapted to tackle Web-scale text collections. We also show a few

---

**Preslav Nakov**, a Senior Scientist at the Qatar Computing Research Institute, part of Qatar Foundation, holds a Ph.D. from the University of California at Berkeley. His research interests include computational linguistics and NLP, machine translation, lexical semantics, Web as a corpus and biomedical text processing.

**Vivi Nastase** is a researcher at the Fondazione Bruno Kessler in Trento, working mainly on lexical semantics, semantic relations, knowledge acquisition and language evolution. She holds a Ph.D. from the University of Ottawa, Canada.

**Diarmuid Ó Séaghdha**, Senior NLP Researcher at VocalIQ and Visiting Industrial Fellow at the University of Cambridge, holds a Ph.D. from the University of Cambridge. His research interests include discourse and dialog, lexical and relational semantics, machine learning for NLP, scientific text mining and social media analysis.

**Stan Szpakowicz**, an emeritus professor of Computer Science at the University of Ottawa, holds a Ph.D. from the University of Warsaw and a D.Sc. from the Polish Academy of Sciences. He has been active in NLP since 1969. His recent interests include lexical resources, semantic relations and emotion analysis.

machine learning techniques which can perform fast and reliable relation extraction by taking advantage of data redundancy and variability.

## Applications of Social Media Text Analysis

*Atefeh Farzindar and Diana Inkpen*

September 18, 2015 – 09:00–12:30

Room 2

Analyzing social media texts is a complex problem that becomes difficult to address using traditional Natural Language Processing (NLP) methods. Our tutorial focuses on presenting new methods for NLP tasks and applications that work on noisy and informal texts, such as the ones from social media.

Automatic processing of large collections of social media texts is important because they contain a lot of useful information, due to the increasing popularity of all types of social media. Use of social media and messaging apps grew 203 percent year-on-year in 2013, with overall app use rising 115 percent over the same period, as reported by Statista, citing data from Flurry Analytics. This growth means that 1.61 billion people are now active in social media around the world and this is expected to advance to 2 billion users in 2016, led by India. The research shows that consumers are now spending daily 5.6 hours on digital media including social media and mobile internet usage.

At the heart of this interest is the ability for users to create and share content via a variety of platforms such as blogs, micro-blogs, collaborative wikis, multimedia sharing sites, social networking sites. The unprecedented volume and variety of user-generated content, as well as the user interaction network constitute new opportunities for understanding social behavior and building socially intelligent systems. Therefore it is important to investigate methods for knowledge extraction from social media data. Furthermore, we can use this information to detect and retrieve more related content about events, such as photos and video clips that have caption texts.

---

Dr. **Atefeh Farzindar** is the CEO of NLP Technologies Inc. and Adjunct Professor at University of Montreal. She has served as Chair of the technology sector of the Language Industry Association Canada (AILIA) (2009-2013), vice president of The Language Technologies Research Centre (LTRC) of Canada (2012-2014) and a member of the Natural Sciences and Engineering Research Council of Canada (NSERC) Computer Science Liaison Committee (2014-2015). Recently, she authored a book chapter in Social Network Integration in Document Summarization, Innovative Document Summarization Techniques: Revolutionizing Knowledge Understanding, IGI Global publisher January 2014.

Dr. **Diana Inkpen** is a Professor the School of Electrical Engineering and Computer Science at the University of Ottawa. Her research interests and expertise are in natural language processing, in particular lexical semantics as applied to near synonyms and nuances of meaning, word and text similarity, classification of texts by emotion and mood, information retrieval from spontaneous speech, extraction of semantic frames, and lexical choice in natural language generation. She published more than 25 journal papers, 85 conference papers, and 6 book chapters. She is an associated editor of the Computational Intelligence and Natural Language Engineering journals.

## **Robust Semantic Analysis of Multiword Expressions with FrameNet**

*Miriam R. L. Petrucci and Valia Kordoni*

September 18, 2015 – 14:00–17:30

Room 2

This tutorial will give participants a solid understanding of the linguistic features of multiword expressions (MWEs), focusing on the semantics of such expressions and their importance for natural language processing and language technology, with particular attention to the way that FrameNet ([framenet.icsi.berkeley.edu](http://framenet.icsi.berkeley.edu)) handles this wide spread phenomenon. Our target audience includes researchers and practitioners of language technology, not necessarily experts in MWEs or knowledgeable about FrameNet, who are interested in NLP tasks that involve or could benefit from considering MWEs as a pervasive phenomenon in human language and communication.

NLP research has been interested in automatic processing of multiword expressions, with reports on and tasks relating to such efforts presented at workshops and conferences for at least ten years (e.g. ACL 2003, LREC 2008, COLING 2010, EACL 2014). Overcoming the challenge of automatically processing MWEs remains elusive in part because of the difficulty in recognizing, acquiring, and interpreting such forms.

Indeed the phenomenon manifests in a range of linguistic forms (as Sag et al. (2001), among many others, have documented), including: noun + noun compounds (e.g. fish knife, health hazard etc.); adjective + noun compounds (e.g. political agenda, national interest, etc.); particle verbs (shut up, take out, etc.); prepositional verbs (e.g. look into, talk into, etc.); VP idioms, such as kick the bucket, and pull someone's leg, along with less obviously idiomatic forms like answer the door, mention someone's name, etc.; expressions that have their own mini-grammars, such as names with honorifics and terms of address (e.g. Rabbi Lord Jonathan Sacks), kinship terms (e.g. second cousin once removed), and time expressions (e.g. January 9, 2015); support verb constructions (e.g. verbs: take a bath, make a promise, etc; and prepositions: in doubt, under review, etc.). Linguists address issues of polysemy, compositionality, idiomticity, and continuity for each type included here.

While native speakers use these forms with ease, the treatment and interpretation of MWEs in computational systems requires considerable effort due to the very issues that concern linguists

---

**Miriam R. L. Petrucci** received her PhD in Linguistics from the University of California, Berkeley. A key member of the team developing FrameNet almost since the project's founding, her research interests include semantics, knowledge base development, grammar and lexis, lexical semantics, Frame Semantics and Construction Grammar

**Valia Kordoni** received her PhD in Computational Linguistics from the University of Essex, UK. She joined the Department of English Studies, Humboldt University Berlin in 2012, where she is Research Professor of Linguistics. Her main research interests are in deep linguistic processing, semantic analysis, and multiword expressions.

## Computational Analysis of Affect and Emotion in Language

*Saif Mohammad and Cecilia Ovesdotter Alm*

September 18, 2015 – 14:00–17:30

Small Auditorium

Computational linguistics has witnessed a surge of interest in approaches to emotion and affect analysis, tackling problems that extend beyond sentiment analysis in depth and complexity. This area involves basic emotions (such as joy, sadness, and fear) as well as any of the hundreds of other emotions humans are capable of (such as optimism, frustration, and guilt), expanding into affective conditions, experiences, and activities. Leveraging linguistic data for computational affect and emotion inference enables opportunities to address a range of affect-related tasks, problems, and non-invasive applications that capture aspects essential to the human condition and individuals' cognitive processes. These efforts enable and facilitate human-centered computing experiences, as demonstrated by applications across clinical, socio-political, artistic, educational, and commercial domains. Efforts to computationally detect, characterize, and generate emotions or affect-related phenomena respond equally to technological needs for personalized, micro-level analytics and broad-coverage, macro-level inference, and they have involved both small and massive amounts of data.

While this is an exciting area with numerous opportunities for members of the ACL community, a major obstacle is its intersection with other investigatory traditions, necessitating knowledge transfer. This tutorial comprehensively integrates relevant concepts and frameworks from linguistics, cognitive science, affective computing, and computational linguistics in order to equip researchers and practitioners with the adequate background and knowledge to work effectively on problems and tasks either directly involving, or benefiting from having an understanding of, affect and emotion analysis.

---

**Saif Mohammad** has research interests in computational linguistics and natural language processing, especially lexical semantics and affect analysis. He develops computational models for sentiment analysis, emotion detection, semantic distance, and lexical-semantic relations such as word-pair antonymy. His team has developed a sentiment analysis system which ranked first in SemEval shared tasks on the sentiment analysis of tweets and on aspect-based sentiment analysis. His word-emotion association resource, the NRC Emotion Lexicon, is widely used for text analysis and information visualization. His recent work on generating music from emotions in text garnered widespread media attention, including articles in Time, LiveScience, io9, The Physics arXiv Blog, PC World, and Popular Science.

**Cecilia Ovesdotter Alm** is a computational linguist dedicated to advancing the understanding of affective and subjective meaning across linguistic modalities and multimodal data. Her work focuses on linguistic annotation and resource development for affect-related problems, as well as computational modeling involving text and speech, image understanding, and linguistic or multimodal sensing in this area. She has published *Affect in Text and Speech* (2009) as well as articles in proceedings and journals, representing over a decade of related research.

There is a substantial body of work in traditional sentiment analysis focusing on positive and negative sentiment. This tutorial covers approaches and features that migrate well to affect analysis. We also discuss key differences from sentiment analysis, and their implications for analyzing affect and emotion.

The tutorial begins with an introduction that highlights opportunities, key terminology, and interesting tasks and challenges (1). The body of the tutorial covers characteristics of emotive language use with emphasis on relevance for computational analysis (2); linguistic data—from conceptual analysis frameworks via useful existing resources to important annotation topics (3); computational approaches for lexical semantic emotion analysis (4); computational approaches for emotion and affect analysis in text (5); visualization methods (6); and a survey of application areas with affect-related problems (7). The tutorial concludes with an outline of future directions and a discussion with participants about the areas relevant to their respective tasks of interest (8).

Besides attending the tutorial, tutorial participants receive electronic copies of tutorial slides, a complete reference list, as well as a categorized annotated bibliography that concentrates on seminal works, recent important publications, and other products and resources for researchers and developers.

# 3

## Workshops

<b>September 17th – 18th, Thursday – Friday</b>		page
MA	WMT: Tenth Workshop on Statistical Machine Translation	24
<b>September 17th – Thursday</b>		
Room 1	WASSA: 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis	29
Room 6	DiscoMT: Second Workshop on Discourse in Machine Translation	31
Room 3	LOUHI: Sixth Workshop on Health Text Mining and Information Analysis	33
<b>September 18th – Friday</b>		
Room 1	LSDSem: First Workshop on Linking Models of Lexical, Sentential and Discourse-level Semantics	35
Room 3	CogACLL: Sixth Workshop on Cognitive Aspects of Computational Language Learning	37
Room 6	VL: Fourth Workshop on Vision and Language	39

## Workshop 1: WMT

### Tenth Workshop on Statistical Machine Translation

---

Organizers: *Ondřej Bojar, Christian Federmann, Barry Haddow, Chris Hokamp, Matthias Huck, Pavel Pecina, Philipp Koehn, Christof Monz, Matteo Negri, Matt Post, Caroline Scarton, Lucia Specia, and Marco Turchi*

Venue: Main Auditorium

## Thursday, September 17, 2015

09:00–09:05 **Opening Remarks**

09:05–09:50 **Session 1: Shared Tasks**

- Findings of the 2015 Workshop on Statistical Machine Translation  
*Ondřej Bojar, Rajen Chatterjee, Christian Federmann, Barry Haddow, Matthias Huck, Chris Hokamp, Philipp Koehn, Varvara Logacheva, Christof Monz, Matteo Negri, Matt Post, Carolina Scarton, Lucia Specia, and Marco Turchi*

09:50–10:30 **Session 2: Data Selection**

09:50–10:10 Statistical Machine Translation with Automatic Identification of Translationese  
*Naama Twitto, Noam Ordan, and Shuly Wintner*

10:10–10:30 Data Selection With Fewer Words

*Philip Resnik, Xiaodong He, and Mari Ostendorf*

10:30–11:00 **Coffee Break**

11:00–12:30 **Session 3: Poster Session - Shared Task: Automatic Post-Editing**

- The FBK Participation in the WMT15 Automatic Post-editing Shared Task  
*Rajen Chatterjee, Marco Turchi, and Matteo Negri*
- USAAR-SAPE: An English-Spanish Statistical Automatic Post-Editing System  
*Santanu Pal, Mihaela Vela, Sudip Kumar Naskar, and Josef van Genabith*
- Why Predicting Post-Edition is so Hard? Failure Analysis of LIMSI Submission to the APE Shared Task  
*Guillaume Wisniewski, Nicolas Pécheux, and François Yvon*

11:00–12:30 **Session 3: Poster Session - Shared Task: Translation**

- DFKI's experimental hybrid MT system for WMT 2015  
*Eleftherios Avramidis, Maja Popović, and Aljoscha Burchardt*
- ParFDA for Fast Deployment of Accurate Statistical Machine Translation Systems, Benchmarks, and Statistics  
*Ergun Bicici, Qun Liu, and Andy Way*
- CUNI in WMT15: Chimera Strikes Again  
*Ondřej Bojar and Aleš Tamchyna*

- CimS - The CIS and IMS Joint Submission to WMT 2015 addressing morphological and syntactic differences in English to German SMT  
*Fabienne Cap, Marion Weller, Anita Ramm, and Alexander Fraser*
- The Karlsruhe Institute of Technology Translation Systems for the WMT 2015  
*Eunah Cho, Thanh-Le Ha, Jan Niehues, Teresa Herrmann, Mohammed Mediani, Yuqi Zhang, and Alex Waibel*
- New Language Pairs in TectoMT  
*Ondřej Dušek, Luís Gomes, Michal Novák, Martin Popel, and Rudolf Rosa*
- Tuning Phrase-Based Segmented Translation for a Morphologically Complex Target Language  
*Stig-Arne Grönroos, Sami Virpioja, and Mikko Kurimo*
- The AFRL-MITLL WMT15 System: There's More than One Way to Decode It!  
*Jeremy Gwinnup, Tim Anderson, Grant Erdmann, Katherine Young, Christina May, Michael Kazi, Elizabeth Salesky, and Brian Thompson*
- The KIT-LIMSI Translation System for WMT 2015  
*Thanh-Le Ha, Quoc-Khanh DO, Eunah Cho, Jan Niehues, Alexandre Allauzen, François Yvon, and Alex Waibel*
- The Edinburgh/JHU Phrase-based Machine Translation Systems for WMT 2015  
*Barry Haddow, Matthias Huck, Alexandra Birch, Nikolay Bogoychev, and Philipp Koehn*
- Montreal Neural Machine Translation Systems for WMT'15  
*Sébastien Jean, Orhan Firat, Kyunghyun Cho, Roland Memisevic, and Yoshua Bengio*
- GF Wide-coverage English-Finnish MT system for WMT 2015  
*Prasanth Kolachina and Aarne Ranta*
- LIMSIWMT'15 : Translation Task  
*Benjamin Marie, Alexandre Allauzen, Franck Burlot, Quoc-Khanh Do, Julia Ive, Matthieu Labeau, Thomas Lavergne, Kevin Löser, Nicolas Pécheux, and François Yvon*
- UdS-Sant: English—German Hybrid Machine Translation System  
*Santanu Pal, Sudip Naskar, and Josef van Genabith*
- The RWTH Aachen German-English Machine Translation System for WMT 2015  
*Jan-Thorsten Peter, Farzad Toutounchi, Joern Wuebker, and Hermann Ney*
- Exact Decoding with Multi Bottom-Up Tree Transducers  
*Daniel Quernheim*
- Sheffield Systems for the Finnish-English WMT Translation Task  
*David Steele, Karin Sim Smith, and Lucia Specia*
- Morphological Segmentation and OPUS for Finnish-English Machine Translation  
*Jörg Tiedemann, Filip Ginter, and Jenna Kanerva*

- Abu-MaTran at WMT 2015 Translation Task: Morphological Segmentation and Web Crawling  
*Raphael Rubino, Tommi Pirinen, Miquel Esplà-Gomis, Nikola Ljubešić, Sergio Ortiz Rojas, Vassilis Papavassiliou, Prokopis Prokopidis, and Antonio Toral*
- The University of Illinois submission to the WMT 2015 Shared Translation Task  
*Lane Schwartz, Bill Bryce, Chase Geigle, Sean Massung, Yisi Liu, Haoruo Peng, Vignesh Raja, Subhro Roy, and Shyam Upadhyay*
- Edinburgh's Syntax-Based Systems at WMT 2015  
*Philip Williams, Rico Sennrich, Maria Nadejde, Matthias Huck, and Philipp Koehn*

12:30–14:00 **Lunch**

14:00–15:30 **Session 4: Invited Talk**

14:00–15:30 *A Practical Guide to Real-Time Neural Translation (Jacob Devlin)*

15:30–16:00 **Coffee Break**

16:00–17:00 **Session 5: Syntax-Based Translation and Rescoring**

16:00–16:20 Hierarchical Machine Translation With Discontinuous Phrases  
*Miriam Kaeshammer*

16:20–16:40 Discontinuous Statistical Machine Translation with Target-Side Dependency Syntax  
*Nina Seemann and Andreas Maletti*

16:40–17:00 ListNet-based MT Rescoring  
*Jan Niehues, Quoc-Khanh DO, Alexandre Allauzen, and Alex Waibel*

## Friday, September 18, 2015

09:00–09:50 **Session 6: Shared Tasks**

09:00–09:20 *Overview of the Quality Estimation Task*

- Results of the WMT15 Metrics Shared Task  
*Miloš Stanojević, Amir Kamran, Philipp Koehn, and Ondřej Bojar*
- Results of the WMT15 Tuning Shared Task  
*Miloš Stanojević, Amir Kamran, and Ondřej Bojar*

09:50–10:30 **Session 7: Translation Modeling**

09:50–10:10 Extended Translation Models in Phrase-based Decoding  
*Andreas Guta, Joern Wuebker, Miguel Graca, Yunsu Kim, and Hermann Ney*

10:10–10:30 Investigations on Phrase-based Decoding with Recurrent Neural Network Language and Translation Models  
*Tamer Alkhouri, Felix Rietig, and Hermann Ney*

10:30–11:00 **Coffee Break**

11:00–12:30 **Session 8: Poster Session - Shared Task: Metrics**

- Multi-level Evaluation for Machine Translation  
*Boxing Chen, Hongyu Guo, and Roland Kuhn*
- VERTa: a Linguistically-motivated Metric at the WMT15 Metrics Task  
*Elisabet Comelles and Jordi Atserias*
- UPF-Cobalt Submission to WMT15 Metrics Task  
*Marina Fomicheva, Núria Bel, Iria da Cunha, and Anton Malinovskiy*
- Machine Translation Evaluation using Recurrent Neural Networks  
*Rohit Gupta, Constantin Orasan, and Josef van Genabith*
- Alignment-based sense selection in METEOR and the RATATOUILLE recipe  
*Benjamin Marie and Marianna Apidianaki*
- chrF: character n-gram F-score for automatic MT evaluation  
*Maja Popović*
- BEER 1.1: ILLC UvA submission to metrics and tuning task  
*Miloš Stanojević and Khalil Sima'an*
- Predicting Machine Translation Adequacy with Document Embeddings  
*Mihaela Vela and Liling Tan*
- LeBLEU: N-gram-based Translation Evaluation Score for Morphologically Complex Languages  
*Sami Virpioja and Stig-Arne Grönroos*
- CASICT-DCU Participation in WMT2015 Metrics Task  
*Hui Yu, Qingsong Ma, Xiaofeng Wu, and Qun Liu*

11:00–12:30 **Session 8: Poster Session - Shared Task: Quality Estimation**

- Referential Translation Machines for Predicting Translation Quality and Related Statistics  
*Ergun Bicici, Qun Liu, and Andy Way*
- UAlacant word-level machine translation quality estimation system at WMT 2015  
*Miquel Esplà-Gomis, Felipe Sánchez-Martínez, and Mikel Forcada*
- QUality Estimation from ScraTCH (QUETCH): Deep Learning for Word-level Translation Quality Estimation  
*Julia Kreutzer, Shigehiko Schamoni, and Stefan Riezler*
- LORIA System for the WMT15 Quality Estimation Shared Task  
*David Langlois*
- Data enhancement and selection strategies for the word-level Quality Estimation  
*Varvara Logacheva, Chris Hokamp, and Lucia Specia*
- USHEF and USAAR-USHEF participation in the WMT15 QE shared task  
*Carolina Scarton, Liling Tan, and Lucia Specia*
- SHEF-NN: Translation Quality Estimation with Neural Networks  
*Kashif Shah, Varvara Logacheva, Gustavo Paetzold, Frédéric Blain, Daniel Beck, Fethi Bougares, and Lucia Specia*
- Strategy-Based Technology for Estimating MT Quality  
*Liugang Shang, Dongfeng Cai, and Duo Ji*

- UGENT-LT3 SCATE System for Machine Translation Quality Estimation  
*Arda Tezcan, Veronique Hoste, Bart Desmet, and Lieve Macken*

**11:00–12:30 Session 8: Poster Session - Shared Task: Tuning**

- Drem: The AFRL Submission to the WMT15 Tuning Task  
*Grant Erdmann and Jeremy Gwinnup*
- MT Tuning on RED: A Dependency-Based Evaluation Metric  
*Liangyou Li, Hui Yu, and Qun Liu*
- Improving evaluation and optimization of MT systems against MEANT  
*Chi-ku Lo, Philipp Dowling, and Dekai Wu*

**12:30–14:00 Lunch****14:00–15:20 Session 9: Evaluation and System Combination**

- An Investigation of Machine Translation Evaluation Metrics in Cross-lingual Question Answering  
*Kyoshiro Sugiyama, Masahiro Mizukami, Graham Neubig, Koichiro Yoshino, Sakriani Sakti, Tomoki Toda, and Satoshi Nakamura*
- Dependency Analysis of Scrambled References for Better Evaluation of Japanese Translation  
*Hideki Isozaki and Natsume Kouchi*
- How do Humans Evaluate Machine Translation  
*Francisco Guzmán, Ahmed Abdelali, Irina Temnikova, Hassan Sajjad, and Stephan Vogel*
- Local System Voting Feature for Machine Translation System Combination  
*Markus Freitag, Jan-Thorsten Peter, Stephan Peitz, Minwei Feng, and Hermann Ney*

**15:20–16:00 Coffee Break****16:00–17:00 Session 10: Closing Session and Open Discussion**

## Workshop 2: WASSA

### 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis

---

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

Venue: Room 1

**Thursday, September 17, 2015**

09:00–09:05 **Opening Remarks**

09:05–09:40 **Invited talk**

- Multilingual Affect Polarity and Valence Prediction in Metaphors  
*Zornitsa Kozareva*

09:40–10:30 **Session 1: Multilingual Sentiment Analysis in Social Media**

09:40–10:10 Sentiment Analysis on Monolingual, Multilingual and Code-Switching Twitter Corpora

*David Vilares, Miguel A. Alonso, and Carlos Gómez-Rodríguez*

10:10–10:30 Connotation in Translation

*Marine Carpuat*

10:30–11:00 **Coffee Break**

11:00–12:30 **Session 2: The Influence of Context for Sentiment Analysis in Social Media**

11:00–11:30 Enhanced Twitter Sentiment Classification Using Contextual Information  
*Soroush Vosoughi and Helen Zhou*

11:30–12:00 Your Sentiment Precedes You: Using an author's historical tweets to predict sarcasm

*Anupam Khattri, Aditya Joshi, Pushpak Bhattacharyya, and Mark Carman*

12:00–12:30 Optimising Agile Social Media Analysis

*Thomas Kober and David Weir*

12:30–14:00 **Lunch Break**

14:00–15:30 **Session 3: Beyond Review Mining**

14:00–14:30 Utilizing review analysis to suggest product advertisement improvements  
*Takaaki Tsunoda, Takashi Inui, and Satoshi Sekine*

14:30–15:00 Towards Opinion Mining from Reviews for the Prediction of Product Rankings

*Wiltrud Kessler, Roman Klinger, and Jonas Kuhn*

15:00–15:30 Classification of deceptive opinions using a low dimensionality representation

*Leticia Cagnina and Paolo Rosso*

15:30–16:00 **Coffee Break**

16:00–17:20 **Session 4: Lexicon Generation and Visualisation for Sentiment Analysis**

16:00–16:30 Extending effect annotation with lexical decomposition  
*Josef Ruppenhofer and Jasper Brandes*

16:30–17:00 Analysing domain suitability of a sentiment lexicon by identifying distributionally bipolar words  
*Lucie Flekova, Daniel Preořiuc-Pietro, and Eugen Ruppert*

17:00–17:20 Imagisaurus: An Interactive Visualizer of Valence and Emotion in the Roget's Thesaurus  
*Saif Mohammad*

17:20–17:30 **Break**

17:30–19:20 **Session 5: Posters**

- Personality Traits on Twitter—or—How to Get 1,500 Personality Tests in a Week  
*Barbara Plank and Dirk Hovy*
- Negation Scope Detection for Twitter Sentiment Analysis  
*Johan Reitan, Jørgen Faret, Björn Gambäck, and Lars Bungum*
- A Linguistically Informed Convolutional Neural Network  
*Sebastian Ebert, Ngoc Thang Vu, and Hinrich Schütze*
- How much does word sense disambiguation help in sentiment analysis of micropost data?  
*Chiraag Sumanth and Diana Inkpen*
- Predicting Ratings for New Movie Releases from Twitter Content  
*Wernard Schmit and Sander Wubben*
- Beyond Sentiment: Social Psychological Analysis of Political Facebook Comments in Hungary  
*Márton Miháltz, Tamás Váradi, István Csértő, Éva Fülöp, Tibor Pólya, and Pál Kővágó*
- Verb-centered Sentiment Inference with Description Logics  
*Manfred Klenner*
- Mining HEXACO personality traits from Enterprise Social Media  
*Priyanka Sinha, Lipika Dey, Pabitra Mitra, and Anupam Basu*
- Opinion Holder and Target Extraction for Verb-based Opinion Predicates – The Problem is Not Solved  
*Michael Wiegand, Marc Schulder, and Josef Ruppenhofer*
- Synthetic Text Generation for Sentiment Analysis  
*Umar Maqsud*
- Detecting speculations, contrasts and conditionals in consumer reviews  
*Maria Skeppstedt, Teri Schamp-Bjerede, Magnus Sahlgren, Carita Paradis, and Andreas Kerren*
- Using Combined Lexical Resources to Identify Hashtag Types  
*Credell Simeon and Robert Hilderman*
- Sentiment Classification via a Response Recalibration Framework  
*Phillip Smith and Mark Lee*

19:20–19:30 **Closing discussion - "Where do we go from here?"**

## Workshop 3: DiscoMT

### Second Workshop on Discourse in Machine Translation

Organizers: *Bonnie Webber, Marine Carpuat, and Andrei Popescu-Belis*

Venue: Room 6

**Thursday, September 17, 2015**

09:00–10:30 **Session 1**

09:00–09:05 *Introduction (Bonnie Webber)*

09:05–09:35 Pronoun-Focused MT and Cross-Lingual Pronoun Prediction: Findings of the 2015 DiscoMT Shared Task on Pronoun Translation  
*Christian Hardmeier, Preslav Nakov, Sara Stymne, Jörg Tiedemann, Yannick Versley, and Mauro Cettolo*

09:35–09:50 Comparison of Coreference Resolvers for Deep Syntax Translation  
*Michał Novák, Dieke Oele, and Gertjan van Noord*

09:50–10:15 Analysing ParCor and its Translations by State-of-the-art SMT Systems  
*Liane Guillou and Bonnie Webber*

10:15–10:30 **Poster Boaster**

10:30–11:00 **Coffee Break**

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

- Document-Level Machine Translation Evaluation with Gist Consistency and Text Cohesion  
*Zhengxian Gong, Min Zhang, and Guodong Zhou*
- The Role of Expectedness in the Implication and Explicitation of Discourse Relations  
*Jet Hoek, Jacqueline Evers-Vermeul, and Ted J.M. Sanders*
- Detecting Document-level Context Triggers to Resolve Translation Ambiguity  
*Laura Mascarell, Mark Fishel, and Martin Volk*
- A Proposal for a Coherence Corpus in Machine Translation  
*Karin Sim Smith, Wilker Aziz, and Lucia Specia*

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

- *On Statistical Machine Translation and Translation Theory (Christian Hardmeier)*
- *Exploration of Inter- and Intralingual Variation of Discourse Phenomena (Ekaterina Lapshinova-Koltunski)*
- *Measuring 'Registerness' in Human and Machine Translation: A Text Classification Approach (Ekaterina Lapshinova-Koltunski and Mihaela Vela)*
- *Translation Model Adaptation Using Genre-Revealing Text Features (Marlies van der Wees, Arianna Bisazza, Christof Monz)*

- *Crosslingual Annotation and Analysis of Implicit Discourse Connectives for Machine Translation (Frances Yung, Kevin Duh, Yuji Matsumoto)*

### 11:00–12:30 Session 2c: Shared Task Posters

- Part-of-Speech Driven Cross-Lingual Pronoun Prediction with Feed-Forward Neural Networks  
*Jimmy Callin, Christian Hardmeier, and Jörg Tiedemann*
- Automatic Post-Editing for the DiscoMT Pronoun Translation Task  
*Liane Guillou*
- A Document-Level SMT System with Integrated Pronoun Prediction  
*Christian Hardmeier*
- Predicting Pronoun Translation Using Syntactic, Morphological and Contextual Features from Parallel Data  
*Sharid Loáiciga*
- Rule-Based Pronominal Anaphora Treatment for Machine Translation  
*Sharid Loáiciga and Eric Wehrli*
- Pronoun Translation and Prediction with or without Coreference Links  
*Ngoc Quang Luong, Lesly Miculicich Werlen, and Andrei Popescu-Belis*
- Predicting Pronouns across Languages with Continuous Word Spaces  
*Ngoc-Quan Pham and Lonneke van der Plas*
- Baseline Models for Pronoun Prediction and Pronoun-Aware Translation  
*Jörg Tiedemann*
- A Maximum Entropy Classifier for Cross-Lingual Pronoun Prediction  
*Dominikus Wetzel, Adam Lopez, and Bonnie Webber*

### 12:30–14:00 Lunch Break

### 14:00–15:30 Session 3

- 14:00–14:25 Measuring ‘Registriness’ in Human and Machine Translation: A Text Classification Approach  
*Ekaterina Lapshinova-Koltunski and Mihaela Vela*
- 14:25–14:50 Translation Model Adaptation Using Genre-Revealing Text Features  
*Marlies van der Wees, Arianna Bisazza, and Christof Monz*
- 14:50–15:15 Crosslingual Annotation and Analysis of Implicit Discourse Connectives for Machine Translation  
*Frances Yung, Kevin Duh, and Yuji Matsumoto*
- 15:15–15:30 Novel Document Level Features for Statistical Machine Translation  
*Rong Zhang and Abraham Ittycheriah*

### 15:30–16:00 Coffee Break

### 16:00–17:30 Session 4

- 16:00–16:25 Exploration of Inter- and Intralingual Variation of Discourse Phenomena  
*Ekaterina Lapshinova-Koltunski*
- 16:25–16:40 On Statistical Machine Translation and Translation Theory  
*Christian Hardmeier*
- 16:40–17:30 Final Discussions and Conclusions (Workshop Organizers and Audience)

## Workshop 4: LOUHI

### Sixth Workshop on Health Text Mining and Information Analysis

---

Organizers: Cyril Grouin, Thierry Hamon, Aurélie Névéol, and Pierre Zweigenbaum

Venue: Room 3

**Thursday, September 17, 2015**

#### 09:00–10:30 Session I - Corpus creation

- In-depth annotation for patient level liver cancer staging  
*Wen-wai Yim, Sharon Kwan, and Meliha Yetisgen*
- Predicting Continued Participation in Online Health Forums  
*Farig Sadeque, Thamar Solorio, Ted Pedersen, Prasha Shrestha, and Steven Bethard*
- Redundancy in French Electronic Health Records: A preliminary study  
*Eva D'hondt, Xavier Tannier, and Aurélie Névéol*
- Is it possible to recover personal health information from an automatically de-identified corpus of French EHRs?  
*Cyril Grouin, Nicolas Griffon, and Aurélie Névéol*

#### 11:00–12:30 Session II - Poster

- An Analysis of Biomedical Tokenization: Problems and Strategies  
*Noa P. Cruz Diaz and Manuel Maña López*
- Annotation of Clinically Important Follow-up Recommendations in Radiology Reports  
*Meliha Yetisgen, Prescott Klassen, Lucas McCarthy, Elena Pellicer, Tom Payne, and Martin Gunn*
- On the Impact of Twitter-based Health Campaigns: A Cross-Country Analysis of Movember  
*Nugroho Dwi Prasetyo, Claudia Hauff, Dong Nguyen, Tijs van den Broek, and Djoerd Hiemstra*
- Exploring Word Embedding for Drug Name Recognition  
*Isabel Segura-Bedmar, Víctor Suárez-Paniagua, and Paloma Martínez*
- Creating a rule based system for text mining of Norwegian breast cancer pathology reports  
*Rebecka Weegar and Hercules Dalianis*
- Parser Adaptation to the Biomedical Domain without Re-Training  
*Jeff Mitchell and Mark Steedman*
- Expanding a dictionary of marker words for uncertainty and negation using distributional semantics  
*Alyaa Alfalahi, Maria Skeppstedt, Rickard Ahlbom, Roza Baskalayci, Aron Henriksson, Lars Asker, Carita Paradis, and Andreas Kerren*

- Held-out versus Gold Standard: Comparison of Evaluation Strategies for Distantly Supervised Relation Extraction from Medline abstracts  
*Roland Roller and Mark Stevenson*
- Checking a structured pathology report for completeness of content using terminological knowledge  
*Sebastian Busse*
- Effectively Crowdsourcing Radiology Report Annotations  
*Anne Cocos, Aaron Masino, Ting Qian, Ellie Pavlick, and Chris Callison-Burch*
- Identifying Key Concepts from EHR Notes Using Domain Adaptation  
*Jiapeng Zheng and Hong Yu*

12:30–14:00 **Lunch break**

14:00–15:30 **Session III - Invited talk**

- Information Extraction from Biomedical Texts: Learning Models with Limited Supervision  
*Marie-Francine Moens*

16:00–17:30 **Session IV - Corpus processing**

- Adverse Drug Event classification of health records using dictionary based pre-processing and machine learning  
*Stefanie Friedrich and Hercules Dalianis*
- NLP—Based Readability Assessment of Health—Related Texts: a Case Study on Italian Informed Consent Forms  
*Giulia Venturi, Tommaso Bellandi, Felice Dell'Orletta, and Simonetta Montemagni*
- Mining and Ranking Biomedical Synonym Candidates from Wikipedia  
*Abhyuday Jagannatha, Jinying Chen, and Hong Yu*
- Representing Clinical Notes for Adverse Drug Event Detection  
*Aron Henriksson*

## Workshop 5: LSDSem

### First Workshop on Linking Models of Lexical, Sentential and Discourse-level Semantics

---

Organizers: *Michael Roth, Annie Louis, Bonnie Webber, and Tim Baldwin*

Venue: Room 1

**Friday, September 18, 2015**

**09:00–10:30 Morning Session**

09:00–09:05 *Introduction (Michael Roth)*

09:05–09:50 *Invited Talk: From Distributed Semantics to Discourse, and Back (Jacob Eisenstein)*

09:50–10:05 An Exploration of Discourse-Based Sentence Spaces for Compositional  
Distributional Semantics  
*Tamara Polajnar, Laura Rimell, and Stephen Clark*

10:05–10:20 Linking discourse modes and situation entity types in a cross-linguistic  
corpus study  
*Kleio-Isidora Mavridou, Annemarie Friedrich, Melissa Peate Sørensen,  
Alexis Palmer, and Manfred Pinkal*

10:20–10:30 Recovering discourse relations: Varying influence of discourse adverbials  
*Hannah Rohde, Anna Dickinson, Chris Clark, Annie Louis, and  
Bonnie Webber*

**10:30–11:00 Coffee Break**

**11:00–12:30 Pre-Lunch Session**

11:00–11:15 Semantics and Discourse Processing for Expressive TTS  
*Rodolfo Delmonte and Rocco Tripodi*

11:15–11:30 Semantically Enriched Models for Modal Sense Classification  
*Mengfei Zhou, Anette Frank, Annemarie Friedrich, and Alexis Palmer*

11:30–11:45 Identification and Disambiguation of Lexical Cues of Rhetorical Relations  
across Different Text Genres  
*Taraneh Khazaei, Lu Xiao, and Robert Mercer*

11:45–11:55 Bridging Sentential and Discourse-level Semantics through Clausal Adjuncts  
*Rashmi Prasad, Bonnie Webber, Alan Lee, Sameer Pradhan, and  
Aravind Joshi*

11:55–12:05 Lexical Level Distribution of Metadiscourse in Spoken Language  
*Rui Correia, Maxine Eskenazi, and Nuno Mamede*

12:05–12:15 Idiom Paraphrases: Seventh Heaven vs Cloud Nine  
*Maria Pershina, Yifan He, and Ralph Grishman*

12:15–12:25 Where Was Alexander the Great in 325 BC? Toward Understanding History  
Text with a World Model  
*Yuki Murakami and Yoshimasa Tsuruoka*

12:30–14:00 **Lunch Break**

14:00–15:30 **Post-Lunch Session**

14:00–14:05 *TextLink: EU COST action on Structuring Discourse in a Multi-Lingual Europe* (Bonnie Webber)

14:05–14:50 *Invited Talk: What Men Say, What Women Hear: Using Semantics To Make Better Sense of Gender Differences in Social Media* (Rada Mihalcea)

14:50–15:00 Predicting word sense annotation agreement  
Héctor Martínez Alonso, Anders Johannsen, Oier Lopez de Lacalle, and Eneko Agirre

15:00–15:10 Distributional Semantics in Use  
Raffaella Bernardi, Gemma Boleda, Raquel Fernandez, and Denis Paperno

15:10–16:00 **Poster Session (Including Coffee Break)**

15:30–16:00 **Coffee Break**

16:00–17:30 **Afternoon Session**

16:00–16:45 *Invited Talk: The (Non)Utility of Semantics for Coreference Resolution* (Michael Strube)

16:45–17:30 *Panel Discussion* (Various speakers)

## **Workshop 6: CogACLL**

### **Sixth Workshop on Cognitive Aspects of Computational Language Learning**

---

Organizers: *Robert Berwick, Anna Korhonen, Alessandro Lenci, Thierry Poibeau, and Aline Villavicencio*

Venue: Room 3

**Friday, September 18, 2015**

09:00–09:10 **Opening and Introduction**

09:10–10:30 **Session 1: Language Processing**

09:10–09:30 Using reading behavior to predict grammatical functions  
*Maria Barrett and Anders Søgaard*

09:30–10:00 Reading metrics for estimating task efficiency with MT output  
*Sigrid Klerke, Sheila Castilho, Maria Barrett, and Anders Søgaard*

10:00–10:30 Evaluating Models of Computation and Storage in Human Sentence Processing  
*Thang Luong, Timothy O'Donnell, and Noah Goodman*

10:30–11:00 **Coffee Break**

11:00–11:50 **Invited Talk by Afra Alishahi**

11:50–12:10 **Session 2: Language Change**

11:50–12:10 An agent-based model of a historical word order change  
*Jelke Bloem, Arjen Versloot, and Fred Weerman*

12:10–13:00 **Session 3: Poster Session**

12:10–13:00 Towards a Model of Prediction-based Syntactic Category Acquisition: First Steps with Word Embeddings  
*Robert Grimm, Giovanni Cassani, Walter Daelemans, and Steven Gillis*

12:10–12:30 Which distributional cues help the most? Unsupervised contexts selection for lexical category acquisition  
*Giovanni Cassani, Robert Grimm, Walter Daelemans, and Steven Gillis*

12:10–13:00 Language Emergence in a Population of Artificial Agents Equipped with the Autotelic Principle  
*Miquel Cornudella and Thierry Poibeau*

12:10–13:00 A Computational Study of Cross-situational Lexical Learning of Brazilian Portuguese  
*Pablo Faria*

12:10–13:00 Units in segmentation: a computational investigation  
*Çağrı Çöltekin*

12:10–13:00 Estimating Grammeme Redundancy by Measuring Their Importance for Syntactic Parser Performance  
*Aleksandrs Berdicevskis*

13:00–14:10 **Lunch**

14:10–15:00 **Invited Talk by Antal van den Bosch**

15:00–15:30 **Session 4: Language Processing II**

15:00–15:30 Improving Coordination on Novel Meaning through Context and Semantic Structure

*Thomas Brochhagen*

15:30–16:00 **Coffee Break**

16:00–17:30 **Session 5: Language Acquisition**

16:00–16:30 Perceptual, conceptual, and frequency effects on error patterns in English color term acquisition

*Barend Beekhuizen and Suzanne Stevenson*

16:30–17:00 Motif discovery in infant- and adult-directed speech

*Bogdan Ludusan, Amanda Seidl, Emmanuel Dupoux, and Alex Cristia*

17:00–17:30 Modeling dative alternations of individual children

*Antal van den Bosch and Joan Bresnan*

17:30–17:35 **Closing Session**

## Workshop 7: VL

### Fourth Workshop on Vision and Language

---

Organizers: Anja Belz, Luisa Coheur, Vittorio Ferrari, Marie-Francine Moens,  
Katerina Pastra, and Ivan Vulic

Venue: Room 6

**Friday, September 18, 2015**

08:45–10:30 **Session 1**

08:45–09:00 *Opening Remarks (Marie-Francine Moens)*

09:00–10:00 *Invited Talk 1: Grounding Distributional Semantics in the Visual World (Marco Baroni)*

10:00–10:30 **Poster Spotlights 1**

10:00–10:05 Visually-Verifiable Textual Entailment: A Challenge Task for Combining  
Language and Vision  
*Jayant Krishnamurthy*

10:05–10:10 Computational Integration of Human Vision and Natural Language through  
Bitext Alignment  
*Preethi Vaidyanathan, Emily Prud'hommeaux, Cecilia O. Alm, Jeff B. Pelz,  
and Anne R. Haake*

10:10–10:15 Towards Reliable Automatic Multimodal Content Analysis  
*Olli Philippe Lautenbacher, Liisa Tiittula, Maija Hirvonen, Jorma Laaksonen,  
and Mikko Kurimo*

10:15–10:20 Lingusitic Analysis of Multi-Modal Recurrent Neural Networks  
*Ákos Kádár, Grzegorz Chrupała, and Afra Alishahi*

10:20–10:25 Defining Visually Descriptive Language  
*Robert Gaizauskas, Josiah Wang, and Arnaud Ramisa*

10:25–10:30 Semantic Tuples for Evaluation of Image to Sentence Generation  
*Lily D. Ellebracht, Arnaud Ramisa, Pranava Swaroop Madhyastha,  
Jose Cordero-Rama, Francesc Moreno-Noguer, and Ariadna Quattoni*

10:30–11:00 **Coffee Break**

11:00–12:15 **Session 2**

11:00–11:25 Image Representations and New Domains in Neural Image Captioning  
*Jack Hessel, Nicolas Savva, and Michael Wilber*

11:25–11:50 Image with a Message: Towards Detecting Non-Literal Image Usages by  
Visual Linking  
*Lydia Weiland, Laura Dietz, and Simone Paolo Ponzetto*

11:50–12:15 **Poster Spotlights 2**

11:50–11:55 Visual Classifier Prediction by Distributional Semantic Embedding of Text  
Descriptions  
*Mohamed Elhoseiny and Ahmed Elgammal*

- 11:55–12:00 Understanding Urban Land Use through the Visualization of Points of Interest  
*Evgheni Polisciuc, Ana Alves, and Penousal Machado*
- 12:00–12:05 Comparing Attribute Classifiers for Interactive Language Grounding  
*Yanchao Yu, Arash Eshghi, and Oliver Lemon*
- 12:05–12:10 *Combining Geometric, Textual and Visual Features for Generating Prepositions in Image Descriptions (to appear in EMNLP 2015)* (Arnau Ramisa, Josiah Wang, Ying Lu, Emmanuel Dellandrea, Francesc Moreno-Noguer and Robert Gaizauskas)
- 12:10–12:15 *From the Virtual to the Real World: Referring to Visible Objects Under Uncertainty in Real-World Spatial Scenes (to appear in EMNLP 2015)* (Dimitra Gkatzia and Verena Rieser)
- 12:15–14:00 Poster Session and Lunch**
- 14:00–15:30 Session 3**
- 14:00–15:00 *Invited Talk 2: The ImageCLEF 2015 Task on Scalable Image Annotation, Localization and Sentence Generation (Krystian Mikolajczyk)*
- 15:00–15:30 Generating Semantically Precise Scene Graphs from Textual Descriptions for Improved Image Retrieval  
*Sebastian Schuster, Ranjay Krishna, Angel Chang, Li Fei-Fei, and Christopher D. Manning*
- 15:30–16:00 Coffee Break**
- 16:00–18:00 Session 4**
- 16:00–16:25 Do Distributed Semantic Models Dream of Electric Sheep? Visualizing Word Representations through Image Synthesis  
*Angeliki Lazaridou, Dat Tien Nguyen, and Marco Baroni*
- 16:25–16:50 A Weighted Combination of Text and Image Classifiers for User Gender Inference  
*Tomoki Taniguchi, Shigeyuki Sakaki, Ryosuke Shigenaka, Yukihiko Tsuboshita, and Tomoko Ohkuma*
- 16:50–17:15 Coupling Natural Language Processing and Animation Synthesis in Portuguese Sign Language Translation  
*Inês Almeida, Luísa Coheur, and Sara Candeias*
- 17:15–17:40 Describing Spatial Relationships between Objects in Images in English and French  
*Anja Belz, Adrian Muscat, Maxime Aberton, and Sami Benjelloun*
- 17:40–18:00 Closing Remarks and Discussion**

## Main Conference: Saturday, September 19

### Overview

---

07:30–18:00 **Registration** Culturgest Foyers

08:00–08:40 **Morning Coffee**

08:40–10:00 **Session P1: Plenary Session**

Main Auditorium

08:40–09:00 *Opening Remarks and Introductory Speeches (General Chair, Program Co-Chairs and Local Co-Chairs)*

09:00–10:00 *Invited Talk: Deep Learning of Semantic Representations (Yoshua Bengio)*

10:00–10:30 **Coffee break**

Foyers

10:30–12:10 **Session 1**

Semantics (Long + TACL Papers)  Main Auditorium	Machine Translation (Long + TACL Papers)  Small Auditorium	NLP for the Web and Social Media, including Computational Social Science (Long Papers)  Room 2	Long Paper Posters  Short Paper Posters  Lower Level Foyer
---	--	--	--

12:10–13:30 **Lunch**

13:30–15:10 **Session 2**

Statistical Models and Machine Learning Methods (Long + TACL Papers)  Main Auditorium	Tagging, Chunking and Parsing (Long +TACL Papers)  Small Auditorium	Summarization (Long Papers)  Room 2	Long Paper Posters  Short Paper Posters  Lower Level Foyer
---	---	---	--

15:10–15:40 **Coffee break**

Foyers

15:40–17:20 **Session 3**

Sentiment Analysis and Opinion Mining (Long Papers)  Main Auditorium	Semantics (Long +TACL Papers)  Small Auditorium	Information Retrieval and Question Answering (Long Papers)  Room 2	Long + TACL Paper Posters  Short Paper Posters  Lower Level Foyer
--	---	--	---

---

## Invited Speaker: Yoshua Bengio

---

### Deep Learning of Semantic Representations

Saturday, September 19, 2015,

Main Auditorium

**Abstract:** The core ingredient of deep learning is the notion of distributed representation. This talk will start by explaining its theoretical advantages, in comparison with non-parametric methods based on counting frequencies of occurrence of observed tuples of values (like with n-grams). The talk will then explain how having multiple levels of representation, i.e., depth, can in principle give another exponential advantage. Neural language models have been extremely successful in recent years but extending their reach from language modeling to machine translation is very appealing because it forces the learned intermediate representations to capture meaning, and we found that the resulting word embeddings are qualitatively different. Recently, we introduced the notion of attention-based encoder-decoder systems, with impressive results on machine translation several language pairs and for mapping an image to a sentence, and these results will conclude the talk.

**Biography:** Yoshua Bengio received a PhD in Computer Science from McGill University, Canada in 1991. After two post-doctoral years, one at M.I.T. with Michael Jordan and one at AT&T Bell Laboratories with Yann LeCun and Vladimir Vapnik, he became professor at the Department of Computer Science and Operations Research at Université de Montréal. He is the author of two books and more than 200 publications, the most cited being in the areas of deep learning, recurrent neural networks, probabilistic learning algorithms, natural language processing and manifold learning. He is among the most cited Canadian computer scientists and is or has been associate editor of the top journals in machine learning and neural networks. Since '2000 he holds a Canada Research Chair in Statistical Learning Algorithms, since '2006 an NSERC Industrial Chair, since '2005 his is a Senior Fellow of the Canadian Institute for Advanced Research and since 2014 he co-directs its program focused on deep learning. He is on the board of the NIPS foundation and has been program chair and general chair for NIPS. He has co-organized the Learning Workshop for 14 years and co-created the new International Conference on Learning Representations. His current interests are centered around a quest for AI through machine learning, and include fundamental questions on deep learning and representation learning, the geometry of generalization in high-dimensional spaces, manifold learning, biologically inspired learning algorithms, and challenging applications of statistical machine learning.

## Session 1 Overview

---

Track A	Track B	Track C	
Semantics (Long + TACL Papers) Main Auditorium	Machine Translation (Long + TACL Papers) Small Auditorium	NLP for the Web and Social Media, including Computational Social Science (Long Papers) Room 2	
Language Understanding for Text-based Games using Deep Reinforcement Learning <i>Narasimhan, Kulkarni, and Barzilay</i>	Dependency Graph-to-String Translation <i>Li, Way, and Liu</i>	Identifying Political Sentiment between Nation States with Social Media <i>Chambers, Bowen, Genco, Tian, Young, Harihara, and Yang</i>	10:30
Distributional vectors encode referential attributes <i>Gupta, Boleda, Baroni, and Padó</i>	Reordering Grammar Induction <i>Stanojević and Sima'an</i>	Open Extraction of Fine-Grained Political Statements <i>Bamman and Smith</i>	10:55
Building a shared world: mapping distributional to model-theoretic semantic spaces <i>Herbelot and Vecchi</i>	Syntax-based Rewriting for Simultaneous Machine Translation <i>He, Grissom II, Morgan, Boyd-Graber, and Daumé III</i>	Using Personal Traits For Brand Preference Prediction <i>Yang, Pan, Mahmud, Yang, and Srinivasan</i>	11:20
[TACL] Deriving Boolean Structures from Distributional Vectors <i>Kruszewski, Paperno, and Baroni</i>	[TACL] Modelling and Optimizing on Syntactic N-Grams for Statistical Machine Translation <i>Sennrich</i>	Semantic Annotation for Microblog Topics Using Wikipedia Temporal Information <i>Tran, Tran, Teka Hadgu, and Jäschke</i>	11:45

**Track D: Long Paper Posters**

Lower Level Foyer

Chair: Saif Mohammad

10:30–12:10

**Poster Cluster 1: Summarization (P1-4)**

- System Combination for Multi-document Summarization  
*K. Hong, M. Marcus, and A. Nenkova*
- Phrase-based Compressive Cross-Language Summarization  
*J.-g. Yao, X. Wan, and J. Xiao*
- Re-evaluating Automatic Summarization with BLEU and 192 Shades of ROUGE  
*Y. Graham*
- Indicative Tweet Generation: An Extractive Summarization Problem?  
*P. Sidhaye and J. C. K. Cheung*

**Poster Cluster 2: Language and Vision (P5)**

- Visual Bilingual Lexicon Induction with Transferred ConvNet Features  
*D. Kiela, I. Vulić, and S. Clark*

**Poster Cluster 3: Sentiment Analysis and Opinion Mining (P6-9)**

- Cross Lingual Sentiment Analysis using Modified BRAE  
*S. Jain and S. Batra*
- Monotone Submodularity in Opinion Summaries  
*J. Jayanth, J. Sundararaj, and P. Bhattacharyya*
- Joint Prediction for Entity/Event-Level Sentiment Analysis using Probabilistic Soft Logic Models  
*L. Deng and J. Wiebe*
- Learning to Recognize Affective Polarity in Similes  
*A. Qadir, E. Riloff, and M. Walker*

**Track E: Short Paper Posters**

Chair: Kevin Duh

Lower Level Foyer

10:30-12:10

**Poster Cluster 1: Language and Vision (P1-3)**

- Cross-document Event Coreference Resolution based on Cross-media Features  
*T. Zhang, H. Li, H. Ji, and S.-F. Chang*
- A Survey of Current Datasets for Vision and Language Research  
*F. Ferraro, N. Mostafazadeh, T.-H. Huang, L. Vanderwende, J. Devlin, M. Galley, and M. Mitchell*
- Combining Geometric, Textual and Visual Features for Predicting Prepositions in Image Descriptions  
*A. Ramisa, J. Wang, Y. Lu, E. Dellandrea, F. Moreno-Noguer, and R. Gaizauskas*

**Poster Cluster 2: Statistical Models and Machine Learning Methods (P4-16)**

- On A Strictly Convex IBM Model 1  
*A. Simion, M. Collins, and C. Stein*
- Factorization of Latent Variables in Distributional Semantic Models  
*A. Österlund, D. Ödling, and M. Sahlgren*
- Non-lexical neural architecture for fine-grained POS Tagging  
*M. Labeau, K. Löser, and A. Aliauzen*
- Online Representation Learning in Recurrent Neural Language Models  
*M. Rei*
- A Model of Zero-Shot Learning of Spoken Language Understanding  
*M. Yazdani and J. Henderson*
- Modeling Tweet Arrival Times using Log-Gaussian Cox Processes  
*M. Lukasik, P. K. SriJith, T. Cohn, and K. Bontcheva*

- Pre-Computable Multi-Layer Neural Network Language Models  
*J. Devlin, C. Quirk, and A. Menezes*
- Birds of a Feather Linked Together: A Discriminative Topic Model using Link-based Priors  
*W. Yang, J. Boyd-Graber, and P. Resnik*
- Aligning Knowledge and Text Embeddings by Entity Descriptions  
*H. Zhong, J. Zhang, Z. Wang, H. Wan, and Z. Chen*
- An Empirical Analysis of Optimization for Max-Margin NLP  
*J. K. Kummerfeld, T. Berg-Kirkpatrick, and D. Klein*
- Learning Better Embeddings for Rare Words Using Distributional Representations  
*I. Sergienya and H. Schütze*
- Composing Relationships with Translations  
*A. Garcia-Duran, A. Bordes, and N. Usunier*
- Noise or additional information? Leveraging crowdsource annotation item agreement for natural language tasks.  
*E. Jamison and I. Gurevych*

## Abstracts: Session 1

---

### Session 1A: Semantics (Long + TACL Papers)

Main Auditorium

Chair: Yoav Artzi

#### Language Understanding for Text-based Games using Deep Reinforcement Learning

*Karthik Narasimhan, Tejas Kulkarni, and Regina Barzilay*

10:30–10:55

In this paper, we consider the task of learning control policies for text-based games. In these games, all interactions in the virtual world are through text and the underlying state is not observed. The resulting language barrier makes such environments challenging for automatic game players. We employ a deep reinforcement learning framework to jointly learn state representations and action policies using game rewards as feedback. This framework enables us to map text descriptions into vector representations that capture the semantics of the game states. We evaluate our approach on two game worlds, comparing against baselines using bag-of-words and bag-of-bigrams for state representations. Our algorithm outperforms the baselines on both worlds demonstrating the importance of learning expressive representations.

#### Distributional vectors encode referential attributes

*Abhijeet Gupta, Gemma Boleda, Marco Baroni, and Sebastian Padó*

10:55–11:20

Distributional methods have proven to excel at capturing fuzzy, graded aspects of meaning (Italy is more similar to Spain than to Germany). In contrast, it is difficult to extract the values of more specific attributes of word referents from distributional representations, attributes of the kind typically found in structured knowledge bases (Italy has 60 million inhabitants). In this paper, we pursue the hypothesis that distributional vectors also implicitly encode referential attributes. We show that a standard supervised regression model is in fact sufficient to retrieve such attributes to a reasonable degree of accuracy. When evaluated on the prediction of both categorical and numeric attributes of countries and cities, the model consistently reduces baseline error by 30%, and is not far from the upper bound. Further analysis suggests that our model is able to “objectify” distributional representations for entities, anchoring them more firmly in the external world in measurable ways.

#### Building a shared world: mapping distributional to model-theoretic semantic spaces

*Aurélie Herbelot and Eva Maria Vecchi*

11:20–11:45

In this paper, we introduce an approach to automatically map a standard distributional semantic space onto a set-theoretic model. We predict that there is a functional relationship between distributional information and vectorial concept representations in which dimensions are predicates and weights are generalised quantifiers. In order to test our prediction, we learn a model of such relationship over a publicly available dataset of feature norms annotated with natural language quantifiers. Our initial experimental results show that, at least for domain-specific data, we can indeed map between formalisms, and generate high-quality vector representations which encapsulate set overlap information. We further investigate the generation of natural language quantifiers from such vectors.

#### [TACL] Deriving Boolean Structures from Distributional Vectors

*Germán Kruszewski, Denis Paperno, and Marco Baroni*

11:45–12:10

Corpus-based distributional semantic models capture degrees of semantic relatedness among the words of very large vocabularies, but have problems with logical phenomena such as entailment, that are instead elegantly handled by model-theoretic approaches, which, in turn, do not scale up. We combine the advantages of the two views by inducing a mapping from distributional vectors of words (or sentences) into a Boolean structure of the kind in which natural language terms are assumed to denote. We evaluate this Boolean Distributional Semantic Model (BDSM) on recognizing entailment between words and sentences. The method achieves results comparable to a state-of-the art SVM, degrades more gracefully when less training data are available and displays interesting qualitative properties.

**Session 1B: Machine Translation (Long + TACL Papers)**

Small Auditorium

Chair: John DeNero

**Dependency Graph-to-String Translation**

Liangyou Li, Andy Way, and Qun Liu

10:30-10:55

Compared to tree grammars, graph grammars have stronger generative capacity over structures. Based on an edge replacement grammar, in this paper we propose to use a synchronous graph-to-string grammar for statistical machine translation. The graph we use is directly converted from a dependency tree by labelling edges. We build our translation model in the log-linear framework with standard features. Large-scale experiments on Chinese–English and German–English tasks show that our model is significantly better than the state-of-the-art hierarchical phrase-based (HPB) model and a recently improved dependency tree-to-string model on BLEU, METEOR and TER scores. Experiments also suggest that our model has better capability to perform long-distance reordering and is more suitable for translating long sentences.

**Reordering Grammar Induction**

Milos Stanojevic and Khalil Sima'an

10:55-11:20

We present a novel approach for unsupervised induction of a Reordering Grammar using a modified form of permutation trees (Zhang and Gildea, 2007), which we apply to preordering in phrase-based machine translation. Unlike previous approaches, we induce in one step both the hierarchical structure and the transduction function over it from word-aligned parallel corpora. Furthermore, our model (1) handles non-ITG reordering patterns (up to 5-ary branching), 2) is learned from all derivations by treating not only labeling but also bracketing as latent variable, (3) is entirely unlexicalized at the level of reordering rules, and (4) requires no linguistic annotation. Our model is evaluated both for accuracy in predicting target order, and for it's impact on translation quality. We report significant performance gains over phrase reordering, and over two known preordering baselines for English-Japanese.

**Syntax-based Rewriting for Simultaneous Machine Translation**

He He, Alvin Grissom II, John Morgan, Jordan Boyd-Graber, and Hal Daumé III 11:20-11:45

Divergent word order between languages causes delay in simultaneous machine translation. We present a sentence rewriting method that generates interpretation-like translations for better speed-accuracy tradeoff. We design grammaticality and meaning-preserving syntactic transformation rules operating on a constituent parse tree. We apply the rules to the reference translation, such that its word order is closer to the source language word order. On Japanese-English translation (two languages with substantially different structure), we show that incorporating the rewritten, more monotonic reference translation into the learning of a phrase-based machine translation system enables it to produce better translations faster than the baseline system that uses gold reference translation only.

**[TACL] Modelling and Optimizing on Syntactic N-Grams for Statistical Machine Translation**

Rico Sennrich

11:45-12:10

The role of language models in SMT is to promote fluent translation output, but traditional n-gram language models are unable to capture fluency phenomena between distant words, such as some morphological agreement phenomena, subcategorisation, and syntactic collocations with string-level gaps. Syntactic language models have the potential to fill this modelling gap. We propose a language model for dependency structures that is relational rather than configurational and thus particularly suited for languages with a (relatively) free word order. It is trainable with Neural Networks, and not only improves over standard n-gram language models, but also outperforms related syntactic language models. We empirically demonstrate its effectiveness in terms of perplexity and as a feature function in string-to-tree SMT from English to German and Russian. We also show that using a syntactic evaluation metric to tune the log-linear parameters of an SMT system further increases translation quality when coupled with a syntactic language model.

**Session 1C: NLP for the Web and Social Media, including Computational Social Science (Long Papers)**

Room 2

Chair: Brendan O'Connor

**Identifying Political Sentiment between Nation States with Social Media***Nathanael Chambers, Victor Bowen, Ethan Genco, Xisen Tian, Eric Young, Ganesh Harihara, and Eugene Yang* 10:30–10:55

This paper describes an approach to large-scale modeling of sentiment analysis for the social sciences. The goal is to model relations between nation states through social media. Many cross-disciplinary applications of NLP involve making predictions (such as predicting political elections), but this paper instead focuses on a model that is applicable to broader analysis. Do citizens express opinions in line with their home country's formal relations? When opinions diverge over time, what is the cause and can social media serve to detect these changes? We describe several learning algorithms to study how the populace of a country discusses foreign nations on Twitter, ranging from state-of-the-art contextual sentiment analysis to some required practical learners that filter irrelevant tweets. We evaluate on standard sentiment evaluations, but we also show strong correlations with two public opinion polls and current international alliance relationships. We conclude with some political science use cases.

**Open Extraction of Fine-Grained Political Statements***David Bamman and Noah A. Smith* 10:55–11:20

Text data has recently been used as evidence in estimating the political ideologies of individuals, including political elites and social media users. While inferences about people are often the intrinsic quantity of interest, we draw inspiration from open information extraction to identify a new task: inferring the political import of propositions like “Obama is a Socialist.” We present several models that exploit the structure that exists between people and the assertions they make to learn latent positions of people and propositions at the same time, and we evaluate them on a novel dataset of propositions judged on a political spectrum.

**Using Personal Traits For Brand Preference Prediction***Chao Yang, Shimei Pan, Jalal Mahmud, Huahai Yang, and Padmini Srinivasan* 11:20–11:45

In this paper, we present the first comprehensive study of the relationship between a person’s traits and his/her brand preferences. In our analysis, we included a large number of character traits such as personality, personal values and individual needs. These trait features were obtained from both a psychometric survey and automated social media analytics. We also included an extensive set of brand names from diverse product categories. From this analysis, we want to shed some light on (1) whether it is possible to use personal traits to infer an individual’s brand preferences (2) whether the trait features automatically inferred from social media are good proxies for the ground truth character traits in brand preference prediction.

**Semantic Annotation for Microblog Topics Using Wikipedia Temporal Information***Tuan Tran, Nam Khanh Tran, Asmelash Teka Hadgu, and Robert Jäschke* 11:45–12:10

Trending topics in microblogs such as Twitter are valuable resources to understand social aspects of real-world events. To enable deep analyses of such trends, semantic annotation is an effective approach; yet the problem of annotating microblog trending topics is largely unexplored by the research community. In this work, we tackle the problem of mapping trending Twitter topics to entities from Wikipedia. We propose a novel model that complements traditional text-based approaches by rewarding entities that exhibit a high temporal correlation with topics during their burst time period. By exploiting temporal information from the Wikipedia edit history and page view logs, we have improved the annotation performance by 17–28%, as compared to the competitive baselines.

**Session 1D: Long Paper Posters**

Lower Level Foyer

Chair: Saif Mohammad

10:30-12:10

**System Combination for Multi-document Summarization***Kai Hong, Mitchell Marcus, and Ani Nenkova*

We present a novel framework of system combination for multi-document summarization. For each input set (input), we generate candidate summaries by combining whole sentences from the summaries generated by different systems. We show that the oracle among these candidates is much better than the summaries that we have combined. We then present a supervised model to select among the candidates. The model relies on a rich set of features that capture content importance from different perspectives. Our model performs better than the systems that we combined based on manual and automatic evaluations. We also achieve very competitive performance on six DUC/TAC datasets, comparable to the state-of-the-art on most datasets.

**Phrase-based Compressive Cross-Language Summarization***Jin-ge Yao, Xiaojun Wan, and Jianguo Xiao*

The task of cross-language document summarization is to create a summary in a target language from documents in a different source language. Previous methods only involve direct extraction of automatically translated sentences from the original documents. Inspired by phrase-based machine translation, we propose a phrase-based model to simultaneously perform sentence scoring, extraction and compression. We design a greedy algorithm to approximately optimize the score function. Experimental results show that our methods outperform the state-of-the-art extractive systems while maintaining similar grammatical quality.

**Re-evaluating Automatic Summarization with BLEU and 192 Shades of ROUGE***Yvette Graham*

We provide an analysis of current evaluation methodologies applied to summarization metrics and identify the following areas of concern: (1) movement away from evaluation by correlation with human assessment; (2) omission of important components of human assessment from evaluations, in addition to large numbers of metric variants; (3) absence of methods of significance testing improvements over a baseline. We outline an evaluation methodology that overcomes all such challenges, providing the first method of significance testing appropriate for evaluation of summarization metrics. Our evaluation reveals for the first time which metric variants significantly outperform others, optimal metric variants distinct from current recommended best variants, as well as machine translation metric BLEU to have performance on-par with ROUGE for the purpose of evaluation of summarization systems. We subsequently replicate a recent large-scale evaluation that relied on, what we now know to be, suboptimal ROUGE variants revealing distinct conclusions about the relative performance of state-of-the-art summarization systems.

**Indicative Tweet Generation: An Extractive Summarization Problem?***Priya Sidhaye and Jackie Chi Kit Cheung*

Social media such as Twitter have become an important method of communication, with potential opportunities for NLG to facilitate the generation of social media content. We focus on the generation of indicative tweets that contain a link to an external web page. While it is natural and tempting to view the linked web page as the source text from which the tweet is generated in an extractive summarization setting, it is unclear to what extent actual indicative tweets behave like extractive summaries. We collect a corpus of indicative tweets with their associated articles and investigate to what extent they can be derived from the articles using extractive methods. We also consider the impact of the formality and genre of the article. Our results demonstrate the limits of viewing indicative tweet generation as extractive summarization, and point to the need for the development of a methodology for tweet generation that is sensitive to genre-specific issues.

**Visual Bilingual Lexicon Induction with Transferred ConvNet Features***Douwe Kiela, Ivan Vulić, and Stephen Clark*

This paper is concerned with the task of bilingual lexicon induction using image-based features. By applying features from a convolutional neural network (CNN), we obtain state-of-the-art performance on a standard dataset, obtaining a 79% relative improvement over previous work which uses bags of visual

words based on SIFT features. The CNN image-based approach is also compared with state-of-the-art linguistic approaches to bilingual lexicon induction, even outperforming these for one of three language pairs on another standard dataset. Furthermore, we shed new light on the type of visual similarity metric to use for genuine similarity versus relatedness tasks, and experiment with using multiple layers from the same network in an attempt to improve performance.

### Cross Lingual Sentiment Analysis using Modified BRAE

*Sarthak Jain and Shashank Batra*

Cross-Lingual Learning provides a mechanism to adapt NLP tools available for label rich languages to achieve similar tasks for label-scarce languages. An efficient cross-lingual tool significantly reduces the cost and effort required to manually annotate data. In this paper, we use the Recursive Autoencoder architecture to develop a Cross Lingual Sentiment Analysis (CLSA) tool using sentence aligned corpora between a pair of resource rich (English) and resource poor(Hindi) language. The system is based on the assumption that semantic similarity between different phrases also implies sentiment similarity in majority of sentences. The resulting system is then analyzed on a newly developed Movie Reviews Dataset in Hindi with labels given on a rating scale and compare performance of our system against existing systems. It is shown that our approach significantly outperforms state of the art systems for Sentiment Analysis, especially when labeled data is scarce.

### Monotone Submodularity in Opinion Summaries

*Jayanth Jayanth, Jayaprakash Sundararaj, and Pushpak Bhattacharyya*

Opinion summarization is the task of producing the summary of a text, such that the summary also preserves the sentiment of the text. Opinion Summarization is thus a trade-off between summarization and sentiment analysis. The demand of compression may drop sentiment bearing sentences, and the demand of sentiment detection may bring in redundant sentences. We harness the power of submodularity to strike a balance between two conflicting requirements. We investigate an incipient class of submodular functions for the problem, and a partial enumeration based greedy algorithm that has performance guarantee of 63%. Our functions generate summaries such that there is good correlation between document sentiment and summary sentiment along with good ROUGE score, which outperforms the-state-of-the-art algorithms.

### Joint Prediction for Entity/Event-Level Sentiment Analysis using Probabilistic Soft Logic Models

*Lingjia Deng and Janyce Wiebe*

In this work, we build an entity/event-level sentiment analysis system, which is able to recognize and infer both explicit and implicit sentiments toward entities and events in the text. We design Probabilistic Soft Logic models that integrate explicit sentiments, inference rules, and +/-effect event information (events that positively or negatively affect entities). The experiments show that the method is able to greatly improve over baseline accuracies in recognizing entity/event-level sentiments.

### Learning to Recognize Affective Polarity in Similes

*Ashequl Qadir, Ellen Riloff, and Marilyn Walker*

A simile is a comparison between two essentially unlike things, such as "Jane swims like a dolphin". Similes often express a positive or negative sentiment toward something, but recognizing the polarity of a simile can depend heavily on world knowledge. For example, "memory like an elephant" is positive, but "memory like a sieve" is negative. Our research explores methods to recognize the polarity of similes on Twitter. We train classifiers using lexical, semantic, and sentiment features, and experiment with both manually and automatically generated training data. Our approach yields good performance at identifying positive and negative similes, and substantially outperforms existing sentiment resources.

**Session 1E: Short Paper Posters**

Lower Level Foyer

Chair: Kevin Duh

10:30–12:10

**Cross-document Event Coreference Resolution based on Cross-media Features***Tongtao Zhang, Hongzhi Li, Heng Ji, and Shih-Fu Chang*

In this paper we focus on a new problem of event coreference resolution across television news videos. Based on the observation that the contents from multiple data modalities are complementary, we develop a novel approach to jointly encode effective features from both closed captions and video key frames. Experiment results demonstrate that visual features provided 7.2% absolute F-score gain on state-of-the-art text based event extraction and coreference resolution.

**A Survey of Current Datasets for Vision and Language Research***Francis Ferraro, Nasrin Mostafazadeh, Ting-Hao Huang, Lucy Vanderwende, Jacob Devlin, Michel Galley, and Margaret Mitchell*

Integrating vision and language has long been a dream in work on artificial intelligence (AI). In the past two years, we have witnessed an explosion of work that brings together vision and language from images to videos and beyond. The available corpora have played a crucial role in advancing this area of research. In this paper, we propose a set of quality metrics for evaluating and analyzing the vision & language datasets and categorize them accordingly. Our analyses show that the most recent datasets have been using more complex language and more abstract concepts, however, there are different strengths and weaknesses in each.

**Combining Geometric, Textual and Visual Features for Predicting Prepositions in Image Descriptions***Arnaud Ramisa, Josiah Wang, Ying Lu, Emmanuel Dellandrea, Francesc Moreno-Noguer, and Robert Gaizauskas*

We investigate the role that geometric, textual and visual features play in the task of predicting a preposition that links two visual entities depicted in an image. The task is an important part of the subsequent process of generating image descriptions. We explore the prediction of prepositions for a pair of entities, both in the case when the labels of such entities are known and unknown. In all situations we found clear evidence that all three features contribute to the prediction task.

**On A Strictly Convex IBM Model 1***Andrei Simion, Michael Collins, and Cliff Stein*

IBM Model 1 is a classical alignment model. Of the first generation word-based SMT models, it was the only such model with a concave objective function. For concave optimization problems like IBM Model 1, we have guarantees on the convergence of optimization algorithms such as Expectation Maximization (EM). However, as was pointed out recently, the objective of IBM Model 1 is not strictly concave and there is quite a bit of alignment quality variance within the optimal solution set. In this work we detail a strictly concave version of IBM Model 1 whose EM algorithm is a simple modification of the original EM algorithm of Model 1 and does not require the tuning of a learning rate or the insertion of an  $\$1\_2\$$  penalty. Moreover, by addressing Model 1's shortcomings, we achieve AER and F-Measure improvements over the classical Model 1 by over 30%

**Factorization of Latent Variables in Distributional Semantic Models***Arvid Österlund, David Ödling, and Magnus Sahlgren*

This paper discusses the use of factorization techniques in distributional semantic models. We focus on a method for redistributing the weight of latent variables, which have previously been shown to improve the performance of distributional semantic models. However, this result has not been replicated and remains poorly understood. We refine the method, and provide additional theoretical justification, as well as empirical results that demonstrate the viability of the proposed approach.

**Non-lexical neural architecture for fine-grained POS Tagging***Matthieu Labeau, Kevin Löser, and Alexandre Allauzen*

In this paper we explore a POS tagging application of neural architectures that can infer word representations from the raw character stream. It relies on two modelling stages that are jointly learnt: a convolutional network that infers a word representation directly from the character stream, followed

by a prediction stage. Models are evaluated on a POS and morphological tagging task for German. Experimental results show that the convolutional network can infer meaningful word representations, while for the prediction stage, a well designed and structured strategy allows the model to outperform state-of-the-art results, without any feature engineering.

### **Online Representation Learning in Recurrent Neural Language Models**

*Marek Rei*

We investigate an extension of continuous online learning in recurrent neural network language models. The model keeps a separate vector representation of the current unit of text being processed and adaptively adjusts it after each prediction. The initial experiments give promising results, indicating that the method is able to increase language modelling accuracy, while also decreasing the parameters needed to store the model along with the computation required at each step.

### **A Model of Zero-Shot Learning of Spoken Language Understanding**

*Majid Yazdani and James Henderson*

When building spoken dialogue systems for a new domain, a major bottleneck is developing a spoken language understanding (SLU) module that handles the new domain's terminology and semantic concepts. We propose a statistical SLU model that generalises to both previously unseen input words and previously unseen output classes by leveraging unlabelled data. After mapping the utterance into a vector space, the model exploits the structure of the output labels by mapping each label to a hyperplane that separates utterances with and without that label. Both these mappings are initialised with unsupervised word embeddings, so they can be computed even for words or concepts which were not in the SLU training data.

### **Modeling Tweet Arrival Times using Log-Gaussian Cox Processes**

*Michał Lukasik, P. K. Srijith, Trevor Cohn, and Kalina Bontcheva*

Research on modeling time series text corpora has typically focused on predicting what text will come next, but less well studied is predicting when the next text event will occur. In this paper we address the latter case, framed as modeling continuous inter-arrival times under a log-Gaussian Cox process, a form of inhomogeneous Poisson process which captures the varying rate at which the tweets arrive over time. In an application to rumour modeling of tweets surrounding the 2014 Ferguson riots, we show how inter-arrival times between tweets can be accurately predicted, and that incorporating textual features further improves predictions.

### **Pre-Computable Multi-Layer Neural Network Language Models**

*Jacob Devlin, Chris Quirk, and Arul Menezes*

In the last several years, neural network models have significantly improved accuracy in a number of NLP tasks. However, one serious drawback that has impeded their adoption in production systems is the slow runtime speed of neural network models compared to alternate models, such as maximum entropy classifiers. In Devlin 2014, the authors presented a simple technique for speeding up feed-forward embedding-based neural network models, where the dot product between each word embedding and part of the first hidden layer are pre-computed offline. However, this technique cannot be used for hidden layers beyond the first. In this paper, we explore a neural network architecture where the embedding layer feeds into multiple hidden layers that are placed "next to" one another so that each can be pre-computed independently. On a large scale language modeling task, this lateral architecture achieves a 10x speedup at runtime and a significant reduction in perplexity when compared to a standard multi-layer network.

### **Birds of a Feather Linked Together: A Discriminative Topic Model using Link-based Priors**

*Weiwei Yang, Jordan Boyd-Graber, and Philip Resnik*

A wide range of applications, from social media to scientific literature analysis, involve graphs in which documents are connected by links. We introduce a topic model for link prediction based on the intuition that linked documents will tend to have similar topic distributions, integrating a max-margin learning criterion and lexical term weights in the loss function. We validate our approach on the tweets from 2,000 Sina Weibo users and evaluate our model's reconstruction of the social network.

### **Aligning Knowledge and Text Embeddings by Entity Descriptions**

*Huaping Zhong, Jianwen Zhang, Zhen Wang, Hai Wan, and Zheng Chen*

We study the problem of jointly embedding a knowledge base and a text corpus. The key issue is the alignment model making sure the vectors of entities, relations and words are in the same space. Wang et al. (2014a) rely on Wikipedia anchors, making the applicable scope quite limited. In this paper we propose a new alignment model based on text descriptions of entities, without dependency on anchors. We require the embedding vector of an entity not only to fit the structured constraints in KBs but also to be equal to the embedding vector computed from the text description. Extensive experiments show that, the proposed approach consistently performs comparably or even better than the method of Wang et al. (2014a), which is encouraging as we do not use any anchor information.

### **An Empirical Analysis of Optimization for Max-Margin NLP**

*Jonathan K. Kummerfeld, Taylor Berg-Kirkpatrick, and Dan Klein*

Despite the convexity of structured max-margin objectives (Taskar et al., 2004, Tschantzidis et al., 2004), the many ways to optimize them are not equally effective in practice. We compare a range of online optimization methods over a variety of structured NLP tasks (coreference, summarization, parsing, etc) and find several broad trends. First, margin methods do tend to outperform both likelihood and the perceptron. Second, for max-margin objectives, primal optimization methods are often more robust and progress faster than dual methods. This advantage is most pronounced for tasks with dense or continuous-valued features. Overall, we argue for a particularly simple online primal subgradient descent method that, despite being rarely mentioned in the literature, is surprisingly effective in relation to its alternatives.

### **Learning Better Embeddings for Rare Words Using Distributional Representations**

*Irina Sergienya and Hinrich Schütze*

There are two main types of word representations: low-dimensional embeddings and high-dimensional distributional vectors, in which each dimension corresponds to a context word. In this paper, we initialize an embedding-learning model with distributional vectors. Evaluation on word similarity shows that this initialization significantly increases the quality of embeddings for rare words.

### **Composing Relationships with Translations**

*Alberto Garcia-Duran, Antoine Bordes, and Nicolas Usunier*

Performing link prediction in Knowledge Bases (KBs) with embedding-based models, like with the model TransE (Bordes et al., 2013) which represents relationships as translations in the embedding space, have shown promising results in recent years. Most of these works focused on modeling single relationships and hence do not take full advantage of the graph structure of KBs. In this paper, we propose an extension of TransE that learns to explicitly model composition of relationships via the addition of their corresponding translation vectors. We show empirically that this allows to improve performance for predicting single relationships as well as compositions of pairs of them.

### **Noise or additional information? Leveraging crowdsource annotation item agreement for natural language tasks.**

*Emily Jamison and Iryna Gurevych*

In order to reduce noise in training data, most natural language crowdsourcing annotation tasks gather redundant labels and aggregate them into an integrated label, which is provided to the classifier. However, aggregation discards potentially useful information from linguistically ambiguous instances. For five natural language tasks, we pass item agreement on to the task classifier via soft labeling and low-agreement filtering of the training dataset. We find a statistically significant benefit from low item agreement training filtering in four of our five tasks, and no systematic benefit from soft labeling.

## Session 2 Overview

---

	<b>Track A</b>	<b>Track B</b>	<b>Track C</b>
	Statistical Models and Machine Learning Methods (Long + TACL Papers) Main Auditorium	Tagging, Chunking and Parsing (Long + TACL Papers) Small Auditorium	Summarization (Long Papers) Room 2
13:30	Evaluation methods for unsupervised word embeddings <i>Schnabel, Labutov, Mimno, and Joachims</i>	Density-Driven Cross-Lingual Transfer of Dependency Parsers <i>Rasooli and Collins</i>	Sentence Compression by Deletion with LSTMs <i>Filippova, Alfonseca, Colmenares, Kaiser, and Vinyals</i>
13:55	Efficient Methods for Incorporating Knowledge into Topic Models <i>Yang, Downey, and Boyd-Graber</i>	A Neural Network Model for Low-Resource Universal Dependency Parsing <i>Duong, Cohn, Bird, and Cook</i>	An Empirical Comparison Between N-gram and Syntactic Language Models for Word Ordering <i>Liu and Zhang</i>
14:20	Traversing Knowledge Graphs in Vector Space <i>Guu, Miller, and Liang</i>	Improved Transition-based Parsing by Modeling Characters instead of Words with LSTMs <i>Ballesteros, Dyer, and Smith</i>	A Neural Attention Model for Abstractive Sentence Summarization <i>Rush, Chopra, and Weston</i>
14:45	[TACL] Improving Topic Models with Latent Feature Word Representations <i>Nguyen, Billingsley, Du, and Johnson</i>	[TACL] Approximation-Aware Dependency Parsing by Belief Propagation <i>Gormley, Dredze, and Eisner</i>	Scientific Article Summarization Using Citation-Context and Article's Discourse Structure <i>Cohan and Goharian</i>

**Track D: Long Paper Posters**

Chair: Nate Chambers

Lower Level Foyer

13:30–15:10

**Poster Cluster 1: Text Mining and NLP Applications (P1-9)**

- Hashtag Recommendation Using Dirichlet Process Mixture Models Incorporating Types of Hashtags  
*Y. Gong, Q. Zhang, and X. Huang*
- A Graph-based Readability Assessment Method using Word Coupling  
*Z. Jiang, G. Sun, Q. Gu, T. Bai, and D. Chen*
- More Features Are Not Always Better: Evaluating Generalizing Models in Incident Type Classification of Tweets  
*A. Schulz, C. Guckelsberger, and B. Schmidt*
- Flexible Domain Adaptation for Automated Essay Scoring Using Correlated Linear Regression  
*P. Phandi, K. M. A. Chai, and H. T. Ng*

- Show Me Your Evidence - an Automatic Method for Context Dependent Evidence Detection  
*R. Rinott, L. Dankin, C. Alzate Perez, M. M. Khapra, E. Aharoni, and N. Slonim*
- Spelling Correction of User Search Queries through Statistical Machine Translation  
*S. Hasan, C. Heger, and S. Mansour*
- Human Evaluation of Grammatical Error Correction Systems  
*R. Grundkiewicz, M. Junczys-Dowmunt, and E. Gillian*
- Learning a Deep Hybrid Model for Semi-Supervised Text Classification  
*A. Ororbia II, C. L. Giles, and D. Reitter*
- Joint Embedding of Query and Ad by Leveraging Implicit Feedback  
*S. Lee and Y. Hu*

**Track E: Short Paper Posters**

Lower Level Foyer

Chair: Wei Lu

13:30–15:10

**Poster Cluster 1: Information Extraction (P1-11)**

- Automatic Extraction of Time Expressions Accross Domains in French Narratives  
*M. D. Tapi Nzali, X. Tannier, and A. Neveol*
- Semi-Supervised Bootstrapping of Relationship Extractors with Distributional Semantics  
*D. S. Batista, B. Martins, and M. J. Silva*
- Extraction and generalisation of variables from scientific publications  
*E. Marsi and P. Öztürk*
- Named entity recognition with document-specific KB tag gazetteers  
*W. Radford, X. Carreras, and J. Henderson*
- "A Spousal Relation Begins with a Deletion of engage and Ends with an Addition of divorce": Learning State Changing Verbs from Wikipedia Revision History  
*D. T. Wijaya, N. Nakashole, and T. Mitchell*
- Improving Distant Supervision for Information Extraction Using Label Propagation Through Lists  
*L. Bing, S. Chaudhari, R. Wang, and W. Cohen*
- An Entity-centric Approach for Overcoming Knowledge Graph Sparsity  
*M. Hegde and P. P. Talukdar*
- Semantic Relation Classification via Convolutional Neural Networks with Simple Negative Sampling  
*K. Xu, Y. Feng, S. Huang, and D. Zhao*
- A Baseline Temporal Tagger for all Languages  
*J. Strötgen and M. Gertz*
- Named Entity Recognition for Chinese Social Media with Jointly Trained Embeddings  
*N. Peng and M. Dredze*

- Inferring Binary Relation Schemas for Open Information Extraction  
*K. Luo, X. Luo, and K. Zhu*

**Poster Cluster 2: Information Retrieval and Question Answering (P12-16)**

- LDTM: A Latent Document Type Model for Cumulative Citation Recommendation  
*J. Wang, D. Song, Z. Zhang, L. Liao, L. Si, and C.-Y. Lin*
- Online Sentence Novelty Scoring for Topical Document Streams  
*S. Lee*
- Global Thread-level Inference for Comment Classification in Community Question Answering  
*S. Joty, A. Barrón-Cedeño, G. Da San Martino, S. Filice, L. Márquez, A. Moschitti, and P. Nakov*
- Key Concept Identification for Medical Information Retrieval  
*J. Zheng and H. Yu*
- Image-Mediated Learning for Zero-Shot Cross-Lingual Document Retrieval  
*R. Funaki and H. Nakayama*

## Abstracts: Session 2

---

### Session 2A: Statistical Models and Machine Learning Methods (Long + TACL Papers)

Main Auditorium

Chair: Alex Klementiev

#### Evaluation methods for unsupervised word embeddings

*Tobias Schnabel, Igor Labutov, David Mimno, and Thorsten Joachims*

13:30–13:55

We present a comprehensive study of evaluation methods for unsupervised embedding techniques that obtain meaningful representations of words from text. Different evaluations result in different orderings of embedding methods, calling into question the common assumption that there is one single optimal vector representation. We present new evaluation techniques that directly compare embeddings with respect to specific queries. These methods reduce bias, provide greater insight, and allow us to solicit data-driven relevance judgments rapidly and accurately through crowdsourcing.

#### Efficient Methods for Incorporating Knowledge into Topic Models

*Yi Yang, Doug Downey, and Jordan Boyd-Graber*

13:55–14:20

Latent Dirichlet allocation (LDA) is a popular topic modeling technique for exploring hidden topics in text corpora. Increasingly, topic modeling is trying to scale to larger topic spaces and use richer forms of prior knowledge, such as word correlations or document labels. However, inference is cumbersome for LDA models with prior knowledge. As a result, LDA models that use prior knowledge only work in small-scale scenarios. In this work, we propose a factor graph framework, Sparse Constrained LDA (SC-LDA), for efficiently incorporating prior knowledge into LDA. In experiments, we evaluate SC-LDA's ability to incorporate word correlation knowledge and document label knowledge on three benchmark datasets. Compared to several baseline methods, SC-LDA achieves comparable performance but runs significantly faster.

#### Traversing Knowledge Graphs in Vector Space

*Kelvin Guu, John Miller, and Percy Liang*

14:20–14:45

Path queries on a knowledge graph can be used to answer compositional questions such as "What languages are spoken by people living in Lisbon?". However, knowledge graphs often have missing facts (edges) which disrupts path queries. Recent models for knowledge base completion impute missing facts by embedding knowledge graphs in vector spaces. We show that these models can be recursively applied to answer path queries, but that they suffer from cascading errors. This motivates a new "compositional" training objective, which dramatically improves all models' ability to answer path queries, in some cases more than doubling accuracy. On a standard knowledge base completion task, we also demonstrate that compositional training acts as a novel form of structural regularization, reliably improving performance across all base models (reducing errors by up to 43%) and achieving new state-of-the-art results.

#### [TACL] Improving Topic Models with Latent Feature Word Representations

*Dat Quoc Nguyen, Richard Billingsley, Lan Du, and Mark Johnson*

14:45–15:10

Probabilistic topic models are widely used to discover latent topics in document collections, while latent feature vector representations of words have been used to obtain high performance in many NLP tasks. In this paper, we extend two different Dirichlet multinomial topic models by incorporating latent feature vector representations of words trained on very large corpora to improve the word-topic mapping learnt on a smaller corpus. Experimental results show that by using information from the external corpora, our new models produce significant improvements on topic coherence, document clustering and document classification tasks, especially on datasets with few or short documents.

**Session 2B: Tagging, Chunking and Parsing (Long +TACL Papers)**

Small Auditorium

Chair: Eugene Charniak

**Density-Driven Cross-Lingual Transfer of Dependency Parsers***Mohammad Sadegh Rasooli and Michael Collins*

13:30–13:55

We present a novel method for the cross-lingual transfer of dependency parsers. Our goal is to induce a dependency parser in a target language of interest without any direct supervision: instead we assume access to parallel translations between the target and one or more source languages, and to supervised parsers in the source language(s). Our key contributions are to show the utility of dense projected structures when training the target language parser, and to introduce a novel learning algorithm that makes use of dense structures. Results on several languages show an absolute improvement of 5.51% in average dependency accuracy over the state-of-the-art method of (Ma and Xia, 2014). Our average dependency accuracy of 82.18% compares favourably to the accuracy of fully supervised methods.

**A Neural Network Model for Low-Resource Universal Dependency Parsing***Long Duong, Trevor Cohn, Steven Bird, and Paul Cook*

13:55–14:20

Accurate dependency parsing requires large treebanks, which are only available for a few languages. We propose a method that takes advantage of shared structure across languages to build a mature parser using less training data. We propose a model for learning a shared “universal” parser that operates over an inter-lingual continuous representation of language, along with language-specific mapping components. Compared with supervised learning, our methods give a consistent 8-10 % improvement across several treebanks in low-resource simulations.

**Improved Transition-based Parsing by Modeling Characters instead of Words with LSTMs***Miguel Ballesteros, Chris Dyer, and Noah A. Smith*

14:20–14:45

We present extensions to a continuous-state dependency parsing method that makes it applicable to morphologically rich languages. Starting with a high-performance transition-based parser that uses long short-term memory (LSTM) recurrent neural networks to learn representations of the parser state, we replace lookup-based word representations with representations constructed from the orthographic representations of the words, also using LSTMs. This allows statistical sharing across word forms that are similar on the surface. Experiments for morphologically rich languages show that the parsing model benefits from incorporating the character-based encodings of words.

**[TACL] Approximation-Aware Dependency Parsing by Belief Propagation***Matthew R. Gormley, Mark Dredze, and Jason Eisner*

14:45–15:10

We show how to train the fast dependency parser of Smith and Eisner (2008) for improved accuracy. This parser can consider higher-order interactions among edges while retaining  $O(n^3)$  runtime. It outputs the parse with maximum expected recall—but for speed, this expectation is taken under a posterior distribution that is constructed only approximately, using loopy belief propagation through structured factors. We show how to adjust the model parameters to compensate for the errors introduced by this approximation, by following the gradient of the actual loss on training data. We find this gradient by backpropagation. That is, we treat the entire parser (approximations and all) as a differentiable circuit, as others have done for loopy CRFs (Domke, 2010; Stoyanov et al., 2011; Domke, 2011; Stoyanov and Eisner, 2012). The resulting parser obtains higher accuracy with fewer iterations of belief propagation than one trained by conditional log-likelihood.

**Session 2C: Summarization (Long Papers)**

Room 2

Chair: Yvette Graham

**Sentence Compression by Deletion with LSTMs**

*Katja Filippova, Enrique Alfonseca, Carlos A. Colmenares, Lukasz Kaiser, and Oriol Vinyals*  
13:30–13:55

We present an LSTM approach to deletion-based sentence compression where the task is to translate a sentence into a sequence of zeros and ones, corresponding to token deletion decisions. We demonstrate that even the most basic version of the system, which is given no syntactic information (no PoS or NE tags, or dependencies) or desired compression length, performs surprisingly well: around 30% of the compressions from a large test set could be regenerated. We compare the LSTM system with a competitive baseline which is trained on the same amount of data but is additionally provided with all kinds of linguistic features. In an experiment with human raters the LSTM-based model outperforms the baseline achieving 4.5 in readability and 3.8 in informativeness.

**An Empirical Comparison Between N-gram and Syntactic Language Models for Word Ordering**

*Jiangming Liu and Yue Zhang* 13:55–14:20

Syntactic language models and N-gram language models have both been used in word ordering. In this paper, we give an empirical comparison between N-gram and syntactic language models on word order task. Our results show that the quality of automatically-parsed training data has a relatively small impact on syntactic models. Both of syntactic and N-gram models can benefit from large-scale raw text. Compared with N-gram models, syntactic models give overall better performance, but they require much more training time. In addition, the two models lead to different error distributions in word ordering. A combination of the two models integrates the advantages of each model, achieving the best result in a standard benchmark.

**A Neural Attention Model for Abstractive Sentence Summarization**

*Alexander M. Rush, Sumit Chopra, and Jason Weston* 14:20–14:45

Summarization based on text extraction is inherently limited, but generation-style abstractive methods have proven challenging to build. In this work, we propose a fully data-driven approach to abstractive sentence summarization. Our method utilizes a local attention-based model that generates each word of the summary conditioned on the input sentence. While the model is structurally simple, it can easily be trained end-to-end and scales to a large amount of training data. The model shows significant performance gains on the DUC-2004 shared task compared with several strong baselines.

**Scientific Article Summarization Using Citation-Context and Article's Discourse Structure**

*Arman Cohan and Nazli Goharian* 14:45–15:10

We propose a summarization approach for scientific articles which takes advantage of the citation-context and the document discourse model. While citations have been previously used in generating scientific summaries, they lack the related context from the referenced article and therefore do not accurately reflect the article's content. Our method overcomes the problem of inconsistency between the citation summary and the article's content by providing context for each citation. We also leverage the inherent scientific article's discourse model for producing better summaries. We show that our proposed method effectively improves over existing summarization approaches (greater than 30% improvement over the best performing baseline) in terms of ROUGE scores on TAC2014 scientific summarization dataset. While the dataset we use for evaluation is in the biomedical domain, most of our approaches are general and therefore adaptable to other domains.

**Session 2D: Long Paper Posters**

Lower Level Foyer

Chair: Nate Chambers

13:30–15:10

**Hashtag Recommendation Using Dirichlet Process Mixture Models Incorporating Types of Hashtags***Yeyun Gong, Qi Zhang, and Xuanjing Huang*

In recent years, the task of recommending hashtags for microblogs has been given increasing attention. Various methods have been proposed to study the problem from different aspects. However, most of the recent studies have not considered the differences in the types or uses of hashtags. In this paper, we introduce a novel nonparametric Bayesian method for this task. Based on the Dirichlet Process Mixture Models (DPMM), we incorporate the type of hashtag as a hidden variable. The results of experiments on the data collected from a real world microblogging service demonstrate that the proposed method outperforms state-of-the-art methods that do not consider these aspects. By taking these aspects into consideration, the relative improvement of the proposed method over the state-of-the-art methods is around 12.2% in F1-score.

**A Graph-based Readability Assessment Method using Word Coupling***Zhiwei Jiang, Gang Sun, Qing Gu, Tao Bai, and Daoxu Chen*

This paper proposes a graph-based readability assessment method using word coupling. Compared to the state-of-the-art methods such as the readability formulae, the word-based and feature-based methods, our method develops a coupled bag-of-words model which combines the merits of word frequencies and text features. Unlike the general bag-of-words model which assumes words are independent, our model correlates the words based on their similarities on readability. By applying TF-IDF (Term Frequency and Inverse Document Frequency), the coupled TF-IDF matrix is built, and used in the graph-based classification framework, which involves graph building, merging and label propagation. Experiments are conducted on both English and Chinese datasets. The results demonstrate both effectiveness and potential of the method.

**More Features Are Not Always Better: Evaluating Generalizing Models in Incident Type Classification of Tweets***Axel Schulz, Christian Guckelsberger, and Benedikt Schmidt*

Social media is a rich source of up-to-date information about events such as incidents. The sheer amount of available information makes machine learning approaches a necessity to process this information further. This learning problem is often concerned with regionally restricted datasets such as data from only one city. Because social media data such as tweets varies considerably across different cities, the training of efficient models requires labeling data from each city of interest, which is costly and time consuming. In this paper, we investigate which features are most suitable for training generalizable models, i.e., models that show good performance across different datasets. We re-implemented the most popular features from the state of the art in addition to other novel approaches, and evaluated them on data from ten different cities. We show that many sophisticated features are not necessarily valuable for training a generalized model and are outperformed by classic features such as plain word-n-grams and character-n-grams.

**Flexible Domain Adaptation for Automated Essay Scoring Using Correlated Linear Regression***Peter Phandi, Kian Ming A. Chai, and Hwee Tou Ng*

Most of the current automated essay scoring (AES) systems are trained using manually graded essays from a specific prompt. These systems experience a drop in accuracy when used to grade an essay from a different prompt. Obtaining a large number of manually graded essays each time a new prompt is introduced is costly and not viable. We propose domain adaptation as a solution to adapt an AES system from an initial prompt to a new prompt. We also propose a novel domain adaptation technique that uses Bayesian linear ridge regression. We evaluate our domain adaptation technique on the publicly available Automated Student Assessment Prize (ASAP) dataset and show that our proposed technique is a competitive default domain adaptation algorithm for the AES task.

## Show Me Your Evidence - an Automatic Method for Context Dependent Evidence Detection

*Ruty Rinott, Lena Dankin, Carlos Alzate Perez, Mitesh M. Khapra, Ehud Aharoni, and Noam Slonim*

Engaging in a debate with oneself or others to take decisions is an integral part of our day-to-day life. A debate on a topic (say, use of performance enhancing drugs) typically proceeds by one party making an assertion/claim (say, PEDs are bad for health) and then providing an evidence to support the claim (say, a 2006 study shows that PEDs have psychiatric side effects). In this work, we propose the task of automatically detecting such evidences from unstructured text that support a given claim. This task has many practical applications in decision support and persuasion enhancement in a wide range of domains. We first introduce an extensive benchmark data set tailored for this task, which allows training statistical models and assessing their performance. Then, we suggest a system architecture based on supervised learning to address the evidence detection task. Finally, promising experimental results are reported.

## Spelling Correction of User Search Queries through Statistical Machine Translation

*Saša Hasan, Carmen Heger, and Saab Mansour*

We use character-based statistical machine translation in order to correct user search queries in the e-commerce domain. The training data is automatically extracted from event logs where users re-issue their search queries with potentially corrected spelling within the same session. We show results on a test set which was annotated by humans and compare against online autocorrection capabilities of three additional web sites. Overall, the methods presented in this paper outperform fully productized spellchecking and autocorrection services in terms of accuracy and F1 score. We also propose novel evaluation steps based on retrieved search results of the corrected queries in terms of quantity and relevance.

## Human Evaluation of Grammatical Error Correction Systems

*Roman Grundkiewicz, Marcin Junczys-Dowmunt, and Edward Gillian*

The paper presents the results of the first large-scale human evaluation of automatic grammatical error correction (GEC) systems. Twelve participating systems and the unchanged input of the CoNLL-2014 shared task have been reassessed in a WMT-inspired human evaluation procedure. Methods introduced for the Workshop of Machine Translation evaluation campaigns have been adapted to GEC and extended where necessary. The produced rankings are used to evaluate standard metrics for grammatical error correction in terms of correlation with human judgment.

## Learning a Deep Hybrid Model for Semi-Supervised Text Classification

*Alexander Ororbia II, C. Lee Giles, and David Reitter*

We present a novel fine-tuning algorithm in a deep hybrid architecture for semi-supervised text classification. During each increment of the online learning process, the fine-tuning algorithm serves as a top-down mechanism for pseudo-jointly modifying model parameters following a bottom-up generative learning pass. The resulting model, trained under what we call the Bottom-Up-Top-Down learning algorithm, is shown to outperform a variety of competitive models and baselines trained across a wide range of splits between supervised and unsupervised training data.

## Joint Embedding of Query and Ad by Leveraging Implicit Feedback

*Sungjin Lee and Yifan Hu*

Sponsored search is at the center of a multibillion dollar market established by search technology. Accurate ad click prediction is a key component for this market to function since the pricing mechanism heavily relies on the estimation of click probabilities. Lexical features derived from the text of both the query and ads play a significant role, complementing features based on historical click information. The purpose of this paper is to explore the use of word embedding techniques to generate effective text features that can capture not only lexical similarity between query and ads but also the latent user intents. We identify several potential weaknesses of the plain application of conventional word embedding methodologies for ad click prediction. These observations motivated us to propose a set of novel joint word embedding methods by leveraging implicit click feedback. We verify the effectiveness of these new word embedding models by adding features derived from the new models to the click prediction system of a commercial search engine. Our evaluation results clearly demonstrate the effectiveness of

the proposed methods. To the best of our knowledge this work is the first successful application of word embedding techniques for the sponsored search task.

**Session 2E: Short Paper Posters**

Lower Level Foyer

Chair: Wei Lu

13:30–15:10

**Automatic Extraction of Time Expressions Accross Domains in French Narratives***Mike Donald Tapi Nzali, Xavier Tannier, and Aurelie Neveol*

The prevalence of temporal references accross all types of natural language utterances makes temporal analysis a key issue in Natural Language Processing. This work addresses three research questions: 1/is temporal expression recognition specific to a particular domain? 2/if so, can we characterize domain specificity? and 3/how can sudomain specificity be integrated in a single tool for unified temporal expression extraction? Herein, we assess temporal expression recognition from documents written in French covering three domains. We present a new corpus of clinical narratives annotated for temporal expressions, and also use existing corpora in the newswire and historical domains. We show that temporal expressions can be extracted with high performance across domains (best F-measure 0.96 obtained with a CRF model on clinical narratives). We argue that domain adaptation for the extraction of temporal expressions can be done with limited efforts and should cover pre-processing as well as temporal specific tasks.

**Semi-Supervised Bootstrapping of Relationship Extractors with Distributional Semantics***David S. Batista, Bruno Martins, and Mário J. Silva*

Semi-supervised bootstrapping techniques for relationship extraction from text iteratively expand a set of initial seed relationships while limiting the semantic drift. We research bootstrapping for relationship extraction using word embeddings to find similar relationships. Experimental results show that relying on word embeddings achieves a better performance on the task of extracting four types of relationships from a collection of newswire documents when compared with a baseline using TF-IDF to find similar relationships.

**Extraction and generalisation of variables from scientific publications***Erwin Marsi and Pinar ÖzTÜRK*

Scientific theories and models in Earth science typically involve changing variables and their complex interactions, including correlations, causal relations and chains of positive/negative feedback loops. Variables tend to be complex rather than atomic entities and expressed as noun phrases containing multiple modifiers, e.g. “oxygen depletion in the upper 500 m of the ocean” or “timing and magnitude of surface temperature evolution in the Southern Hemisphere in deglacial proxy records”. Text mining from Earth science literature is therefore significantly different from biomedical text mining and requires different approaches and methods. Our approach aims at automatically locating and extracting variables and their direction of variation: increasing, decreasing or just changing. Variables are initially extracted by matching tree patterns onto the syntax trees of the source texts. Next, variables are generalised in order to enhance their similarity, facilitating hierarchical search and inference. This generalisation is accomplished by progressive pruning of syntax trees using a set of tree transformation operations. Text mining results are presented as a browsable variable hierarchy which allows users to inspect all mentions of a particular variable type in the text as well as any generalisations or specialisations. The approach is demonstrated on a corpus of 10k abstracts of Nature publications in the field of Marine science. We discuss experiences with this early prototype and outline a number of possible improvements and directions for future research.

**Named entity recognition with document-specific KB tag gazetteers***Will Radford, Xavier Carreras, and James Henderson*

We consider a novel setting for Named Entity Recognition (NER) where we have access to document-specific knowledge base tags. These tags consist of a canonical name from a knowledge base (KB) and entity type, but are not aligned to the text. We explore how to use KB tags to create document-specific gazetteers at inference time to improve NER. We find that this kind of supervision helps recognise organisations more than standard wide- coverage gazetteers. Moreover, augmenting document-specific gazetteers with KB information lets users specify fewer tags for the same performance, reducing cost.

**“A Spousal Relation Begins with a Deletion of engage and Ends with an Addition of divorce”: Learning State Changing Verbs from Wikipedia Revision History***Derry Tanti Wijaya, Ndapandula Nakashole, and Tom Mitchell*

Learning to determine when the time-varying facts of a Knowledge Base (KB) have to be updated is a challenging task. We propose to learn state changing verbs from Wikipedia edit history. When a state-changing event, such as a marriage or death, happens to an entity, the infobox on the entity's Wikipedia page usually gets updated. At the same time, the article text may be updated with verbs either being added or deleted to reflect the changes made to the infobox. We use Wikipedia edit history to distantly supervise a method for automatically learning verbs and state changes. Additionally, our method uses constraints to effectively map verbs to infobox changes. We observe in our experiments that when state-changing verbs are added or deleted from an entity's Wikipedia page text, we can predict the entity's infobox updates with 88% precision and 76% recall. One compelling application of our verbs is to incorporate them as triggers in methods for updating existing KBs, which are currently mostly static.

### **Improving Distant Supervision for Information Extraction Using Label Propagation Through Lists**

*Lidong Bing, Sneha Chaudhari, Richard Wang, and William Cohen*

Because of polysemy, distant labeling for information extraction leads to noisy training data. We describe a procedure for reducing this noise by using label propagation on a graph in which the nodes are entity mentions, and mentions are coupled when they occur in coordinate list structures. We show that this labeling approach leads to good performance even when off-the-shelf classifiers are used on the distantly-labeled data.

### **An Entity-centric Approach for Overcoming Knowledge Graph Sparsity**

*Manjunath Hegde and Partha P. Talukdar*

Automatic construction of knowledge graphs (KGs) from unstructured text has received considerable attention in recent research, resulting in the construction of several KGs with millions of entities (nodes) and facts (edges) among them. Unfortunately, such KGs tend to be severely sparse in terms of number of facts known for a given entity, i.e., have low knowledge density. For example, the NELL KG consists of only 1.34 facts per entity. Unfortunately, such low knowledge density makes it challenging to use such KGs in real-world applications. In contrast to best-effort extraction paradigms followed in the construction of such KGs, in this paper we argue in favor of ENTITY Centric Expansion (ENTICE), an entity-centric KG population frame-work, to alleviate the low knowledge density problem in existing KGs. By using ENTICE, we are able to increase NELL's knowledge density by a factor of 7.7 at 75.5% accuracy. Additionally, we are also able to extend the ontology discovering new relations and entities.

### **Semantic Relation Classification via Convolutional Neural Networks with Simple Negative Sampling**

*Kun Xu, Yansong Feng, Songfang Huang, and Dongyan Zhao*

Syntactic features play an essential role in identifying relationship in a sentence. Previous neural network models often suffer from irrelevant information introduced when subjects and objects are in a long distance. In this paper, we propose to learn more robust relation representations from the shortest dependency path through a convolution neural network. We further propose a straightforward negative sampling strategy to improve the assignment of subjects and objects. Experimental results show that our method outperforms the state-of-the-art methods on the SemEval-2010 Task 8 dataset.

### **A Baseline Temporal Tagger for all Languages**

*Jannik Strötgen and Michael Gertz*

Temporal taggers are usually developed for a certain language. Besides English, only few languages have been addressed, and only the temporal tagger HeidelTime covers several languages. While this tool was manually extended to these languages, there have been earlier approaches for automatic extensions to a single target language. In this paper, we present an approach to extend HeidelTime to all languages in the world. Our evaluation shows promising results, in particular considering that our approach neither requires language skills nor training data, but results in a baseline tagger for 200+ languages.

### **Named Entity Recognition for Chinese Social Media with Jointly Trained Embeddings**

*Nanyun Peng and Mark Dredze*

We consider the task of named entity recognition for Chinese social media. The long line of work in Chinese NER has focused on formal domains, and NER for social media has been largely restricted to

English. We present a new corpus of Weibo messages annotated for both name and nominal mentions. Additionally, we evaluate three types of neural embedding for representing Chinese text. Finally, we propose a joint training objective for the embeddings that makes use of both (NER) labeled and unlabeled raw text. Our methods yield a 9% improvement over a state-of-the-art baseline.

### **Inferring Binary Relation Schemas for Open Information Extraction**

*Kangqi Luo, Xusheng Luo, and Kenny Zhu*

This paper presents a framework to model the semantic representation of binary relations produced by open information extraction systems. For each binary relation, we infer a set of preferred types on the two arguments simultaneously, and generate a ranked list of type pairs which we call schemas. All inferred types are drawn from the Freebase type taxonomy, which are human readable. Our system collects 171,168 binary relations from ReVerb, and is able to produce top-ranking relation schemas with a mean reciprocal rank of 0.337.

### **LDTM: A Latent Document Type Model for Cumulative Citation Recommendation**

*Jingang Wang, Dandan Song, Zhiwei Zhang, Lejian Liao, Luo Si, and Chin-Yew Lin*

This paper studies Cumulative Citation Recommendation (CCR) - given an entity in Knowledge Bases, how to effectively detect its potential citations from volume text streams. Most previous approaches treated all kinds of features indifferently to build a global relevance model, in which the prior knowledge embedded in documents cannot be exploited adequately. To address this problem, we propose a latent document type discriminative model by introducing a latent layer to capture the correlations between documents and their underlying types. The model can better adjust to different types of documents and yield flexible performance when dealing with a broad range of document types. An extensive set of experiments has been conducted on TREC-KBA-2013 dataset, and the results demonstrate that this model can yield a significant performance gain on recommendation quality as compared to the state-of-the-art.

### **Online Sentence Novelty Scoring for Topical Document Streams**

*Sungjin Lee*

The enormous amount of information on the Internet has raised the challenge of highlighting new information in the context of already viewed content. This type of intelligent interface can save users time and prevent frustration. Our goal is to scale out novelty detection to large web properties like Google and Yahoo News. We present a set of efficient, light-weight features for online novelty scoring and a fast nonlinear feature transformation method using Deep Neural Network. Our experimental results on the TREC 2004 datasets show that the proposed method is not only efficient but also very powerful, significantly surpassing the best challenge system at TREC 2004.

### **Global Thread-level Inference for Comment Classification in Community Question Answering**

*Shafiq Joty, Alberto Barrón-Cedeño, Giovanni Da San Martino, Simone Filice, Lluís Márquez, Alessandro Moschitti, and Preslav Nakov*

Community question answering, a recent evolution of question answering in the Web context, allows a user to quickly consult the opinion of a number of people on a particular topic, thus taking advantage of the wisdom of the crowd. Here we try to help the user by deciding automatically which answers are good and which are bad for a given question. In particular, we focus on exploiting the output structure at the thread level in order to make more consistent global decisions. More specifically, we exploit the relations between pairs of comments at any distance in the thread, which we incorporate in a graph-cut and in an ILP frameworks. The evaluation on the benchmark dataset of SemEval-2015 Task 3 confirms the importance of using thread-level information, which allows us to improve over the state of the art.

### **Key Concept Identification for Medical Information Retrieval**

*Jiaping Zheng and Hong Yu*

The difficult language in Electronic Health Records (EHRs) presents a challenge to patients' understanding of their own conditions. One approach to lowering the barrier is to provide tailored patient education based on their own EHR notes. We are developing a system to retrieve EHR note-tailored online consumer oriented health education materials. We explored topic model and key concept identification methods to construct queries from the EHR notes. Our experiments show that queries using identified

key concepts with pseudo-relevance feedback significantly outperform (over 10-fold improvement) the baseline system of using the full text note.

**Image-Mediated Learning for Zero-Shot Cross-Lingual Document Retrieval**

*Ruka Funaki and Hideki Nakayama*

We propose an image-mediated learning approach for cross-lingual document retrieval where no or only a few parallel corpora are available. Using the images in image-text documents of each language as the hub, we derive a common semantic subspace bridging two languages by means of generalized canonical correlation analysis. For the purpose of evaluation, we create and release a new document dataset consisting of three types of data (English text, Japanese text, and images). Our approach substantially enhances retrieval accuracy in zero-shot and few-shot scenarios where text-to-text examples are scarce.

## Session 3 Overview

---

Track A	Track B	Track C	
Sentiment Analysis and Opinion Mining (Long Papers) Main Auditorium	Semantics (Long +TACL Papers) Small Auditorium	Information Retrieval and Question Answering (Long Papers) Room 2	
Detecting Risks in the Banking System by Sentiment Analysis Nopp and Hanbury	A large annotated corpus for learning natural language inference Bowman, Angeli, Potts, and Manning	Name List Only? Target Entity Disambiguation in Short Texts Cao, Li, Guo, Bai, Ji, and Tang	15:40
Sentiment Flow - A General Model of Web Review Argumentation Wachsmuth, Kiesel, and Stein	Question-Answer Driven Semantic Role Labeling: Using Natural Language to Annotate Natural Language He, Lewis, and Zettlemoyer	Biography-Dependent Collaborative Entity Archiving for Slot Filling Hong, Wang, Chen, Wang, Zhang, and Ji	16:05
Neural Networks for Open Domain Targeted Sentiment Zhang, Zhang, and Vo	[TACL] It's All Fun and Games until Someone Annotates: Video Games with a Purpose for Linguistic Annotation. Jurgens and Navigli	Stochastic Top-k ListNet Luo, Wang, Liu, and Pan	16:30
Extracting Condition-Opinion Relations Toward Fine-grained Opinion Mining Nakayama and Fujii	[TACL] Semantic Proto-Roles Reisinger, Rudinger, Ferraro, Rawlins, and Van Durme	Exploring Markov Logic Networks for Question Answering Khot, Balasubramanian, Gribkoff, Sabharwal, Clark, and Etzioni	16:55

**Track D: Long + TACL Paper Posters**

Chair: Einat Minkov

Lower Level Foyer

15:40–17:20

**Poster Cluster 1: Information Extraction (P1-9)**

- Language and Domain Independent Entity Linking with Quantified Collective Validation  
*H. Wang, J. G. Zheng, X. Ma, P. Fox, and H. Ji*
- Modeling Relation Paths for Representation Learning of Knowledge Bases  
*Y. Lin, Z. Liu, H. Luan, M. Sun, S. Rao, and S. Liu*
- Corpus-level Fine-grained Entity Typing Using Contextual Information  
*Y. Yaghoobzadeh and H. Schütze*
- Knowledge Base Unification via Sense Embeddings and Disambiguation  
*C. Delli Bovi, L. Espinosa Anke, and R. Navigli*

- Open-Domain Name Error Detection using a Multi-Task RNN  
*H. Cheng, H. Fang, and M. Ostendorf*
- Extracting Relations between Non-Standard Entities using Distant Supervision and Imitation Learning  
*I. Augenstein, A. Vlachos, and D. Maynard*
- Sieve-Based Spatial Relation Extraction with Expanding Parse Trees  
*J. D'Souza and V. Ng*
- [TACL] Cross-Document Co-Reference Resolution using Sample-Based Clustering with Knowledge Enrichment  
*S. Dutta and G. Weikum*
- [TACL] Combining Minimally-supervised Methods for Arabic Named Entity Recognition  
*M. Althobaiti, U. Kruschwitz, and M. Poesio*

**Track E: Short Paper Posters**

Chair: Wei Xu

Lower Level Foyer

15:40-17:20

**Poster Cluster 1: Text Mining and NLP Applications (P1-13)**

- Mr. Bennet, his coachman, and the Archbishop walk into a bar but only one of them gets recognized: On The Difficulty of Detecting Characters in Literary Texts  
*H. Vala, D. Jurgens, A. Piper, and D. Ruths*
- Convolutional Sentence Kernel from Word Embeddings for Short Text Categorization  
*J. Kim, F. Rousseau, and M. Vazirgiannis*
- Predicting the Structure of Cooking Recipes  
*J. Jermsurawong and N. Habash*
- TSDPMM: Incorporating Prior Topic Knowledge into Dirichlet Process Mixture Models for Text Clustering  
*L. Hu, J. Li, X. Li, C. Shao, and X. Wang*
- Sentence Modeling with Gated Recursive Neural Network  
*X. Chen, X. Qiu, C. Zhu, S. Wu, and X. Huang*
- Learning Timeline Difference for Text Categorization  
*F. Fukumoto and Y. Suzuki*
- Summarizing Topical Contents from PubMed Documents Using a Thematic Analysis  
*S. Kim, L. Yeganova, and W. J. Wilbur*
- Recognizing Biographical Sections in Wikipedia  
*A. Palmero Aprosio and S. Tonelli*
- Learn to Solve Algebra Word Problems Using Quadratic Programming  
*L. Zhou, S. Dai, and L. Chen*
- An Unsupervised Method for Discovering Lexical Variations in Roman Urdu Informal Text  
*A. Rafaei, A. Qayyum, M. Moeenuddin, A. Karim, H. Sajjad, and F. Kamiran*

- Component-Enhanced Chinese Character Embeddings  
*Y. Li, W. Li, F. Sun, and S. Li*
- Multi-label Text Categorization with Joint Learning Predictions-as-Features Method  
*L. Li, H. Wang, X. Sun, B. Chang, S. Zhao, and L. Sha*
- A Framework for Comparing Groups of Documents  
*A. Maiya*

## Abstracts: Session 3

### Session 3A: Sentiment Analysis and Opinion Mining (Long Papers)

Main Auditorium

Chair: Janyce Wiebe

#### Detecting Risks in the Banking System by Sentiment Analysis

Clemens Nopp and Allan Hanbury

15:40–16:05

In November 2014, the European Central Bank (ECB) started to directly supervise the largest banks in the Eurozone via the Single Supervisory Mechanism (SSM). While supervisory risk assessments are usually based on quantitative data and surveys, this work explores whether sentiment analysis is capable of measuring a bank's attitude and opinions towards risk by analyzing text data. For realizing this study, a collection consisting of more than 500 CEO letters and outlook sections extracted from bank annual reports is built up. Based on these data, two distinct experiments are conducted. The evaluations find promising opportunities, but also limitations for risk sentiment analysis in banking supervision. At the level of individual banks, predictions are relatively inaccurate. In contrast, the analysis of aggregated figures revealed strong and significant correlations between uncertainty or negativity in textual disclosures and the quantitative risk indicator's future evolution. Risk sentiment analysis should therefore rather be used for macroprudential analyses than for assessments of individual banks.

#### Sentiment Flow - A General Model of Web Review Argumentation

Henning Wachsmuth, Johannes Kiesel, and Benno Stein

16:05–16:30

Web reviews have been intensively studied in argumentation-related tasks such as sentiment analysis. However, due to their focus on content-based features, many sentiment analysis approaches are effective only for reviews from those domains they have been specifically modeled for. This paper puts its focus on domain independence and asks whether a general model can be found for how people argue in web reviews. Our hypothesis is that people express their global sentiment on a topic with similar sequences of local sentiment independent of the domain. We model such sentiment flow robustly under uncertainty through abstraction. To test our hypothesis, we predict global sentiment based on sentiment flow. In systematic experiments, we improve over the domain independence of strong baselines. Our findings suggest that sentiment flow qualifies as a general model of web review argumentation.

#### Neural Networks for Open Domain Targeted Sentiment

Meishan Zhang, Yue Zhang, and Duy Tin Vo

16:30–16:55

Open domain targeted sentiment is the joint information extraction task that finds target mentions together with the sentiment towards each mention from a text corpus. The task is typically modeled as a sequence labeling problem, and solved using state-of-the-art labelers such as CRF. We empirically study the effect of word embeddings and automatic feature combinations on the task by extending a CRF baseline using neural networks, which have demonstrated large potentials for sentiment analysis. Results show that the neural model can give better results by significantly increasing the recall. In addition, we propose a novel integration of neural and discrete features, which combines their relative advantages, leading to significantly higher results compared to both baselines.

#### Extracting Condition-Opinion Relations Toward Fine-grained Opinion Mining

Yuki Nakayama and Atsushi Fujii

16:55–17:20

A fundamental issue in opinion mining is to search a corpus for opinion units, each of which typically comprises the evaluation by an author for a target object from an aspect, such as "This hotel is in a good location". However, few attempts have been made to address cases where the validity of an evaluation is restricted on a condition in the source text, such as "for traveling with small kids". In this paper, we propose a method to extract condition-opinion relations from online reviews, which enables fine-grained analysis for the utility of target objects depending the user attribute, purpose, and situation. Our method uses supervised machine learning to identify sequences of words or phrases that comprise conditions for opinions. We propose several features associated with lexical and syntactic information, and show their effectiveness experimentally.

**Session 3B: Semantics (Long +TACL Papers)**

Small Auditorium

Chair: Regina Barzilay

**A large annotated corpus for learning natural language inference**

*Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning* 15:40-16:05

Understanding entailment and contradiction is fundamental to understanding natural language, and inference about entailment and contradiction is a valuable testing ground for the development of semantic representations. However, machine learning research in this area has been dramatically limited by the lack of large-scale resources. To address this, we introduce the Stanford Natural Language Inference corpus, a new, freely available collection of labeled sentence pairs, written by humans doing a novel grounded task based on image captioning. At 570K pairs, it is two orders of magnitude larger than all other resources of its type. This increase in scale allows lexicalized classifiers to outperform some sophisticated existing entailment models, and it allows a neural network-based model to perform competitively on natural language inference benchmarks for the first time.

**Question-Answer Driven Semantic Role Labeling: Using Natural Language to Annotate Natural Language**

*Luheng He, Mike Lewis, and Luke Zettlemoyer* 16:05-16:30

This paper introduces the task of question-answer driven semantic role labeling (QA-SRL), where question-answer pairs are used to represent predicate-argument structure. For example, the verb “introduce” in the previous sentence would be labeled with the questions “What is introduced?”, and “What introduces something?”, each paired with the phrase from the sentence that gives the correct answer. Posing the problem this way allows the questions themselves to define the set of possible roles, without the need for predefined frame or thematic role ontologies. It also allows for scalable data collection by annotators with very little training and no linguistic expertise. We gather data in two domains, newswire text and Wikipedia articles, and introduce simple classifier-based models for predicting which questions to ask and what their answers should be. Our results show that non-expert annotators can produce high quality QA-SRL data, and also establish baseline performance levels for future work on this task.

**[TACL] It's All Fun and Games until Someone Annotates: Video Games with a Purpose for Linguistic Annotation.**

*David Jurgens and RobertoNavigli* 16:30-16:55

Annotated data is prerequisite for many NLP applications. Acquiring large-scale annotated corpora is a major bottleneck, requiring significant time and resources. Recent work has proposed turning annotation into a game to increase its appeal and lower its cost; however, current games are largely text-based and closely resemble traditional annotation tasks. We propose a new linguistic annotation paradigm that produces annotations from playing graphical video games. The effectiveness of this design is demonstrated using two video games: one to create a mapping from WordNet senses to images, and a second game that performs Word Sense Disambiguation. Both games produce accurate results. The first game yields annotation quality equal to that of experts and a cost reduction of 73% over equivalent crowdsourcing; the second game provides a 16.3% improvement in accuracy over current state-of-the-art sense disambiguation games with WordNet.

**[TACL] Semantic Proto-Roles**

*Drew Reisinger, Rachel Rudinger, Francis Ferraro, Kyle Rawlins, and Benjamin Van Durme* 16:55-17:20

We present the first large-scale, corpus based verification of Dowty's seminal theory of proto-roles. Our results demonstrate both the need for and the feasibility of a property-based annotation scheme of semantic relationships, as opposed to the currently dominant notion of categorical roles.

## Session 3C: Information Retrieval and Question Answering (Long Papers)

Room 2

Chair: Scott Wen-tau Yih

### Name List Only? Target Entity Disambiguation in Short Texts

Yixin Cao, Juanzi Li, Xiaofei Guo, Shuanhu Bai, Heng Ji, and Jie Tang

15:40–16:05

Target entity disambiguation (TED), the task of identifying target entities of the same domain, has been recognized as a critical step in various important applications. In this paper, we propose a graph-based model called TremenRank to collectively identify target entities in short texts given a name list only. TremenRank propagates trust within the graph, allowing for an arbitrary number of target entities and texts using inverted index technology. Furthermore, we design a multi-layer directed graph to assign different trust levels to short texts for better performance. The experimental results demonstrate that our model outperforms state-of-the-art methods with an average gain of 24.8% in accuracy and 15.2% in the F1-measure on three datasets in different domains.

### Biography-Dependent Collaborative Entity Archiving for Slot Filling

Yu Hong, Xiaobin Wang, Yadong Chen, Jian Wang, Tongtao Zhang, and Heng Ji 16:05–16:30

Current studies on Knowledge Base Population (KBP) tasks, such as slot filling, show the particular importance of entity-oriented automatic relevant document acquisition. Richer, diverse and reliable relevant documents satisfy the fundamental requirement that a KBP system explores the attributes of an entity, such as provenance-based background knowledge extraction (e.g., a person's religion, origin, etc.). Towards the bottleneck problem between comprehensiveness and definiteness of acquisition, we propose a fuzzy-to-exact matching based collaborative archiving method. In particular we introduce topic modeling methodologies into entity profiling, so as to build a bridge between fuzzy and exact matching. On one side of the bridge, we employ the topics in a small-scale set of high-quality relevant documents (i.e., exact matching results) to summarize the life slices of a target entity (i.e., so-called biography). On the other side, we use the biography as a reliable reference material to detect new truly relevant documents from a large-scale semi-finished pseudo-feedback (i.e., fuzzy matching results). We leverage the archiving method in state-of-the-art slot filling systems. Experiments on TAC-KBP data show significant improvement.

### Stochastic Top-k ListNet

Tianyi Luo, Dong Wang, Rong Liu, and Yiqiao Pan

16:30–16:55

ListNet is a well-known listwise learning to rank model and has gained much attention in recent years. A particular problem of ListNet, however, is the high computation complexity in model training, mainly due to the large number of object permutations involved in computing the gradients. This paper proposes a stochastic ListNet approach which computes the gradient within a bounded permutation subset. It significantly reduces the computation complexity of model training and allows extension to Top-k models, which is impossible with the conventional implementation based on full-set permutations. Meanwhile, the new approach utilizes partial ranking information of human labels, which helps improve model quality. Our experiments demonstrated that the stochastic ListNet method indeed leads to better ranking performance and speeds up the model training remarkably.

### Exploring Markov Logic Networks for Question Answering

Tushar Khot, Niranjan Balasubramanian, Eric Gribkoff, Ashish Sabharwal, Peter Clark, and Oren Etzioni 16:55–17:20

Elementary-level science exams pose significant knowledge acquisition and reasoning challenges for automatic question answering. We develop a system that reasons with knowledge derived from textbooks, represented in a subset of first-order logic. Automatic extraction, while scalable, often results in knowledge that is incomplete and noisy, motivating use of reasoning mechanisms that handle uncertainty. Markov Logic Networks (MLNs) seem a natural model for expressing such knowledge, but the exact way of leveraging MLNs is by no means obvious. We investigate three ways of applying MLNs to our task. First, we simply use the extracted science rules directly as MLN clauses and exploit the structure present in hard constraints to improve tractability. Second, we interpret science rules as describing prototypical entities, resulting in a drastically simplified but brittle network. Our third approach, called Praline, uses MLNs to align lexical elements as well as define and control how inference should be performed in this task. Praline demonstrates a 15% accuracy boost and a 10x reduction in runtime as compared to other MLN-based methods, and comparable accuracy to word-based baseline approaches.

**Session 3D: Long + TACL Paper Posters**

Lower Level Foyer

Chair: Einat Minkov

15:40–17:20

**Language and Domain Independent Entity Linking with Quantified Collective Validation***Han Wang, Jin Guang Zheng, Xiaogang Ma, Peter Fox, and Heng Ji*

Linking named mentions detected in a source document to an existing knowledge base provides disambiguated entity referents for the mentions. This allows better document analysis, knowledge extraction and knowledge base population. Most of the previous research extensively exploited the linguistic features of the source documents in a supervised or semi-supervised way. These systems therefore cannot be easily applied to a new language or domain. In this paper, we present a novel unsupervised algorithm named Quantified Collective Validation that avoids excessive linguistic analysis on the source documents and fully leverages the knowledge base structure for the entity linking task. We show our approach achieves state-of-the-art English entity linking performance and demonstrate successful deployment in a new language (Chinese) and two new domains (Biomedical and Earth Science). All the experiment datasets and system demonstration are available at [http://tw.rpi.edu/web/doc/hanwang\\_emnlp\\_2015](http://tw.rpi.edu/web/doc/hanwang_emnlp_2015) for research purpose.

**Modeling Relation Paths for Representation Learning of Knowledge Bases***Yankai Lin, Zhiyuan Liu, Huanbo Luan, Maosong Sun, Siwei Rao, and Song Liu*

Representation learning of knowledge bases (KBs) aims to embed both entities and relations into a low-dimensional space. Most existing methods only consider direct relations in representation learning. We argue that multiple-step relation paths also contain rich inference patterns between entities, and propose a path-based representation learning model. This model considers relation paths as translations between entities for representation learning, and addresses two key challenges: (1) Since not all relation paths are reliable, we design a path-constraint resource allocation algorithm to measure the reliability of relation paths. (2) We represent relation paths via semantic composition of relation embeddings. Experimental results on real-world datasets show that, as compared with baselines, our model achieves significant and consistent improvements on knowledge base completion and relation extraction from text.

**Corpus-level Fine-grained Entity Typing Using Contextual Information***Yadollah Yaghoobzadeh and Hinrich Schütze*

This paper addresses the problem of corpus-level entity typing, i.e., inferring from a large corpus that an entity is a member of a class such as “food” or “artist”. The application of entity typing we are interested in is knowledge base completion, specifically, to learn which classes an entity is a member of. We propose FIGMENT to tackle this problem. FIGMENT is embedding-based and combines (i) a global model that scores based on aggregated contextual information of an entity and (ii) a context model that first scores the individual occurrences of an entity and then aggregates the scores. In our evaluation, FIGMENT strongly outperforms an approach to entity typing that relies on relations obtained by an open information extraction system.

**Knowledge Base Unification via Sense Embeddings and Disambiguation***Claudio Delli Bovi, Luis Espinosa Anke, and Roberto Navigli*

We present KB-Unify, a novel approach for integrating the output of different Open Information Extraction systems into a single unified and fully disambiguated knowledge repository. KB-Unify consists of three main steps: (1) disambiguation of relation argument pairs via a sense-based vector representation and a large unified sense inventory; (2) ranking of semantic relations according to their degree of specificity; (3) cross-resource relation alignment and merging based on the semantic similarity of domains and ranges. We tested KB-Unify on a set of four heterogeneous knowledge bases, obtaining high-quality results. We discuss and provide evaluations at each stage, and release output and evaluation data for the use and scrutiny of the community.

**Open-Domain Name Error Detection using a Multi-Task RNN***Hao Cheng, Hao Fang, and Mari Ostendorf*

Out-of-vocabulary name errors in speech recognition create significant problems for downstream language processing, but the fact that they are rare poses challenges for automatic detection, particularly

in an open-domain scenario. To address this problem, a multi-task recurrent neural network language model for sentence-level name detection is proposed for use in combination with out-of-vocabulary word detection. The sentence-level model is also effective for leveraging external text data. Experiments show a 26% improvement in name-error detection F-score over a system using n-gram lexical features.

### **Extracting Relations between Non-Standard Entities using Distant Supervision and Imitation Learning**

*Isabelle Augenstein, Andreas Vlachos, and Diana Maynard*

Distantly supervised approaches have become popular in recent years as they allow training relation extractors without text-bound annotation, using instead known relations from a knowledge base and a large textual corpus from an appropriate domain. While state of the art distant supervision approaches use off-the-shelf named entity recognition and classification NERC systems to identify relation arguments, discrepancies in domain or genre between the data used for NERC training and the intended domain for the relation extractor can lead to low performance. This is particularly problematic for “non-standard” named entities such as album which would fall into the MISC category. We propose to ameliorate this issue by jointly training the named entity classifier and the relation extractor using imitation learning which reduces structured prediction learning to classification learning. We further experiment with Web features different features and compare against using two off-the-shelf supervised NERC systems, Stanford NER and FIGER, for named entity classification. Our experiments show that imitation learning improves average precision by 4 points over an one-stage classification model, while removing Web features results in a 6 points reduction. Compared to using FIGER and Stanford NER, average precision is 10 points and 19 points higher with our imitation learning approach.

### **Sieve-Based Spatial Relation Extraction with Expanding Parse Trees**

*Jennifer D’Souza and Vincent Ng*

Spatial relation extraction is the under-investigated task of identifying the relations on spatial elements. A key challenge introduced by the recent SpaceEval shared task on spatial relation extraction is the identification of MOVELINKs, a type of spatial relation in which up to eight spatial elements can participate. To handle the complexity of extracting MOVELINKs, we combine two ideas that have been successfully applied to information extraction tasks, namely tree kernels and multi-pass sieves, proposing the use of an expanding parse tree as a novel structured feature for training MOVELINK classifiers. Our approach yields state-of-the-art results on two key subtasks in SpaceEval.

### **[TACL] Cross-Document Co-Reference Resolution using Sample-Based Clustering with Knowledge Enrichment**

*Sourav Dutta and Gerhard Weikum*

Identifying and linking named entities across information sources is the basis of knowledge acquisition and at the heart of Web search, recommendations, and analytics. An important problem in this context is cross-document co-reference resolution (CCR): computing equivalence classes of textual mentions denoting the same entity, within and across documents. Prior methods employ ranking, clustering, or probabilistic graphical models using syntactic features and distant features from knowledge bases. However, these methods exhibit limitations regarding run-time and robustness. This paper presents the CROCS framework for unsupervised CCR, improving the state of the art in two ways. First, we extend the way knowledge bases are harnessed, by constructing a notion of semantic summaries for intra-document co-reference chains using co-occurring entity mentions belonging to different chains. Second, we reduce the computational cost by a new algorithm that embeds sample-based bisection, using spectral clustering or graph partitioning, in a hierarchical clustering process. This allows scaling up CCR to large corpora. Experiments with three datasets show significant gains in output quality, compared to the best prior methods, and the run-time efficiency of CROCS.

### **[TACL] Combining Minimally-supervised Methods for Arabic Named Entity Recognition**

*Maha Althobaiti, Udo Kruschwitz, and Massimo Poesio*

Supervised methods can achieve high performance on NLP tasks, such as Named Entity Recognition (NER), but new annotations are required for every new domain and/or genre change. This has motivated research in minimally supervised methods such as semisupervised learning and distant learning, but neither technique has yet achieved performance levels comparable to those of supervised methods. Semi-supervised methods tend to have very high precision but comparatively low recall, whereas distant

learning tends to achieve higher recall but lower precision. This complementarity suggests that better results may be obtained by combining the two types of minimally supervised methods. In this paper we present a novel approach to Arabic NER using a combination of semi-supervised and distant learning techniques. We trained a semi-supervised NER classifier and another one using distant learning techniques, and then combined them using a variety of classifier combination schemes, including the Bayesian Classifier Combination (BCC) procedure recently proposed for sentiment analysis. According to our results, the BCC model leads to an increase in performance of 8 percentage points over the best base classifiers.

**Session 3E: Short Paper Posters**

Lower Level Foyer

Chair: Wei Xu

15:40–17:20

**Mr. Bennet, his coachman, and the Archbishop walk into a bar but only one of them gets recognized: On The Difficulty of Detecting Characters in Literary Texts***Hardik Vala, David Jurgens, Andrew Piper, and Derek Ruths*

Characters are fundamental to literary analysis. Current approaches are heavily reliant on NER to identify characters, causing many to be overlooked. We propose a novel technique for character detection, achieving significant improvements over state of the art on multiple datasets.

**Convolutional Sentence Kernel from Word Embeddings for Short Text Categorization***Jonghoon Kim, Francois Rousseau, and Michalis Vazirgiannis*

This paper introduces a convolutional sentence kernel based on word embeddings. Our kernel overcomes the sparsity issue that arises when classifying short documents or in case of little training data. Experiments on six sentence datasets showed statistically significant higher accuracy over the standard linear kernel with n-gram features and other proposed models.

**Predicting the Structure of Cooking Recipes***Jermsak Jermsurawong and Nizar Habash*

Cooking recipes exist in abundance; but due to their unstructured text format, they are hard to study quantitatively beyond treating them as simple bags of words. In this paper, we propose an ingredient-instruction dependency tree data structure to represent recipes. The proposed representation allows for more refined comparison of recipes and recipe-parts, and is a step towards semantic representation of recipes. Furthermore, we build a parser that maps recipes into the proposed representation. The parser's edge prediction accuracy of 93.5% improves over a strong baseline of 85.7% (54.5% error reduction).

**TSDPMM: Incorporating Prior Topic Knowledge into Dirichlet Process Mixture Models for Text Clustering***Linmei Hu, Juanzi Li, Xiaoli Li, Chao Shao, and Xuzhong Wang*

Dirichlet process mixture model (DPMM) has great potential for detecting the underlying structure of data. Extensive studies have applied it for text clustering in terms of topics. However, due to the unsupervised nature, the topic clusters are always less satisfactory. Considering that people often have some prior knowledge about which potential topics should exist in given data, we aim to incorporate such knowledge into the DPMM to improve text clustering. We propose a novel model TSDPMM based on a new seeded Pólya urn scheme. Experimental results on document clustering across three datasets demonstrate our proposed TSDPMM significantly outperforms state-of-the-art DPMM model and can be applied in a lifelong learning framework.

**Sentence Modeling with Gated Recursive Neural Network***Xinchi Chen, Xipeng Qiu, Chenxi Zhu, Shiyu Wu, and Xuanjing Huang*

Recently, neural network based sentence modeling methods have achieved great progress. Among these methods, the recursive neural networks (RecNNs) can effectively model the combination of the words in sentence. However, RecNNs need a given external topological structure, like syntactic tree. In this paper, we propose a gated recursive neural network (GRNN) to model sentences, which employs a full binary tree (FBT) structure to control the combinations in recursive structure. By introducing two kinds of gates, our model can better model the complicated combinations of features. Experiments on three text classification datasets show the effectiveness of our model.

**Learning Timeline Difference for Text Categorization***Fumiyo Fukumoto and Yoshimi Suzuki*

This paper addresses text categorization problem that training data may derive from a different time period from the test data. We present a learning framework which extends a boosting technique to learn accurate model for timeline adaptation. The results showed that the method was comparable to the current state-of-the-art biased-SVM method, especially the method is effective when the creation time period of the test data differs greatly from the training data.

## **Summarizing Topical Contents from PubMed Documents Using a Thematic Analysis**

*Sun Kim, Lana Yeganova, and W John Wilbur*

Improving the search and browsing experience in PubMed is a key component in helping users detect information of interest. In particular, when exploring a novel field, it is important to provide a comprehensive view for a specific subject. One solution for providing this panoramic picture is to find sub-topics from a set of documents. We propose a method that finds sub-topics that we refer to as themes and computes representative titles based on a set of documents in each theme. The method combines a thematic clustering algorithm and the Pool Adjacent Violators algorithm to induce significant themes. Then, for each theme, a title is computed using PubMed document titles and theme-dependent term scores. We tested our system on five disease sets from OMIM and evaluated the results based on normalized pointwise mutual information and MeSH terms. For both performance measures, the proposed approach outperformed LDA. The quality of theme titles were also evaluated by comparing them with manually created titles.

## **Recognizing Biographical Sections in Wikipedia**

*Alessio Palmero Aprosio and Sara Tonelli*

Wikipedia is the largest collection of encyclopedic data ever written in the history of humanity. Thanks to its coverage and its availability in machine-readable format, it has become a primary resource for large-scale research in historical and cultural studies. In this work, we focus on the subset of pages describing persons, and we investigate the task of recognizing biographical sections from them: given a person's page, we identify the list of sections where information about her/his life is present. We model this as a sequence classification problem, and propose a supervised setting, in which the training data are acquired automatically. Besides, we show that six simple features extracted only from the section titles are very informative and yield good results well above a strong baseline.

## **Learn to Solve Algebra Word Problems Using Quadratic Programming**

*Lipu Zhou, Shuaixiang Dai, and Liwei Chen*

This paper presents a new algorithm to automatically solve algebra word problems. Our algorithm solves a word problem via analyzing a hypothesis space containing all possible equation systems generated by assigning the numbers in the word problem into a set of equation system templates extracted from the training data. To obtain a robust decision surface, we train a log-linear model to make the margin between the correct assignments and the false ones as large as possible. This results in a quadratic programming (QP) problem which can be efficiently solved. Experimental results show that our algorithm achieves 79.7% accuracy, about 10% higher than the state-of-the-art baseline (Kushman et al., 2014)

## **An Unsupervised Method for Discovering Lexical Variations in Roman Urdu Informal Text**

*Abdul Rafee, Abdul Qayyum, Muhammad Moeenuddin, Asim Karim, Hassan Sajjad, and Faisal Kamiran*

We present an unsupervised method to find lexical variations in Roman Urdu informal text. Our method includes a phonetic algorithm UrduPhone, a featurebased similarity function, and a clustering algorithm Lex-C. UrduPhone encodes roman Urdu strings to their phonetic equivalent representations. This produces an initial grouping of different spelling variations of a word. The similarity function incorporates word features and their context. Lex-C is a variant of k-medoids clustering algorithm that group lexical variations. It incorporates a similarity threshold to balance the number of clusters and their maximum similarity. We test our system on two datasets of SMS and blogs and show an f-measure gain of up to 12% from baseline systems.

## **Component-Enhanced Chinese Character Embeddings**

*Yanran Li, Wenjie Li, Fei Sun, and Sujian Li*

Distributed word representations are very useful for capturing semantic information and have been successfully applied in many NLP tasks, especially on English. In this work, we innovatively develop two component-enhanced Chinese character embedding models and their bi-gram extensions. Distinguished from English word embeddings, our models explore the compositions of Chinese characters, which often serve as semantic indicators inherently. The evaluations on both word similarity and text classification demonstrate the effectiveness of our models.

**Multi-label Text Categorization with Joint Learning Predictions-as-Features Method**

*Li Li, Houfeng Wang, Xu Sun, Baobao Chang, Shi Zhao, and Lei Sha*

Multi-label text categorization is a type of text categorization, where each document is assigned to one or more categories. Recently, a series of methods have been developed, which train a classifier for each label, organize the classifiers in a partially ordered structure and take predictions produced by the former classifiers as the latter classifiers' features. These predictions-as-features style methods model high order label dependencies and obtain high performance. Nevertheless, the predictions-as-features methods suffer a drawback. When training a classifier for one label, the predictions-as-features methods can model dependencies between former labels and the current label, but they can't model dependencies between the current label and the latter labels. To address this problem, we propose a novel joint learning algorithm that allows the feedbacks to be propagated from the classifiers for latter labels to the classifier for the current label. We conduct experiments using real-world textual data sets, and these experiments illustrate the predictions-as-features models trained by our algorithm outperform the original models.

**A Framework for Comparing Groups of Documents**

*Arun Maiya*

We present a general framework for comparing multiple groups of documents. A bipartite graph model is proposed where document groups are represented as one node set and the comparison criteria are represented as the other node set. Using this model, we present basic algorithms to extract insights into similarities and differences among the document groups. Finally, we demonstrate the versatility of our framework through an analysis of NSF funding programs for basic research.

## Main Conference: Sunday, September 20

### Overview

---

07:30–18:00	<b>Registration</b>	<i>Culturgest Foyers</i>													
08:00–09:00	<b>Morning Coffee</b>	<i>Foyers</i>													
09:00–10:00	<b>Session P2: Plenary Session</b>	<i>Main Auditorium</i>													
	<i>Invited Talk: Measuring How Elected Officials and Constituents Communicate (Justin Grimmer)</i>														
10:00–10:30	<b>Coffee break</b>	<i>Foyers</i>													
10:30–12:10	<b>Session 4</b>	<table border="1"> <tbody> <tr> <td>Information Extraction (Long Papers)</td> <td>Statistical Models and Machine Learning Methods (Long Papers)</td> <td>Discourse (Long +TACL Papers)</td> <td>Long Paper Posters</td> </tr> <tr> <td><i>Main Auditorium</i></td> <td><i>Small Auditorium</i></td> <td><i>Room 2</i></td> <td><i>Short Paper Posters</i></td> </tr> <tr> <td></td> <td></td> <td></td> <td><i>Lower Level Foyer</i></td> </tr> </tbody> </table>		Information Extraction (Long Papers)	Statistical Models and Machine Learning Methods (Long Papers)	Discourse (Long +TACL Papers)	Long Paper Posters	<i>Main Auditorium</i>	<i>Small Auditorium</i>	<i>Room 2</i>	<i>Short Paper Posters</i>				<i>Lower Level Foyer</i>
Information Extraction (Long Papers)	Statistical Models and Machine Learning Methods (Long Papers)	Discourse (Long +TACL Papers)	Long Paper Posters												
<i>Main Auditorium</i>	<i>Small Auditorium</i>	<i>Room 2</i>	<i>Short Paper Posters</i>												
			<i>Lower Level Foyer</i>												
12:10–12:50	<b>Lunch</b>														
12:50–13:30	<b>Session P3: SIGDAT business meeting</b>	<i>Main Auditorium</i>													
13:30–15:10	<b>Session 5</b>	<table border="1"> <tbody> <tr> <td>Text Mining and NLP Applications (Long + TACL Papers)</td> <td>Semantics (Long +TACL Papers)</td> <td>Phonology and Word Segmentation (Long Papers)</td> <td>Long Paper Posters</td> </tr> <tr> <td><i>Main Auditorium</i></td> <td><i>Small Auditorium</i></td> <td><i>Room 2</i></td> <td><i>Short Paper Posters</i></td> </tr> <tr> <td></td> <td></td> <td></td> <td><i>Lower Level Foyer</i></td> </tr> </tbody> </table>		Text Mining and NLP Applications (Long + TACL Papers)	Semantics (Long +TACL Papers)	Phonology and Word Segmentation (Long Papers)	Long Paper Posters	<i>Main Auditorium</i>	<i>Small Auditorium</i>	<i>Room 2</i>	<i>Short Paper Posters</i>				<i>Lower Level Foyer</i>
Text Mining and NLP Applications (Long + TACL Papers)	Semantics (Long +TACL Papers)	Phonology and Word Segmentation (Long Papers)	Long Paper Posters												
<i>Main Auditorium</i>	<i>Small Auditorium</i>	<i>Room 2</i>	<i>Short Paper Posters</i>												
			<i>Lower Level Foyer</i>												
15:10–15:40	<b>Coffee break</b>	<i>Foyers</i>													
15:40–17:20	<b>Session 6</b>	<table border="1"> <tbody> <tr> <td>Machine Translation (Long Papers)</td> <td>Sentiment Analysis and Opinion Mining / Tagging, Chunking and Parsing (Long Papers)</td> <td>Language and Vision / Information Extraction (Long Papers)</td> <td>Long Paper Posters</td> </tr> <tr> <td><i>Main Auditorium</i></td> <td><i>Small Auditorium</i></td> <td><i>Room 2</i></td> <td><i>Short Paper Posters</i></td> </tr> <tr> <td></td> <td></td> <td></td> <td><i>Lower Level Foyer</i></td> </tr> </tbody> </table>		Machine Translation (Long Papers)	Sentiment Analysis and Opinion Mining / Tagging, Chunking and Parsing (Long Papers)	Language and Vision / Information Extraction (Long Papers)	Long Paper Posters	<i>Main Auditorium</i>	<i>Small Auditorium</i>	<i>Room 2</i>	<i>Short Paper Posters</i>				<i>Lower Level Foyer</i>
Machine Translation (Long Papers)	Sentiment Analysis and Opinion Mining / Tagging, Chunking and Parsing (Long Papers)	Language and Vision / Information Extraction (Long Papers)	Long Paper Posters												
<i>Main Auditorium</i>	<i>Small Auditorium</i>	<i>Room 2</i>	<i>Short Paper Posters</i>												
			<i>Lower Level Foyer</i>												
19:00–23:00	<b>Conference Dinner</b>	<i>Pateo Alfacinha</i>													

## **Invited Speaker: Justin Grimmer**

---

### **Measuring How Elected Officials and Constituents Communicate**

Sunday, September 20, 2015,

Main Auditorium

**Abstract:** This talk will show how elected officials use communication to cultivate support with constituents, how constituents express their views to elected officials, and why biases in both kinds of communication matter for political representation. To demonstrate the bias and its effects, I propose to use novel collections of political texts and new text as data methods. Using the new data and methods, I will show how the incentives of communication contribute to perceptions of an angry public and vitriolic politicians. Among elected officials, the ideologically extreme members of Congress disproportionately participate in policy debates, resulting in political debates that occur between the most extreme members of each party. Among constituents, the most ideologically extreme and angry voters disproportionately contact their member of Congress, creating the impression of a polarized and vitriolic public. The talk will explain how the findings help us to understand how representation occurs in American politics, while also explaining how computational tools can help address questions in the social sciences.

**Biography:** Justin Grimmer is an associate professor of political science at Stanford University. His research examines how representation occurs in American politics using new statistical methods. His first book *Representational Style in Congress: What Legislators Say and Why It Matters* (Cambridge University Press, 2013) shows how senators define the type of representation they provide constituents and how this affects constituents' evaluations and was awarded the 2014 Richard Fenn Prize. His second book *The Impression of Influence: How Legislator Communication and Government Spending Cultivate a Personal Vote* (Princeton University Press, 2014 with Sean J. Westwood and Solomon Messing) demonstrates how legislators ensure they receive credit for government actions. His work has appeared in the *American Political Science Review*, *American Journal of Political Science*, *Journal of Politics*, *Political Analysis*, *Proceedings of the National Academy of Sciences*, *Regulation and Governance*, and *Poetics*.

## Session 4 Overview

---

Track A	Track B	Track C	
<i>Information Extraction (Long Papers)</i> Main Auditorium	<i>Statistical Models and Machine Learning Methods (Long Papers)</i> Small Auditorium	<i>Discourse (Long +TACL Papers)</i> Room 2	
C3EL: A Joint Model for Cross-Document Co-Reference Resolution and Entity Linking <i>Dutta and Weikum</i>	How Much Information Does a Human Translator Add to the Original? <i>Zoph, Ghazvininejad, and Knight</i>	Discourse parsing for multi-party chat dialogues <i>Afantenos, Kow, Asher, and Perret</i>	10:30
Joint Mention Extraction and Classification with Mention Hypergraphs <i>Lu and Roth</i>	Hierarchical Recurrent Neural Network for Document Modeling <i>Lin, Liu, Yang, Li, Zhou, and Li</i>	Joint prediction in MST-style discourse parsing for argumentation mining <i>Peldszus and Stede</i>	10:55
FINET: Context-Aware Fine-Grained Named Entity Typing <i>Del Corro, Abujabal, Gemulla, and Weikum</i>	Auto-Sizing Neural Networks: With Applications to n-gram Language Models <i>Murray and Chiang</i>	[TACL] One Vector is Not Enough: Entity-Augmented Distributed Semantics for Discourse Relations <i>Ji and Eisenstein</i>	11:20
Joint Entity Recognition and Disambiguation <i>Luo, Huang, Lin, and Nie</i>	Dual Decomposition Inference for Graphical Models over Strings <i>Peng, Cotterell, and Eisner</i>	[TACL] Latent Structures for Coreference Resolution <i>Martschat and Strube</i>	11:45

**Track D: Long Paper Posters**

Lower Level Foyer

Chair: Sam Bowman

10:30–12:10

**Poster Cluster 1: Semantics (P1-8)**

- Feature-Rich Two-Stage Logistic Regression for Monolingual Alignment  
*M. A. Sultan, S. Bethard, and T. Sumner*
- Semantic Role Labeling with Neural Network Factors  
*N. FitzGerald, O. Täckström, K. Ganchev, and D. Das*
- RELLY: Inferring Hypernym Relationships Between Relational Phrases  
*A. Grycner, G. Weikum, J. Pujara, J. Foulds, and L. Getoor*
- Mise en Place: Unsupervised Interpretation of Instructional Recipes  
*C. Kiddon, G. T. Ponnuraj, L. Zettlemoyer, and Y. Choi*
- Semantic Framework for Comparison Structures in Natural Language  
*O. Bakhshandeh and J. Allen*

- Sarcastic or Not: Word Embeddings to Predict the Literal or Sarcastic Meaning of Words  
*D. Ghosh, W. Guo, and S. Muresan*
- Incorporating Trustiness and Collective Synonym/Contrastive Evidence into Taxonomy Construction  
*T. Luu Anh, J.-j. Kim, and S. K. Ng*
- Learning to Automatically Solve Logic Grid Puzzles  
*A. Mitra and C. Baral*

**Track E: Short Paper Posters**

Chair: Maja Popovic

Lower Level Foyer

10:30-12:10

**Poster Cluster 1: Machine Translation and Multilinguality (P1-13)**

- Improving fast\_align by Reordering  
*C. Ding, M. Utiyama, and E. Sumita*
- Touch-Based Pre-Post-Editing of Machine Translation Output  
*B. Marie and A. Max*
- A Discriminative Training Procedure for Continuous Translation Models  
*Q.-K. DO, A. Allauzen, and F. Yvon*
- System Combination for Machine Translation through Paraphrasing  
*W.-Y. Ma and K. McKeown*
- Hierarchical Incremental Adaptation for Statistical Machine Translation  
*J. Wuebker, S. Green, and J. DeNero*
- ReVal: A Simple and Effective Machine Translation Evaluation Metric Based on Recurrent Neural Networks  
*R. Gupta, C. Orasan, and J. van Genabith*
- Investigating Continuous Space Language Models for Machine Translation Quality Estimation  
*K. Shah, R. W. M. Ng, F. Bougares, and L. Specia*
- Supervised Phrase Table Triangulation with Neural Word Embeddings for Low-Resource Languages  
*T. Levinboim and D. Chiang*
- Translation Invariant Word Embeddings  
*K. Huang, M. Gardner, E. Papalexakis, C. Faloutsos, N. Sidiropoulos, T. Mitchell, P. P. Talukdar, and X. Fu*
- Hierarchical Phrase-based Stream Decoding  
*A. Finch, X. Wang, M. Utiyama, and E. Sumita*
- Rule Selection with Soft Syntactic Features for String-to-Tree Statistical Machine Translation  
*F. Braune, N. Seemann, and A. Fraser*

- Motivating Personality-aware Machine Translation  
*S. Mirkin, S. Nowson, C. Brun, and J. Perez*
- Trans-gram, Fast Cross-lingual Word-embeddings  
*J. Coulmance, J.-M. Marty, G. Wenzek, and A. Benhalloum*

**Poster Cluster 2: Computational Psycholinguistics (P14-16)**

- The Overall Markedness of Discourse Relations  
*L. Jin and M.-C. de Marneffe*
- Experiments in Open Domain Deception Detection  
*V. Pérez-Rosas and R. Mihalcea*
- A model of rapid phonotactic generalization  
*T. Linzen and T. O'Donnell*

## Abstracts: Session 4

---

### Session 4A: Information Extraction (Long Papers)

Main Auditorium

Chair: Heng Ji

#### C3EL: A Joint Model for Cross-Document Co-Reference Resolution and Entity Linking

*Sourav Dutta and Gerhard Weikum*

10:30–10:55

Cross-document co-reference resolution (CCR) computes equivalence classes over textual mentions denoting the same entity in a document corpus. Named-entity linking (NEL) disambiguates mentions onto entities present in a knowledge base (KB) or maps them to null if not present in the KB. Traditionally, CCR and NEL have been addressed separately. However, such approaches miss out on the mutual synergies if CCR and NEL were performed jointly. This paper proposes C3EL, an unsupervised framework combining CCR and NEL for jointly tackling both problems. C3EL incorporates results from the CCR stage into NEL, and vice versa: additional global context obtained from CCR improves the feature space and performance of NEL, while NEL in turn provides distant KB features for already disambiguated mentions to improve CCR. The CCR and NEL steps are interleaved in an iterative algorithm that focuses on the highest-confidence still unresolved mentions in each iteration. Experimental results on two different corpora, news-centric and web-centric, demonstrate significant gains over state-of-the-art baselines for both CCR and NEL.

#### Joint Mention Extraction and Classification with Mention Hypergraphs

*Wei Lu and Dan Roth*

10:55–11:20

We present a novel model for the task of joint mention extraction and classification. Unlike existing approaches, our model is able to effectively capture overlapping mentions whose lengths are unbounded. Our model is highly scalable, with a time complexity that is linear in the number of words in the input sentence and linear in the number of possible mention classes. The model can be extended to additionally capture mention heads explicitly in a joint manner under the same time complexity. We demonstrate the effectiveness of our model through extensive experiments on standard datasets.

#### FINET: Context-Aware Fine-Grained Named Entity Typing

*Luciano Del Corro, Abdalghani Abujabal, Rainer Gemulla, and Gerhard Weikum* 11:20–11:45

We propose FINET, a system for detecting the types of named entities in short inputs—such as sentences or tweets—with respect to WordNet’s super fine-grained type system. FINET generates candidate types using a sequence of multiple extractors, ranging from explicitly mentioned types to implicit types, and subsequently selects the most appropriate using ideas from word-sense disambiguation. FINET combats data scarcity and noise from existing systems: It does not rely on supervision in its extractors and generates training data for type selection from WordNet and other resources. FINET supports the most fine-grained type system so far, including types with no annotated training data. Our experiments indicate that FINET outperforms state-of-the-art methods in terms of recall, precision, and granularity of extracted types.

#### Joint Entity Recognition and Disambiguation

*Gang Luo, Xiaojiang Huang, Chin-Yew Lin, and Zaiqing Nie*

11:45–12:10

Extracting named entities in text and linking extracted names to a given knowledge base are fundamental tasks in applications for text understanding. Existing systems typically run a named entity recognition (NER) model to extract entity names first, then run an entity linking model to link extracted names to a knowledge base. NER and linking models are usually trained separately, and the mutual dependency between the two tasks is ignored. We propose JERL, Joint Entity Recognition and Linking, to jointly model NER and linking tasks and capture the mutual dependency between them. It allows the information from each task to improve the performance of the other. To the best of our knowledge, JERL is the first model to jointly optimize NER and linking tasks together completely. In experiments on the CoNLL’03/AIDA data set, JERL outperforms state-of-art NER and linking systems, and we find improvements of 0.4% absolute F1 for NER on CoNLL’03, and 0.36% absolute precision1 for linking on AIDA.

## Session 4B: Statistical Models and Machine Learning Methods (Long Papers)

Small Auditorium

Chair: Chris Dyer

### How Much Information Does a Human Translator Add to the Original?

Barret Zoph, Marjan Ghazvininejad, and Kevin Knight

10:30–10:55

We ask how much information a human translator adds to an original text, and we provide a bound. We address this question in the context of bilingual text compression: given a source text, how many bits of additional information are required to specify the target text produced by a human translator? We develop new compression algorithms and establish a benchmark task.

### Hierarchical Recurrent Neural Network for Document Modeling

Rui Lin, Shujie Liu, Muyun Yang, Mu Li, Ming Zhou, and Sheng Li

10:55–11:20

This paper proposes a novel hierarchical recurrent neural network language model (HRNNLM) for document modeling. After establishing a RNN to capture the coherence between sentences in a document, HRNNLM integrates it as the sentence history information into the word level RNN to predict the word sequence with cross-sentence contextual information. A two-step training approach is designed, in which sentence-level and word-level language models are approximated for the convergence in a pipeline style. Examined by the standard sentence ordering scenario, HRNNLM is proved for its better accuracy in modeling the sentence coherence. And at the word level, experimental results also indicate a significant lower model perplexity, followed by a practical better translation result when applied to a Chinese-English document translation reranking task.

### Auto-Sizing Neural Networks: With Applications to n-gram Language Models

Kenton Murray and David Chiang

11:20–11:45

Neural networks have been shown to improve performance across a range of natural-language tasks while addressing some issues with traditional models such as size. However, designing and training them can be complicated. Frequently, researchers resort to repeated experimentation across a range of parameters to pick optimal settings. In this paper, we address the issue of choosing the correct number of units in the hidden layers. We introduce a method for automatically adjusting network size by pruning out hidden units through  $\ell_\infty$ , 1 and  $\ell_2$ , 1 regularization. We apply this method to language modeling and demonstrate its ability to correctly choose the number of hidden units while maintaining perplexity. We also include these models in a machine translation decoder and show that these smaller neural models maintain the significant improvements of their unpruned versions.

### Dual Decomposition Inference for Graphical Models over Strings

Nanyun Peng, Ryan Cotterell, and Jason Eisner

11:45–12:10

We investigate dual decomposition for joint MAP inference of many strings. Given an arbitrary graphical model, we decompose it into small acyclic sub-models, whose MAP configurations can be found by finite-state composition and dynamic programming. We force the solutions of these subproblems to agree on overlapping variables, by tuning Lagrange multipliers for an adaptively expanding set of variable-length n-gram count features. This is the first inference method for arbitrary graphical models over strings that does not require approximations such as random sampling, message simplification, or a bound on string length. Provided that the inference method terminates, it gives a certificate of global optimality (though MAP inference in our setting is undecidable in general). On our global phonological inference problems, it always terminates, and achieves more accurate results than max-product and sum-product loopy belief propagation.

**Session 4C: Discourse (Long +TACL Papers)**

Room 2

Chair: Pascal Denis

**Discourse parsing for multi-party chat dialogues**

Stergos Afantinos, Eric Kow, Nicholas Asher, and Jérémie Perret

10:30-10:55

In this paper we present the first ever, to the best of our knowledge, discourse parser for multi-party chat dialogues. Discourse in multi-party dialogues dramatically differs from monologues since threaded conversations are commonplace rendering prediction of the discourse structure compelling. Moreover, the fact that our data come from chats renders the use of syntactic and lexical information useless since people take great liberties in expressing themselves lexically and syntactically. We use the dependency parsing paradigm as has been done in the past (Muller et al., 2012; Li et al., 2014). We learn local probability distributions and then use MST for decoding. We achieve 0.680 F\_1 on unlabelled structures and 0.516 F\_1 on fully labeled structures which is better than many state of the art systems for monologues, despite the inherent difficulties that multi-party chat dialogues have.

**Joint prediction in MST-style discourse parsing for argumentation mining**

Andreas Peldszus and Manfred Stede

10:55-11:20

We introduce a new approach to argumentation mining that we applied to a parallel German/English corpus of short texts annotated with argumentation structure. We focus on structure prediction, which we break into a number of subtasks: relation identification, central claim identification, role classification, and function classification. Our new model jointly predicts different aspects of the structure by combining the different subtask predictions in the edge weights of an evidence graph; we then apply a standard MST decoding algorithm. This model not only outperforms two reasonable baselines and two data-driven models of global argument structure for the difficult subtask of relation identification, but also improves the results for central claim identification and function classification and it compares favorably to a complex mstparser pipeline.

**[TACL] One Vector is Not Enough: Entity-Augmented Distributed Semantics for Discourse Relations**

Yangfeng Ji and Jacob Eisenstein

11:20-11:45

Discourse relations bind smaller linguistic units into coherent texts. Automatically identifying discourse relations is difficult, because it requires understanding the semantics of the linked arguments. A more subtle challenge is that it is not enough to represent the meaning of each argument of a discourse relation, because the relation may depend on links between lower-level components, such as entity mentions. Our solution computes distributed meaning representations for each discourse argument by composition up the syntactic parse tree. We also perform a downward compositional pass to capture the meaning of coreferent entity mentions. Implicit discourse relations are then predicted from these two representations, obtaining substantial improvements on the Penn Discourse Treebank.

**[TACL] Latent Structures for Coreference Resolution**

Sebastian Martschat and Michael Strube

11:45-12:10

Machine learning approaches to coreference resolution vary greatly in the modeling of the problem: while early approaches operated on the mention pair level, current research focuses on ranking architectures and antecedent trees. We propose a unified representation of different approaches to coreference resolution in terms of the structure they operate on. We represent several coreference resolution approaches proposed in the literature in our framework and evaluate their performance. Finally, we conduct a systematic analysis of the output of these approaches, highlighting differences and similarities.

**Session 4D: Long Paper Posters**

Lower Level Foyer

Chair: Sam Bowman

10:30–12:10

**Feature-Rich Two-Stage Logistic Regression for Monolingual Alignment***Md Arafat Sultan, Steven Bethard, and Tamara Sumner*

Monolingual alignment is the task of pairing semantically similar units from two pieces of text. We report a top-performing supervised aligner that operates on short text snippets. We employ a large feature set to (1) encode similarities among semantic units (words and named entities) in context, and (2) address cooperation and competition for alignment among units in the same snippet. These features are deployed in a two-stage logistic regression framework for alignment. On two benchmark data sets, our aligner achieves F1 scores of 92.1% and 88.5%, with statistically significant error reductions of 4.8% and 7.3% over the previous best aligner. It produces top results in extrinsic evaluation as well.

**Semantic Role Labeling with Neural Network Factors***Nicholas FitzGerald, Oscar Täckström, Kuzman Ganchev, and Dipanjan Das*

We present a new method for semantic role labeling in which arguments and semantic roles are jointly embedded in a shared vector space for a given predicate. These embeddings belong to a neural network, whose output represents the potential functions of a graphical model designed for the SRL task. We consider both local and structured learning methods and obtain state-of-the-art results on standard PropBank and FrameNet corpora with a straightforward product-of-experts model. We further show how the model can learn jointly from PropBank and FrameNet annotations to obtain additional improvements on the smaller FrameNet dataset.

**RELLY: Inferring Hypernym Relationships Between Relational Phrases***Adam Grycner, Gerhard Weikum, Jay Pujara, James Foulds, and Lise Getoor*

Relational phrases (e.g., “got married to”) and their hypernyms (e.g., “is a relative of”) are central for many tasks including question answering, open information extraction, paraphrasing, and entailment detection. This has motivated the development of several linguistic resources (e.g. DIRT, PATTY, and WiseNet) which systematically collect and organize relational phrases. These resources have demonstrable practical benefits, but are each limited due to noise, sparsity, or size. We present a new general-purpose method, RELLY, for constructing a large hypernymy graph of relational phrases with high-quality subsumptions using collective probabilistic programming techniques. Our graph induction approach integrates small high-precision knowledge bases together with large automatically curated resources, and reasons collectively to combine these resources into a consistent graph. Using RELLY, we construct a high-coverage, high-precision hypernymy graph consisting of 20K relational phrases and 35K hypernymy links. Our evaluation indicates a hypernymy link precision of 78%, and demonstrates the value of this resource for a document-relevance ranking task.

**Mise en Place: Unsupervised Interpretation of Instructional Recipes***Chloé Kiddon, Ganesh Thandavam Ponnuraj, Luke Zettlemoyer, and Yejin Choi*

We present an unsupervised hard EM approach to automatically mapping instructional recipes to action graphs, which define what actions should be performed on which objects and in what order. Recovering such structures can be challenging, due to unique properties of procedural language where, for example, verbal arguments are commonly elided when they can be inferred from context and disambiguation often requires world knowledge. Our probabilistic model incorporates aspects of procedural semantics and world knowledge, such as likely locations and selectional preferences for different actions. Experiments with cooking recipes demonstrate the ability to recover high quality action graphs, outperforming a strong sequential baseline by 8 points in F1, while also discovering general-purpose knowledge about cooking.

**Semantic Framework for Comparison Structures in Natural Language***Omid Bakhshandeh and James Allen*

Comparison is one of the most important phenomena in language for expressing objective and subjective facts about various entities. Systems that can understand and reason over comparative structure can play a major role in the applications which require deeper understanding of language. In this paper we present a novel semantic framework for representing the meaning of comparative structures in natural language, which models comparisons as predicate-argument pairs inter-connected with semantic roles.

Our framework supports not only adjectival, but also adverbial, nominal, and verbal comparatives. With this paper, we provide a novel dataset of gold-standard comparison structures annotated according to our semantic framework.

### **Sarcastic or Not: Word Embeddings to Predict the Literal or Sarcastic Meaning of Words**

*Debanjan Ghosh, Weiwei Guo, and Smaranda Muresan*

Sarcasm is generally characterized as a figure of speech that involves the substitution of a literal by a figurative meaning, which is usually the opposite of the original literal meaning. We re-frame the sarcasm detection task as a type of word sense disambiguation problem, where the sense of a word is either literal or sarcastic. We call this the Literal/Sarcastic Sense Disambiguation (LSSD) task. We address two issues: 1) how to collect a set of target words that can have either literal or sarcastic meanings depending on context; and 2) given an utterance and a target word, how to automatically detect whether the target word is used in the literal or the sarcastic sense. For the latter, we investigate several distributional semantics methods and show that a Support Vector Machines (SVM) classifier with a modified kernel using word embeddings achieves a 7-10% F1 improvement over a strong lexical baseline.

### **Incorporating Trustiness and Collective Synonym/Contrastive Evidence into Taxonomy Construction**

*Tuan Luu Anh, Jung-jae Kim, and See Kiong Ng*

Taxonomy plays an important role in many applications by organizing domain knowledge into a hierarchy of is-a relations between terms. Previous works on the taxonomic relation identification from text corpora lack in two aspects: 1) They do not consider the trustiness of individual source texts, which is important to filter out incorrect relations from unreliable sources. 2) They also do not consider collective evidence from synonyms and contrastive terms, where synonyms may provide additional supports to taxonomic relations, while contrastive terms may contradict them. In this paper, we present a method of taxonomic relation identification that incorporates the trustiness of source texts measured with such techniques as PageRank and knowledge-based trust, and the collective evidence of synonyms and contrastive terms identified by linguistic pattern matching and machine learning. The experimental results show that the proposed features can consistently improve performance up to 4%-10% of F-measure.

### **Learning to Automatically Solve Logic Grid Puzzles**

*Arindam Mitra and Chitta Baral*

Logic grid puzzle is a genre of logic puzzles in which we are given (in a natural language) a scenario, the object to be deduced and certain clues. The reader has to figure out the solution using the clues provided and some generic domain constraints. In this paper, we present a system, Logicia, that takes a logic grid puzzle and the set of elements in the puzzle and tries to solve it by translating it to the knowledge representation and reasoning language of Answer Set Programming (ASP) and then using an ASP solver. The translation to ASP involves extraction of entities and their relations from the clues. For that we use a novel learning based approach which uses varied supervision, including the entities present in a clue and the expected representation of a clue in ASP. Our system, Logicia, learns to automatically translate a clue with 81.11% accuracy and is able to solve 71% of the problems of a corpus. This is the first learning system that can solve logic grid puzzles described in natural language in a fully automated manner. The code and the data will be made publicly available at <http://bioai.lab.asu.edu/logicgridpuzzles>.

**Session 4E: Short Paper Posters**

Lower Level Foyer

Chair: Maja Popovic

10:30-12:10

**Improving fast\_align by Reordering***Chenchen Ding, Masao Utiyama, and Eiichiro Sumita*

`fast_align` is a simple, fast, and efficient approach for word alignment based on the IBM model 2. `fast_align` performs well for language pairs with relatively similar word orders; however, it does not perform well for language pairs with drastically different word orders. We propose a segmenting-reversing reordering process to solve this problem by alternately applying `fast_align` and reordering source sentences during training. Experimental results with Japanese-English translation demonstrate that the proposed approach improves the performance of `fast_align` significantly without the loss of efficiency. Experiments using other languages are also reported.

**Touch-Based Pre-Post-Editing of Machine Translation Output***Benjamin Marie and Aurélien Max*

We introduce pre-post-editing, possibly the most basic form of interactive translation, as a touch-based interaction with iteratively improved translation hypotheses prior to classical post-editing. We report simulated experiments that yield very large improvements on classical evaluation metrics (up to 21 BLEU) as well as on a parameterized variant of the TER metric that takes into account the cost of matching/touching tokens, confirming the promising prospects of the novel translation scenarios offered by our approach.

**A Discriminative Training Procedure for Continuous Translation Models***Quoc-Khanh DO, Alexandre Alauzen, and François Yvon*

Continuous-space translation models have recently emerged as extremely powerful ways to boost the performance of existing translation systems. A simple, yet effective way to integrate such models in inference is to use them in an  $n$ -best rescoring step. In this paper, we focus on this scenario and show that the performance gains in rescoring can be greatly increased when the neural network is trained jointly with all the other model parameters, using an appropriate objective function. Our approach is validated on two domains, where it outperforms strong baselines.

**System Combination for Machine Translation through Paraphrasing***Wei-Yun Ma and Kathleen McKeown*

In this paper, we propose a paraphrasing model to address the task of system combination for machine translation. We dynamically learn hierarchical paraphrases from target hypotheses and form a synchronous context-free grammar to guide a series of transformations of target hypotheses into fused translations. The model is able to exploit phrasal and structural system-weighted consensus and also to utilize existing information about word ordering present in the target hypotheses. In addition, to consider a diverse set of plausible fused translations, we develop a hybrid combination architecture, where we paraphrase every target hypothesis using different fusing techniques to obtain fused translations for each target, and then make the final selection among all fused translations. Our experimental results show that our approach can achieve a significant improvement over combination baselines.

**Hierarchical Incremental Adaptation for Statistical Machine Translation***Joern Wuebker, Spence Green, and John DeNero*

We present an incremental adaptation approach for statistical machine translation that maintains a flexible hierarchical domain structure within a single consistent model. Both weights and rules are updated incrementally on a stream of post-edits. Our multi-level domain hierarchy allows the system to adapt simultaneously towards local context at different levels of granularity, including genres and individual documents. Our experiments show consistent improvements in translation quality from all components of our approach.

**ReVal: A Simple and Effective Machine Translation Evaluation Metric Based on Recurrent Neural Networks***Rohit Gupta, Constantin Orasan, and Josef van Genabith*

Many state-of-the-art Machine Translation (MT) evaluation metrics are complex, involve extensive external resources (e.g. for paraphrasing) and require tuning to achieve best results. We present a simple

alternative approach based on dense vector spaces and recurrent neural networks (RNNs), in particular Long Short Term Memory (LSTM) networks. For WMT-14, our new metric scores best for two out of five language pairs, and overall best and second best on all language pairs, using Spearman and Pearson correlation, respectively. We also show how training data is computed automatically from WMT ranks data.

### **Investigating Continuous Space Language Models for Machine Translation Quality Estimation**

*Kashif Shah, Raymond W. M. Ng, Fethi Bougares, and Lucia Specia*

We present the novel features trained with a deep neural network for Machine Translation (MT) Quality Estimation (QE). The features are learned with a Continuous Space Language Model (CSLM) by estimating the probabilities of the source and target segments. These new features along with other available MT system-independent features are investigated on series of datasets with various quality labels for QE including Post Editing Effort (PEE) , Human Translation Edit Rate (HTER), Post Editing Time (PET) and METEOR score. The results show the significant improvements in predictions over the baseline as well as systems trained on previously available features set for all WMT QE tasks. More notably, the addition of newly proposed features beat previous best systems by a significant margin on official WMT12 and WMT14 post editing effort prediction tasks.

### **Supervised Phrase Table Triangulation with Neural Word Embeddings for Low-Resource Languages**

*Tomer Levinboim and David Chiang*

In this paper, we develop a supervised learning technique that improves noisy phrase translation scores obtained by phrase table triangulation. In particular, we extract word translation distributions from small amounts of source-target bilingual data (a dictionary or a parallel corpus) with which we learn to assign better scores to translation candidates obtained by triangulation. Our method is able to gain improvement in translation quality on two tasks: (1) On Malagasy-to-French translation via English, we use only 1k dictionary entries to gain +0.5 BLEU over triangulation. (2) On Spanish-to-French via English we use only 4k sentence pairs to gain +0.7 BLEU over triangulation interpolated with a phrase table extracted from the same 4k sentence pairs.

### **Translation Invariant Word Embeddings**

*Kejun Huang, Matt Gardner, Evangelos Papalexakis, Christos Faloutsos, Nikos Sidiropoulos, Tom Mitchell, Partha P. Talukdar, and Xiao Fu*

This work focuses on the task of finding latent vector representations of the words in a corpus. In particular, we address the issue of what to do when there are multiple languages in the corpus. Prior work has, among other techniques, used canonical correlation analysis to project pre-trained vectors in two languages into a common space. We propose a simple and scalable method that is inspired by the notion that the learned vector representations should be invariant to translation between languages. We show empirically that our method outperforms prior work on multilingual tasks, matches the performance of prior work on monolingual tasks, and scales linearly with the size of the input data (and thus the number of languages being embedded).

### **Hierarchical Phrase-based Stream Decoding**

*Andrew Finch, Xiaolin Wang, Masao Utiyama, and Eiichiro Sumita*

This paper proposes a method for hierarchical phrase-based stream decoding. A stream decoder is able to take a continuous stream of tokens as input, and segments this stream into word sequences that are translated and output as a stream of target word sequences. Phrase-based stream decoding techniques have been shown to be effective as a means of simultaneous interpretation. In this paper we transfer the essence of this idea into the framework of hierarchical machine translation. The hierarchical decoding framework organizes the decoding process into a chart; this structure is naturally suited to the process of stream decoding, leading to an efficient stream decoding algorithm that searches a restricted subspace containing only relevant hypotheses. Furthermore, the decoder allows more explicit access to the word re-ordering process that is of critical importance in decoding while interpreting. The decoder was evaluated on TED talk data for English-Spanish and English-Chinese. Our results show that like the phrase-based stream decoder, the hierarchical is capable of approaching the performance of the underlying hierarchical phrase-based machine translation decoder, at useful levels of latency. In addition

the hierarchical approach appeared to be robust to the difficulties presented by the more challenging English-Chinese task.

## **Rule Selection with Soft Syntactic Features for String-to-Tree Statistical Machine Translation**

*Fabienne Braune, Nina Seemann, and Alexander Fraser*

In syntax-based machine translation, rule selection is the task of choosing the correct target side of a translation rule among rules with the same source side. We define a discriminative rule selection model for systems that have syntactic annotation on the target language side (string-to-tree). This is a new and clean way to integrate soft source syntactic constraints into string-to-tree systems as features of the rule selection model. We release our implementation as part of Moses.

## **Motivating Personality-aware Machine Translation**

*Shachar Mirkin, Scott Nowson, Caroline Brun, and Julien Perez*

Language use is known to be influenced by personality traits as well as by socio-demographic characteristics such as age or mother tongue. As a result, it is possible to automatically identify these traits of the author from her texts. It has recently been shown that knowledge of such dimensions can improve performance in NLP tasks such as topic and sentiment modeling. We posit that machine translation is another application that should be personalized. In order to motivate this, we explore whether translation preserves demographic and psychometric traits. We show that, largely, both translation of the source training data into the target language, and the target test data into the source language has a detrimental effect on the accuracy of predicting author traits. We argue that this supports the need for personal and personality-aware machine translation models.

## **Trans-gram, Fast Cross-lingual Word-embeddings**

*Jocelyn Coulmance, Jean-Marc Marty, Guillaume Wenzek, and Amine Benhalloum*

We introduce “Trans-gram”, a simple and computationally-efficient method to simultaneously learn and align word-embeddings for a variety of languages, using only monolingual data and a smaller set of sentence-aligned data. We use our new method to compute aligned word-embeddings for twenty-one languages using English as a pivot language. We show that some linguistic features are aligned across languages for which we do not have aligned data, even though those properties do not exist in the pivot language. We also achieve state of the art results on standard cross-lingual text classification and word translation tasks.

## **The Overall Markedness of Discourse Relations**

*Lifeng Jin and Marie-Catherine de Marneffe*

Discourse relations can be categorized as continuous or discontinuous in the hypothesis of continuity (Murray, 1997), with continuous relations expressing normal succession of events in discourse such as temporal, spatial or causal. Asr and Demberg (2013) propose a markedness measure to test the prediction that discontinuous relations may have more unambiguous connectives, but restrict the markedness calculation to relations with explicit connectives only. This paper extends their measure to explicit and implicit relations and shows that results from this extension better fit the continuity hypothesis predictions both for the English Penn Discourse (Prasad et al., 2008) and the Chinese Discourse (Zhou and Xue, 2015) Treebanks.

## **Experiments in Open Domain Deception Detection**

*Verónica Pérez-Rosas and Rada Mihalcea*

The widespread use of deception in online sources has motivated the need for methods to automatically profile and identify deceivers. This work explores deception, gender and age detection in short texts using a machine learning approach. First, we collect a new open domain deception dataset also containing demographic data such as gender and age. Second, we extract feature sets including n-grams, shallow and deep syntactic features, semantic features, and syntactic complexity and readability metrics. Third, we build classifiers that aim to predict deception, gender, and age. Our findings show that while deception detection can be performed in short texts even in the absence of a pre-determined domain, gender and age prediction in deceptive texts is a challenging task. We further explore the linguistic differences in deceptive content that relate to deceivers gender and age and find evidence that both age and gender play an important role in people’s word choices when fabricating lies.

**A model of rapid phonotactic generalization***Tal Linzen and Timothy O'Donnell*

The phonotactics of a language describes the ways in which the sounds of the language combine to form possible morphemes and words. Humans can learn phonotactic patterns at the level of abstract classes, generalizing across sounds (e.g., "words can end in a voiced stop"). Moreover, they rapidly acquire these generalizations, even before they acquire soundspecific patterns. We present a probabilistic model intended to capture this earlyabstraction phenomenon. The model represents both abstract and concrete generalizations in its hypothesis space from the outset of learning. This-combined with a parsimony bias in favor of compact descriptions of the input data-leads the model to favor rapid abstraction in a way similar to human learners.

## Session 5 Overview

---

Track A	Track B	Track C	
Text Mining and NLP Applications (Long + TACL Papers) Main Auditorium	Semantics (Long +TACL Papers) Small Auditorium	Phonology and Word Segmentation (Long Papers) Room 2	
[TACL] Unsupervised Identification of Translationese <i>Rabinovich and Wintner</i>	Parsing English into Abstract Meaning Representation Using Syntax-Based Machine Translation <i>Pust, Hermjakob, Knight, Marcu, and May</i>	Do we need bigram alignment models? On the effect of alignment quality on transduction accuracy in G2P <i>Eger</i>	13:30
Automatically Solving Number Word Problems by Semantic Parsing and Reasoning <i>Shi, Wang, Lin, Liu, and Rui</i>	The Forest Convolutional Network: Compositional Distributional Semantics with a Neural Chart and without Binarization <i>Le and Zuidema</i>	Keyboard Logs as Natural Annotations for Word Segmentation <i>Takahasi and Mori</i>	13:55
[TACL] Which Step Do I Take First? Troubleshooting with Bayesian Models <i>Louis and Lapata</i>	Alignment-Based Compositional Semantics for Instruction Following <i>Andreas and Klein</i>	Long Short-Term Memory Neural Networks for Chinese Word Segmentation <i>Chen, Qiu, Zhu, Liu, and Huang</i>	14:20
[TACL] Problems in Current Text Simplification Research: New Data Can Help <i>Xu, Callison-Burch, and Napoles</i>	[TACL] Context-aware Frame-Semantic Role Labeling <i>Roth and Lapata</i>	Semi-supervised Chinese Word Segmentation based on Bilingual Information <i>Chen and Xu</i>	14:45

**Track D: Long Paper Posters**

Lower Level Foyer

Chair: Boxing Chen

13:30–15:10

**Poster Cluster 1: Machine Translation and Multilinguality (P1-8)**

- Hierarchical Back-off Modeling of Hiero Grammar based on Non-parametric Bayesian Model  
*H. Kamigaito, T. Watanabe, H. Takamura, M. Okumura, and E. Sumita*
- Consistency-Aware Search for Word Alignment  
*S. Shen, Y. Liu, M. Sun, and H. Luan*
- Graph-Based Collective Lexical Selection for Statistical Machine Translation  
*J. Su, D. Xiong, S. Huang, X. Han, and J. Yao*
- Bilingual Correspondence Recursive Autoencoder for Statistical Machine Translation  
*J. Su, D. Xiong, B. Zhang, Y. Liu, J. Yao, and M. Zhang*

- How to Avoid Unwanted Pregnancies: Domain Adaptation using Neural Network Models  
*S. Joty, H. Sajjad, N. Durrani, K. Al-Mannai, A. Abdelali, and S. Vogel*
- Detecting Content-Heavy Sentences: A Cross-Language Case Study  
*J. J. Li and A. Nenkova*
- Search-Aware Tuning for Hierarchical Phrase-based Decoding  
*F. Zhai, L. Huang, and K. Zhao*
- Part-of-speech Taggers for Low-resource Languages using CCA Features  
*Y.-B. Kim, B. Snyder, and R. Sarikaya*

**Track E: Short Paper Posters**

Chair: Stephen Clark

Lower Level Foyer

13:30–15:10

---

**Poster Cluster 1: Tagging, Syntax and Parsing (P1-12)**

- An Improved Tag Dictionary for Faster Part-of-Speech Tagging  
*R. Moore*
- Improving Arabic Diacritization through Syntactic Analysis  
*A. Shahrour, S. Khalifa, and N. Habash*
- Combining Discrete and Continuous Features for Deterministic Transition-based Dependency Parsing  
*M. Zhang and Y. Zhang*
- Efficient Inner-to-outer Greedy Algorithm for Higher-order Labeled Dependency Parsing  
*X. Ma and E. Hovy*
- Online Updating of Word Representations for Part-of-Speech Tagging  
*W. Yin, T. Schnabel, and H. Schütze*
- Empty Category Detection using Path Features and Distributed Case Frames  
*S. Takeno, M. Nagata, and K. Yamamoto*
- Foreebank: Syntactic Analysis of Customer Support Forums  
*R. Kaljahi, J. Foster, J. Roturier, C. Ribeyre, T. Lynn, and J. Le Roux*
- Semi-supervised Dependency Parsing using Bilexical Contextual Features from Auto-Parsed Data  
*E. Kiperwasser and Y. Goldberg*
- Improved Transition-Based Parsing and Tagging with Neural Networks  
*C. Alberti, D. Weiss, G. Coppola, and S. Petrov*
- Syntactic Parse Fusion  
*D. K. Choe, D. McClosky, and E. Charniak*
- Not All Contexts Are Created Equal: Better Word Representations with Variable Attention  
*W. Ling, Y. Tsvetkov, S. Amir, R. Fernandez, C. Dyer, A. W. Black, I. Trancoso, and C.-C. Lin*

- **An Improved Non-monotonic Transition System for Dependency Parsing**  
*M. Honnibal and M. Johnson*

## Abstracts: Session 5

---

### Session 5A: Text Mining and NLP Applications (Long + TACL Papers)

Main Auditorium

Chair: Marie-Francine Moens

#### [TACL] Unsupervised Identification of Translationese

Ella Rabinovich and Shuly Wintner

13:30–13:55

Translated texts are distinctively different from original ones, to the extent that supervised text classification methods can distinguish between them with high accuracy. These differences were proven useful for statistical machine translation. However, it has been suggested that the accuracy of translation detection deteriorates when the classifier is evaluated outside the domain it was trained on. We show that this is indeed the case, in a variety of evaluation scenarios. We then show that unsupervised classification is highly accurate on this task. We suggest a method for determining the correct labels of the clustering outcomes, and then use the labels for voting, improving the accuracy even further. Moreover, we suggest a simple method for clustering in the challenging case of mixed-domain datasets, in spite of the dominance of domain-related features over translation-related ones. The result is an effective, fully-unsupervised method for distinguishing between original and translated texts that can be applied to new domains with reasonable accuracy.

#### Automatically Solving Number Word Problems by Semantic Parsing and Reasoning

Shuming Shi, Yuehui Wang, Chin-Yew Lin, Xiaojiang Liu, and Yong Rui 13:55–14:20

This paper presents a semantic parsing and reasoning approach to automatically solving math word problems. A new meaning representation language is designed to bridge natural language text and math expressions. A CFG parser is implemented based on 9,600 semi-automatically created grammar rules. We conduct experiments on a test set of over 1,500 number word problems (i.e., verbally expressed number problems) and yield 95.4% precision and 60.2% recall.

#### [TACL] Which Step Do I Take First? Troubleshooting with Bayesian Models

Annie Louis and Mirella Lapata

14:20–14:45

Online discussion forums and community question-answering websites provide one of the primary avenues for online users to share information. In this paper, we propose text mining techniques which aid users navigate troubleshooting-oriented data such as questions asked on forums and their suggested solutions. We introduce Bayesian generative models of the troubleshooting data and apply them to two interrelated tasks (a) predicting the complexity of the solutions (e.g., plugging a keyboard in the computer is easier compared to installing a special driver) and (b) presenting them in a ranked order from least to most complex. Experimental results show that our models are on par with human performance on these tasks, while outperforming baselines based on solution length or readability.

#### [TACL] Problems in Current Text Simplification Research: New Data Can Help

Wei Xu, Chris Callison-Burch, and Courtney Napoles

14:45–15:10

Simple Wikipedia has dominated simplification research in the past 5 years. In this opinion paper, we argue that focusing on Wikipedia limits simplification research. We back up our arguments with corpus analysis and by highlighting statements that other researchers have made in the simplification literature. We introduce a new simplification dataset that is a significant improvement over Simple Wikipedia, and present a novel quantitative-comparative approach to study the quality of simplification data resources.

**Session 5B: Semantics (Long +TACL Papers)**

Small Auditorium

Chair: Benjamin Van Durme

**Parsing English into Abstract Meaning Representation Using Syntax-Based Machine Translation***Michael Pust, Ulf Hermjakob, Kevin Knight, Daniel Marcu, and Jonathan May* 13:30–13:55

We present a parser for Abstract Meaning Representation (AMR). We treat English-to-AMR conversion within the framework of string-to-tree, syntax-based machine translation (SBMT). To make this work, we transform the AMR structure into a form suitable for the mechanics of SBMT and useful for modeling. We introduce an AMR-specific language model and add data and features drawn from semantic resources. Our resulting AMR parser significantly improves upon state-of-the-art results.

**The Forest Convolutional Network: Compositional Distributional Semantics with a Neural Chart and without Binarization***Phong Le and Willem Zuidema* 13:55–14:20

According to the principle of compositionality, the meaning of a sentence is computed from the meaning of its parts and the way they are syntactically combined. In practice, however, the syntactic structure is computed by automatic parsers which are far-from-perfect and not tuned to the specifics of the task. Current recursive neural network (RNN) approaches for computing sentence meaning therefore run into a number of practical difficulties, including the need to carefully select a parser appropriate for the task, deciding how and to what extent syntactic context modifies the semantic composition function, as well as on how to transform parse trees to conform to the branching settings (typically, binary branching) of the RNN. This paper introduces a new model, the Forest Convolutional Network, that avoids all of these challenges, by taking a parse forest as input, rather than a single tree, and by allowing arbitrary branching factors. We report improvements over the state-of-the-art in sentiment analysis and question classification.

**Alignment-Based Compositional Semantics for Instruction Following***Jacob Andreas and Dan Klein* 14:20–14:45

This paper describes an alignment-based model for interpreting natural language instructions in context. We approach instruction following as a search over plans, scoring sequences of actions conditioned on structured observations of text and the environment. By explicitly modeling both the low-level compositional structure of individual actions and the high-level structure of full plans, we are able to learn both grounded representations of sentence meaning and pragmatic constraints on interpretation. To demonstrate the model's flexibility, we apply it to a diverse set of benchmark tasks. On every task, we outperform strong task-specific baselines, and achieve several new state-of-the-art results.

**[TACL] Context-aware Frame-Semantic Role Labeling***Michael Roth and Mirella Lapata* 14:45–15:10

Frame semantic representations have been useful in several applications ranging from text-to-scene generation, to question answering and social network analysis. Predicting such representations from raw text is, however, a challenging task and corresponding models are typically only trained on a small set of sentence-level annotations. In this paper, we present a semantic role labeling system that takes into account sentence and discourse context. We introduce several new features which we motivate based on linguistic insights and experimentally demonstrate that they lead to significant improvements over the current state-of-the-art in FrameNet-based semantic role labeling.

**Session 5C: Phonology and Word Segmentation (Long Papers)**

Room 2

Chair: Yue Zhang

**Do we need bigram alignment models? On the effect of alignment quality on transcription accuracy in G2P**

Steffen Eger

13:30–13:55

We investigate the need for bigram alignment models and the benefit of supervised alignment techniques in grapheme-to-phoneme (G2P) conversion. Moreover, we quantitatively estimate the relationship between alignment quality and overall G2P system performance. We find that, in English, bigram alignment models do perform better than unigram alignment models on the G2P task. Moreover, we find that supervised alignment techniques may perform considerably better than their unsupervised brethren and that few manually aligned training pairs suffice for them to do so. Finally, we estimate a highly significant impact of alignment quality on overall G2P transcription performance and that this relationship is linear in nature.

**Keyboard Logs as Natural Annotations for Word Segmentation**

Fumihiko Takahasi and Shinsuke Mori

13:55–14:20

In this paper we propose a framework to improve word segmentation accuracy using input method logs. An input method is software used to type sentences in languages which have far more characters than the number of keys on a keyboard. The main contributions of this paper are: 1) an input method server that proposes word candidates which are not included in the vocabulary, 2) a publicly usable input method that logs user behavior (like typing and selection of word candidates), and 3) a method for improving word segmentation by using these logs. We conducted word segmentation experiments on tweets from Twitter, and showed that our method improves accuracy in this domain. Our method itself is domain-independent and only needs logs from the target domain.

**Long Short-Term Memory Neural Networks for Chinese Word Segmentation**

Xinchi Chen, Xipeng Qiu, Chenxi Zhu, Pengfei Liu, and Xuanjing Huang 14:20–14:45

Currently most of state-of-the-art methods for Chinese word segmentation are based on supervised learning, whose features are mostly extracted from local context. These methods cannot utilize the long distance information which is also crucial for word segmentation. In this paper, we propose a novel neural network model for Chinese word segmentation, which adopts the long short-term memory (LSTM) neural network to keep the previous important information in memory cell and avoids the limit of window size of local context. Experiments on PKU, MSRA and CTB6 benchmark datasets show that our model outperforms the previous neural network models and state-of-the-art methods.

**Semi-supervised Chinese Word Segmentation based on Bilingual Information**

Wei Chen and Bo Xu

14:45–15:10

This paper presents a bilingual semi-supervised Chinese word segmentation (CWS) method that leverages the natural segmenting information of English sentences. The proposed method involves learning three levels of features, namely, character-level, phrase-level and sentence-level, provided by multiple sub-models. We use a sub-model of conditional random fields (CRF) to learn monolingual grammars, a sub-model based on character-based alignment to obtain explicit segmenting knowledge, and another sub-model based on transliteration similarity to detect out-of-vocabulary (OOV) words. Moreover, we propose a sub-model leveraging neural network to ensure the proper treatment of the semantic gap and a phrase-based translation sub-model to score the translation probability of the Chinese segmentation and its corresponding English sentences. A cascaded log-linear model is employed to combine these features to segment bilingual unlabeled data, the results of which are used to justify the original supervised CWS model. The evaluation shows that our method results in superior results compared with those of the state-of-the-art monolingual and bilingual semi-supervised models that have been reported in the literature.

**Session 5D: Long Paper Posters**

Lower Level Foyer

Chair: Boxing Chen

13:30–15:10

**Hierarchical Back-off Modeling of Hiero Grammar based on Non-parametric Bayesian Model***Hidetaka Kamigaito, Taro Watanabe, Hiroya Takamura, Manabu Okumura, and Eiichiro Sumita*

In hierarchical phrase-based machine translation, a rule table is automatically learned by heuristically extracting synchronous rules from a parallel corpus. As a result, spuriously many rules are extracted which may be composed of various incorrect rules. The larger rule table incurs more run time for decoding and may result in lower translation quality. To resolve the problems, we propose a hierarchical back-off model for Hiero grammar, an instance of a synchronous context free grammar (SCFG), on the basis of the hierarchical Pitman-Yor process. The model can extract a compact rule and phrase table without resorting to any heuristics by hierarchically backing off to smaller phrases under SCFG. Inference is efficiently carried out using two-step synchronous parsing of Xiao et al., (2012) combined with slice sampling. In our experiments, the proposed model achieved higher or at least comparable translation quality against a previous Bayesian model on various language pairs; German/French/Spanish/Japanese-English. When compared against heuristic models, our model achieved comparable translation quality on a full size German-English language pair in Europarl v7 corpus with significantly smaller grammar size; less than 10% of that for heuristic model.

**Consistency-Aware Search for Word Alignment***Shiqi Shen, Yang Liu, Maosong Sun, and Huanbo Luan*

As conventional word alignment search algorithms usually ignore the consistency constraint in translation rule extraction, improving alignment accuracy does not necessarily increase translation quality. We propose to use coverage, which reflects how well extracted phrases can recover the training data, to enable word alignment to model consistency and correlate better with machine translation. This can be done by introducing an objective that maximizes both alignment model score and coverage. We introduce an efficient algorithm to calculate coverage on the fly during search. Experiments show that our consistency-aware search algorithm significantly outperforms both generative and discriminative alignment approaches across various languages and translation models.

**Graph-Based Collective Lexical Selection for Statistical Machine Translation***Jinsong Su, Deyi Xiong, Shujian Huang, Xianpei Han, and Junfeng Yao*

Lexical selection is of great importance to statistical machine translation. In this paper, we propose a graph-based framework for collective lexical selection. The framework is established on a translation graph that captures not only local associations between source-side content words and their target translations but also target-side global dependencies in terms of relatedness among target items. We also introduce a random walk style algorithm to collectively identify translations of source-side content words that are strongly related in translation graph. We validate the effectiveness of our lexical selection framework on Chinese-English translation. Experiment results with large-scale training data show that our approach significantly improves lexical selection.

**Bilingual Correspondence Recursive Autoencoder for Statistical Machine Translation***Jinsong Su, Deyi Xiong, Biao Zhang, Yang Liu, Junfeng Yao, and Min Zhang*

Learning semantic representations and tree structures of bilingual phrases is beneficial for statistical machine translation. In this paper, we propose a new neural network model called Bilingual Correspondence Recursive Autoencoder (BCorrRAE) to model bilingual phrases in translation. We incorporate word alignments into BCorrRAE to allow it freely access bilingual constraints at different levels. BCorrRAE minimizes a joint objective on the combination of a recursive autoencoder reconstruction error, a structural alignment consistency error and a cross-lingual reconstruction error so as to not only generate alignment-consistent phrase structures, but also capture different levels of semantic relations within bilingual phrases. In order to examine the effectiveness of BCorrRAE, we incorporate both semantic and structural similarity features built on bilingual phrase representations and tree structures learned by BCorrRAE into a state-of-the-art SMT system. Experiments on NIST Chinese-English test sets show that our model achieves a substantial improvement of up to 1.55 BLEU points over the baseline.

**How to Avoid Unwanted Pregnancies: Domain Adaptation using Neural Network Models**

*Shafiq Joty, Hassan Sajjad, Nadir Durrani, Kamla Al-Mannai, Ahmed Abdelali, and Stephan Vogel*

We present novel models for domain adaptation based on the neural network joint model (NNJM). Our models maximize the cross entropy by regularizing the loss function with respect to in-domain model. Domain adaptation is carried out by assigning higher weight to out-domain sequences that are similar to the in-domain data. In our alternative model we take a more restrictive approach by additionally penalizing sequences similar to the out-domain data. Our models achieve better perplexities than the baseline NNJM models and give improvements of up to 0.5 and 0.6 BLEU points in Arabic-to-English and English-to-German language pairs, on a standard task of translating TED talks.

**Detecting Content-Heavy Sentences: A Cross-Language Case Study**

*Junyi Jessy Li and Ani Nenkova*

The information conveyed by some sentences would be more easily understood by a reader if it were expressed in multiple sentences. We call such sentences content heavy: these are possibly grammatical but difficult to comprehend, cumbersome sentences. In this paper we introduce the task of detecting content-heavy sentences in cross-lingual context. Specifically we develop methods to identify sentences in Chinese for which English speakers would prefer translations consisting of more than one sentence. We base our analysis and definitions on evidence from multiple human translations and reader preferences on flow and understandability. We show that machine translation quality when translating content heavy sentences is markedly worse than overall quality and that this type of sentence are fairly common in Chinese news. We demonstrate that sentence length and punctuation usage in Chinese are not sufficient clues for accurately detecting heavy sentences and present a richer classification model that accurately identifies these sentences.

**Search-Aware Tuning for Hierarchical Phrase-based Decoding**

*Feifei Zhai, Liang Huang, and Kai Zhao*

Parameter tuning is a key problem for statistical machine translation (SMT). Most popular parameter tuning algorithms for SMT are agnostic of the decoding algorithm, resulting in parameters vulnerable to search errors in decoding. The recent research of “search-aware tuning” (Liu and Huang, 2014) address this problem by forcing the tuning algorithms to consider the translation derivations in every decoding step to make promising partial translations more likely to survive the in-exact decoding beam. We extend this approach from phrase-based translation to syntax-based translation by generalizing the translation quality metrics for partial translations to handle tree-structured derivations in a way inspired by inside-outside algorithm. Our approach is simple to use and can be applied to all conventional parameter tuning methods as a plugin. Extensive experiments on Chinese-to- English translation show significant BLEU improvements in MERT and MIRA.

**Part-of-speech Taggers for Low-resource Languages using CCA Features**

*Young-Bum Kim, Benjamin Snyder, and Ruhi Sarikaya*

In this paper, we address the challenge of creating accurate and robust part- of- speech taggers for low-resource languages. We propose a method that leverages existing parallel data between the target language and a large set of resource-rich languages without ancillary resources such as tag dictionaries. Crucially, we use CCA to induce latent word representations that incorporate cross-genre distributional cues, as well as projected tags from a full array of resource-rich languages. We develop a probability-based confidence model to identify words with highly likely tag projections and use these words to train a multi-class SVM using the CCA features. Our method yields average performance of 85% accuracy for languages with almost no resources, outperforming a state-of-the-art partially-observed CRF model.

**Session 5E: Short Paper Posters**

Lower Level Foyer

Chair: Stephen Clark

13:30–15:10

**An Improved Tag Dictionary for Faster Part-of-Speech Tagging***Robert Moore*

Ratnaparkhi (1996) introduced a method of inferring a tag dictionary from annotated data to speed up part-of-speech tagging by limiting the set of possible tags for each word. While Ratnaparkhi's tag dictionary makes tagging faster but less accurate, an alternative tag dictionary that we recently proposed (Moore, 2014) makes tagging as fast as with Ratnaparkhi's tag dictionary, but with no decrease in accuracy. In this paper, we show that a very simple semi-supervised variant of Ratnaparkhi's method results in a much tighter tag dictionary than either Ratnaparkhi's or our previous method, with accuracy as high as with our previous tag dictionary but much faster tagging—more than 100,000 tokens per second in Perl.

**Improving Arabic Diacritization through Syntactic Analysis***Anas Shahrour, Salam Khalifa, and Nizar Habash*

We present an approach to Arabic automatic diacritization that integrates syntactic analysis with morphological tagging through improving the prediction of case and state features. Our best system increases the accuracy of word diacritization by 2.5% absolute on all words, and 5.2% absolute on nominals over a state-of-the-art baseline. Similar increases are shown on the full morphological analysis choice.

**Combining Discrete and Continuous Features for Deterministic Transition-based Dependency Parsing***Meishan Zhang and Yue Zhang*

We investigate a combination of a traditional linear sparse feature model and a multi-layer neural network model for deterministic transition-based dependency parsing, by integrating the sparse features into the neural model. Correlations are drawn between the hybrid model and previous work on integrating word embedding features into a discrete linear model. By analyzing the results of various parsers on web-domain parsing, we show that the integrated model is a better way to combine traditional and embedding features compared with previous methods.

**Efficient Inner-to-outer Greedy Algorithm for Higher-order Labeled Dependency Parsing***Xuezhe Ma and Eduard Hovy*

Many NLP systems use dependency parsers as critical components. Joint learning parsers usually achieve better parsing accuracies than two-stage methods. However, classical joint parsing algorithms significantly increase computational complexity, which makes joint learning impractical. In this paper, we proposed an efficient dependency parsing algorithm that is capable of capturing multiple edge-label features, while maintaining low computational complexity. We evaluate our parser on 14 different languages. Our parser consistently obtains more accurate results than three baseline systems and three popular, off-the-shelf parsers.

**Online Updating of Word Representations for Part-of-Speech Tagging***Wenpeng Yin, Tobias Schnabel, and Hinrich Schütze*

We propose online unsupervised domain adaptation (DA), which is performed incrementally as data comes in and is applicable when batch DA is not possible. In a part-of-speech (POS) tagging evaluation, we find that online unsupervised DA performs as well as batch DA.

**Empty Category Detection using Path Features and Distributed Case Frames***Shunsuke Takeno, Masaaki Nagata, and Kazuhide Yamamoto*

We describe an approach for machine learning-based empty category detection on the phrase structure analysis of Japanese. The problem are formalized as tree node classification, we find that conjunction between path features and other node component features, namely, head word feature, child feature and empty category feature, are highly effective. We also find that a set of dot products between the word embeddings for a verb and those for case particles can be used as a substitution for case frames. The experiment results showed that the proposed method outperformed the previous state of the art from 68.6% to 73.2% in F-measure.

**Foreebank: Syntactic Analysis of Customer Support Forums**

*Rasoul Kaljahi, Jennifer Foster, Johann Roturier, Corentin Ribeyre, Teresa Lynn, and Joseph Le Roux*

We present a new treebank of English and French technical forum content which has been annotated for grammatical errors and phrase structure. This double annotation allows us to empirically measure the effect of errors on parsing performance. While it is slightly easier to parse the corrected versions of the forum sentences, the errors are not the main factor in making this kind of text hard to parse.

**Semi-supervised Dependency Parsing using Bilexical Contextual Features from Auto-Parsed Data**

*Eliyahu Kiperwasser and Yoav Goldberg*

We present a semi-supervised approach to improve dependency parsing accuracy by using bilexical statistics derived from auto-parsed data. The method is based on estimating the attachment potential of head-modifier words, by taking into account not only the head and modifier words themselves, but also the words surrounding the head and the modifier. When integrating the learned statistics as features in a graph-based parsing model, we observe nice improvements in accuracy when parsing various English datasets.

**Improved Transition-Based Parsing and Tagging with Neural Networks**

*Chris Alberti, David Weiss, Greg Coppola, and Slav Petrov*

We extend and improve upon recent work in structured training for neural network transition-based dependency parsing. We do this by experimenting with novel features, additional transition systems and by testing on a wider array of languages. In particular, we introduce set-valued features to encode the predicted morphological properties and part-of-speech confusion sets of the words being parsed. We also investigate the use of joint parsing and part-of-speech tagging in the neural paradigm. Finally, we conduct a multi-lingual evaluation that demonstrates the robustness of the overall structured neural approach, as well as the benefits of the extensions proposed in this work. Our research further demonstrates the breadth of the applicability of neural network methods to dependency parsing, as well as the ease with which new features can be added to neural parsing models.

**Syntactic Parse Fusion**

*Do Kook Choe, David McClosky, and Eugene Charniak*

Model combination techniques have consistently shown state-of-the-art performance across multiple tasks, including syntactic parsing. However, they dramatically increase runtime and can be difficult to employ in practice. We demonstrate that applying constituency model combination techniques to n-best lists instead of n different parsers results in significant parsing accuracy improvements. Parses are weighted by their probabilities and combined using an adapted version of Sagae and Lavie (2006). These accuracy gains come with marginal computational costs and are obtained on top of existing parsing techniques such as discriminative reranking and self-training, resulting in state-of-the-art accuracy: 92.6% on WSJ section 23. On out-of-domain corpora, accuracy is improved by 0.4% on average. We empirically confirm that six well-known n-best parsers benefit from the proposed methods across six domains.

**Not All Contexts Are Created Equal: Better Word Representations with Variable Attention**

*Wang Ling, Yulia Tsvetkov, Silvio Amir, Ramon Fernandez, Chris Dyer, Alan W Black, Isabel Trancoso, and Chu-Cheng Lin*

We introduce an extension to the bag-of-words model for learning words representations that take into account both syntactic and semantic properties within language. This is done by employing an attention model, finds within the contextual words, the words that are relevant for each prediction. The general intuition of our model is that some words are only relevant for predicting local context (e.g. function words), while other words are more suited for determining global context, such as the topic of the document. Experiments performed on both semantically and syntactically oriented tasks show gains using our model over the existing bag of words model. Furthermore, compared to other more sophisticated models, our model scales better as we increase the size of the context of the model.

**An Improved Non-monotonic Transition System for Dependency Parsing**

*Matthew Honnibal and Mark Johnson*

Transition-based dependency parsers usually use transition systems that monotonically extend partial parse states until they identify a complete parse tree. Honnibal et al. (2013) showed that greedy parsing accuracy can be improved by adding additional non-monotonic transitions that permit the parser to “repair” earlier parsing mistakes by “over-writing” earlier parsing decisions. This increases the size of the set of complete parse trees that each partial parse state can derive, enabling such a parser to escape the “garden paths” that can trap monotonic greedy transition-based dependency parsers. We describe a new set of non-monotonic transitions that permits a partial parse state to derive a larger set of completed parse trees than previous work, which allows our parser to escape from a larger set of garden paths. A parser with our new non-monotonic transition system has 91.85% directed attachment accuracy, an improvement of 0.6% over a comparable parser using the standard monotonic arc-eager transitions.

## Session 6 Overview

---

	<b>Track A</b>	<b>Track B</b>	<b>Track C</b>
	Machine Translation (Long Papers) Main Auditorium	Sentiment Analysis and Opinion Mining / Tagging, Chunking and Parsing (Long Papers) Small Auditorium	Language and Vision / Information Extraction (Long Papers) Room 2
15:40	Improving Statistical Machine Translation with a Multilingual Paraphrase Database <i>Mehdizadeh Seraj, Siahbani, and Sarkar</i>	Document Modeling with Gated Recurrent Neural Network for Sentiment Classification <i>Tang, Qin, and Liu</i>	Solving Geometry Problems: Combining Text and Diagram Interpretation <i>Seo, Hajishirzi, Farhadi, Etzioni, and Malcolm</i>
16:05	Learning Semantic Representations for Nonterminals in Hierarchical Phrase-Based Translation <i>Wang, Xiong, and Zhang</i>	Fine-grained Opinion Mining with Recurrent Neural Networks and Word Embeddings <i>Liu, Joty, and Meng</i>	Do You See What I Mean? Visual Resolution of Linguistic Ambiguities <i>Berzak, Barbu, Harari, Katz, and Ullman</i>
16:30	A Comparison between Count and Neural Network Models Based on Joint Translation and Reordering Sequences <i>Guta, Alkhouri, Peter, Wuebker, and Ney</i>	Joint A* CCG Parsing and Semantic Role Labelling <i>Lewis, He, and Zettlemoyer</i>	Efficient and Expressive Knowledge Base Completion Using Subgraph Feature Extraction <i>Gardner and Mitchell</i>
16:55	Effective Approaches to Attention-based Neural Machine Translation <i>Luong, Pham, and Manning</i>	Improving Semantic Parsing with Enriched Synchronous Context-Free Grammar <i>Li, Zhu, Lu, and Zhou</i>	Representing Text for Joint Embedding of Text and Knowledge Bases <i>Toutanova, Chen, Pantel, Poon, Choudhury, and Gamon</i>

**Track D: Long Paper Posters**

Lower Level Foyer

Chair: Noah Smith

15:40–17:20

---

### Poster Cluster 1: Statistical Models and Machine Learning Methods (P1-10)

- A Utility Model of Authors in the Scientific Community  
*Y. Sim, B. Routledge, and N. A. Smith*
- Finding Function in Form: Compositional Character Models for Open Vocabulary Word Representation  
*W. Ling, C. Dyer, A. W. Black, I. Trancoso, R. Fernandez, S. Amir, L. Marujo, and T. Luis*
- Syntax-Aware Multi-Sense Word Embeddings for Deep Compositional Models of Meaning  
*J. Cheng and D. Kartsaklis*

- Conversation Trees: A Grammar Model for Topic Structure in Forums  
*A. Louis and S. B. Cohen*
- Fast, Flexible Models for Discovering Topic Correlation across Weakly-Related Collections  
*J. Zhang, A. Gerow, J. Altosaar, J. Evans, and R. Jean So*
- Molding CNNs for text: non-linear, non-consecutive convolutions  
*T. Lei, R. Barzilay, and T. Jaakkola*
- Multi-Perspective Sentence Similarity Modeling with Convolutional Neural Networks  
*H. He, K. Gimpel, and J. Lin*
- Posterior calibration and exploratory analysis for natural language processing models  
*K. Nguyen and B. O'Connor*
- A Generative Word Embedding Model and its Low Rank Positive Semidefinite Solution  
*S. Li, J. Zhu, and C. Miao*
- Reading Documents for Bayesian Online Change Point Detection  
*T. Kim and J. Choi*

**Track E: Short Paper Posters**

Chair: Ido Dagan

Lower Level Foyer

15:40–17:20

**Poster Cluster 1: Semantics (P1-13)**

- Recognizing Textual Entailment Using Probabilistic Inference  
*L. Sha, S. Li, B. Chang, Z. Sui, and T. Jiang*
- Chinese Semantic Role Labeling with Bidirectional Recurrent Neural Networks  
*Z. Wang, T. Jiang, B. Chang, and Z. Sui*
- Unsupervised Negation Focus Identification with Word-Topic Graph Model  
*B. Zou, G. Zhou, and Q. Zhu*
- Reverse-engineering Language: A Study on the Semantic Compositionality of German Compounds  
*C. Dima*
- Event Detection and Factuality Assessment with Non-Expert Supervision  
*K. Lee, Y. Artzi, Y. Choi, and L. Zettlemoyer*
- Large-Scale Acquisition of Entailment Pattern Pairs by Exploiting Transitivity  
*J. Kloetzer, K. Torisawa, C. Hashimoto, and J.-H. Oh*
- Context-Dependent Knowledge Graph Embedding  
*Y. Luo, Q. Wang, B. Wang, and L. Guo*
- Learning to Identify the Best Contexts for Knowledge-based WSD  
*E. Wasserman Pritsker, W. Cohen, and E. Minkov*

- Measuring Prerequisite Relations Among Concepts  
*C. Liang, Z. Wu, W. Huang, and C. L. Giles*
- Adapting Phrase-based Machine Translation to Normalise Medical Terms in Social Media Messages  
*N. Limsopatham and N. Collier*
- Script Induction as Language Modeling  
*R. Rudinger, P. Rastogi, F. Ferraro, and B. Van Durme*
- Online Learning of Interpretable Word Embeddings  
*H. Luo, Z. Liu, H. Luan, and M. Sun*
- A Strong Lexical Matching Method for the Machine Comprehension Test  
*E. Smith, N. Greco, M. Bosnjak, and A. Vlachos*

## Abstracts: Session 6

---

### Session 6A: Machine Translation (Long Papers)

Main Auditorium

Chair: Lucia Specia

#### **Improving Statistical Machine Translation with a Multilingual Paraphrase Database**

*Ramtin Mehdizadeh Seraj, Maryam Siahbani, and Anoop Sarkar*

15:40–16:05

The multilingual Paraphrase Database (PPDB) is a freely available automatically created resource of paraphrases in multiple languages. In statistical machine translation, paraphrases can be used to provide translation for out-of-vocabulary (OOV) phrases. In this paper, we show that a graph propagation approach that uses PPDB paraphrases can be used to improve overall translation quality. We provide an extensive comparison with previous work and show that our PPDB-based method improves the BLEU score by up to 1.79 percent points. We show that our approach improves on the state of the art in three different settings: when faced with limited amount of parallel training data; a domain shift between training and test data; and handling a morphologically complex source language. Our PPDB-based method outperforms the use of distributional profiles from monolingual source data.

#### **Learning Semantic Representations for Nonterminals in Hierarchical Phrase-Based Translation**

*Xing Wang, Deyi Xiong, and Min Zhang*

16:05–16:30

In hierarchical phrase-based translation, coarse-grained nonterminal Xs may generate inappropriate translations due to the lack of sufficient information for phrasal substitution. In this paper we propose a framework to refine nonterminals in hierarchical translation rules with realvalued semantic representations. The semantic representations are learned via a weighted mean value and a minimum distance method using phrase vector representations obtained from large scale monolingual corpus. Based on the learned semantic vectors, we build a semantic nonterminal refinement model to measure semantic similarities between phrasal substitutions and nonterminal Xs in translation rules. Experiment results on Chinese-English translation show that the proposed model significantly improves translation quality on NIST test sets.

#### **A Comparison between Count and Neural Network Models Based on Joint Translation and Reordering Sequences**

*Andreas Guta, Tamer Alkhouri, Jan-Thorsten Peter, Joern Wuebker, and Hermann Ney* 16:30–16:55

We propose a conversion of bilingual sentence pairs and the corresponding word alignments into novel linear sequences. These are joint translation and reordering (JTR) uniquely defined sequences, combining interdepending lexical and alignment dependencies on the word level into a single framework. They are constructed in a simple manner while capturing multiple alignments and empty words. JTR sequences can be used to train a variety of models. We investigate the performances of n-gram models with modified Kneser-Ney smoothing, feed-forward and recurrent neural network architectures when estimated on JTR sequences, and compare them to the operation sequence model (Durrani et al., 2013). Evaluations on the IWSLT German-English, WMT German-English and BOLT Chinese-English tasks show that JTR models improve state-of-the-art phrase-based systems by up to 2.2 BLEU.

#### **Effective Approaches to Attention-based Neural Machine Translation**

*Thang Luong, Hieu Pham, and Christopher D. Manning*

16:55–17:20

An attentional mechanism has lately been used to improve neural machine translation (NMT) by selectively focusing on parts of the source sentence during translation. However, there has been little work exploring useful architectures for attention-based NMT. This paper examines two simple and effective classes of attentional mechanism: a global approach which always attends to all source words and a local one that only looks at a subset of source words at a time. We demonstrate the effectiveness of both approaches over the WMT translation tasks between English and German in both directions. With local attention, we achieve a significant gain of 5.0 BLEU points over non-attentional systems which already incorporate known techniques such as dropout. Our ensemble model using different attention

architectures has established a new state-of-the-art result in the WMT'15 English to German translation task with 25.9 BLEU points, an improvement of 1.0 BLEU points over the existing best system backed by NMT and an n-gram reranker.

## Session 6B: Sentiment Analysis and Opinion Mining / Tagging, Chunking and Parsing (Long Papers)

Small Auditorium

Chair: Bing Liu and Dan Bikell

### Document Modeling with Gated Recurrent Neural Network for Sentiment Classification

Duyu Tang, Bing Qin, and Ting Liu

15:40–16:05

Document level sentiment classification remains a challenge: encoding the intrinsic relations between sentences in the semantic meaning of a document. To address this, we introduce a neural network model to learn vector-based document representation in a unified, bottom-up fashion. The model first learns sentence representation with convolutional neural network or long short-term memory. Afterwards, semantics of sentences and their relations are adaptively encoded in document representation with gated recurrent neural network. We conduct document level sentiment classification on four large-scale review datasets from IMDB and Yelp Dataset Challenge. Experimental results show that: (1) our neural model shows superior performances over several state-of-the-art algorithms; (2) gated recurrent neural network dramatically outperforms standard recurrent neural network in document modeling for sentiment classification.

### Fine-grained Opinion Mining with Recurrent Neural Networks and Word Embeddings

Pengfei Liu, Shafiq Joty, and Helen Meng

16:05–16:30

The tasks in fine-grained opinion mining can be regarded as either a token-level sequence labeling problem or as a semantic compositional task. We propose a general class of discriminative models based on recurrent neural networks (RNNs) and word embeddings that can be successfully applied to such tasks without any task-specific feature engineering effort. Our experimental results on the task of opinion target identification show that RNNs, without using any hand-crafted features, outperform feature-rich CRF-based models. Our framework is flexible, allows us to incorporate other linguistic features, and achieves results that rival the top performing systems in SemEval-2014.

### Joint A\* CCG Parsing and Semantic Role Labelling

Mike Lewis, Luheng He, and Luke Zettlemoyer

16:30–16:55

Joint models of syntactic and semantic parsing have the potential to improve performance on both tasks - but to date, the best results have been achieved with pipelines. We introduce a joint model using CCG, which is motivated by the close link between CCG syntax and semantics. Semantic roles are recovered by labelling the deep dependency structures produced by the grammar. Furthermore, because CCG is lexicalized, we show it is possible to factor the parsing model over words and introduce a new A\* parsing algorithm - which we demonstrate is faster and more accurate than adaptive supertagging. Our joint model is the first to substantially improve both syntactic and semantic accuracy over a comparable pipeline, and also achieves state-of-the-art results for a non-ensemble semantic role labelling model.

### Improving Semantic Parsing with Enriched Synchronous Context-Free Grammar

Junhui Li, Muhua Zhu, Wei Lu, and Guodong Zhou

16:55–17:20

Semantic parsing maps a sentence in natural language into a structured meaning representation. Previous studies show that semantic parsing with synchronous context-free grammars (SCFGs) achieves favorable performance over most other alternatives. Motivated by the observation that the performance of semantic parsing with SCFGs is closely tied to the translation rules, this paper explores to extend translation rules with high quality and increased coverage in three ways. First, we introduce structure informed non-terminals, better guiding the parsing in favor of well formed structure, instead of using a uniformed non-terminal in SCFGs. Second, we examine the difference between word alignments for semantic parsing and statistical machine translation (SMT) to better adapt word alignment in SMT to semantic parsing. And finally, we address the unknown word translation issue via synthetic translation rules. Evaluation on the standard GeoQuery benchmark dataset shows that our approach achieves the state-of-the-art across various languages, including English, German and Greek.

**Session 6C: Language and Vision / Information Extraction (Long Papers)**

Room 2

Chair: Meg Mitchell and Partha Talukdar

**Solving Geometry Problems: Combining Text and Diagram Interpretation***Minjoon Seo, Hannaneh Hajishirzi, Ali Farhadi, Oren Etzioni, and Clint Malcolm* 15:40-16:05

This paper introduces GEOS, the first automated system to solve unaltered SAT geometry questions by combining text understanding and diagram interpretation. We model the problem of understanding geometry questions as submodular optimization, and identify a formal problem description likely to be compatible with both the question text and diagram. GEOS then feeds the description to a geometric solver that attempts to determine the correct answer. In our experiments, GEOS achieves a 49% score on official SAT questions, and a score of 61% on practice questions. Finally, we show that by integrating textual and visual information, GEOS boosts the accuracy of dependency and semantic parsing of the question text.

**Do You See What I Mean? Visual Resolution of Linguistic Ambiguities***Yevgeni Berzak, Andrei Barbu, Daniel Harari, Boris Katz, and Shimon Ullman* 16:05-16:30

Understanding language goes hand in hand with the ability to integrate complex contextual information obtained via perception. In this work, we present a novel task for grounded language understanding: disambiguating a sentence given a visual scene which depicts one of the possible interpretations of that sentence. To this end, we introduce a new multimodal corpus containing ambiguous sentences, representing a wide range of syntactic, semantic and discourse ambiguities, coupled with videos that visualize the different interpretations for each sentence. We address this task by extending a vision model which determines if a sentence is depicted by a video. We demonstrate how such a model can be adjusted to recognize different interpretations of the same underlying sentence, allowing to disambiguate sentences in a unified fashion across the different ambiguity types.

**Efficient and Expressive Knowledge Base Completion Using Subgraph Feature Extraction***Matt Gardner and Tom Mitchell*

16:30-16:55

We explore some of the practicalities of using random walk inference methods, such as the Path Ranking Algorithm (PRA), for the task of knowledge base completion. We show that the random walk probabilities computed (at great expense) by PRA provide no discernible benefit to performance on this task, and so they can safely be dropped. This allows us to define a simpler algorithm for generating feature matrices from graphs, which we call subgraph feature extraction (SFE). In addition to being conceptually simpler than PRA, SFE is much more efficient, reducing computation by an order of magnitude, and more expressive, allowing for much richer features than just paths between two nodes in a graph. We show experimentally that this technique gives substantially better performance than PRA and its variants, improving mean average precision from .432 to .528 on a knowledge base completion task using the NELL knowledge base.

**Representing Text for Joint Embedding of Text and Knowledge Bases***Kristina Toutanova, Danqi Chen, Patrick Pantel, Hoifung Poon, Pallavi Choudhury, and Michael Gamon*

16:55-17:20

Models that learn to represent textual and knowledge base relations in the same continuous latent space are able to perform joint inferences among the two kinds of relations and obtain high accuracy on knowledge base completion (Riedel et al. 2013). In this paper we propose a model that captures the compositional structure of textual relations, and jointly optimizes entity, knowledge base, and textual relation representations. The proposed model significantly improves performance over a model that does not share parameters among textual relations with common sub-structure.

**Session 6D: Long Paper Posters**

Lower Level Foyer

Chair: Noah Smith

15:40–17:20

**A Utility Model of Authors in the Scientific Community***Yanchuan Sim, Bryan Routledge, and Noah A. Smith*

Authoring a scientific paper is a complex process involving many decisions. We introduce a probabilistic model of some of the important aspects of that process: that authors have individual preferences, that writing a paper requires trading off among the preferences of authors as well as extrinsic rewards in the form of community response to their papers, that preferences (of individuals and the community) and tradeoffs vary over time. Variants of our model lead to improved predictive accuracy of citations given texts and texts given authors. Further, our model's posterior suggests an interesting relationship between seniority and author choices.

**Finding Function in Form: Compositional Character Models for Open Vocabulary Word Representation***Wang Ling, Chris Dyer, Alan W Black, Isabel Trancoso, Ramon Fernandez, Silvio Amir, Luis Marujo, and Tiago Luis*

We introduce a model for constructing vector representations of words by composing characters using bidirectional LSTMs. Relative to traditional word representation models that have independent vectors for each word type, our model requires only a single vector per character type and a fixed set of parameters for the compositional model. Despite the compactness of this model and, more importantly, the arbitrary nature of the form–function relationship in language, our “composed” word representations yield state-of-the-art results in language modeling and part-of-speech tagging. Benefits over traditional baselines are particularly pronounced in morphologically rich languages (e.g., Turkish).

**Syntax-Aware Multi-Sense Word Embeddings for Deep Compositional Models of Meaning***Jianpeng Cheng and Dimitri Kartsaklis*

Deep compositional models of meaning acting on distributional representations of words in order to produce vectors of larger text constituents are evolving to a popular area of NLP research. We detail a compositional distributional framework based on a rich form of word embeddings that aims at facilitating the interactions between words in the context of a sentence. Embeddings and composition layers are jointly learned against a generic objective that enhances the vectors with syntactic information from the surrounding context. Furthermore, each word is associated with a number of senses, the most plausible of which is selected dynamically during the composition process. We evaluate the produced vectors qualitatively and quantitatively with positive results. At the sentence level, the effectiveness of the framework is demonstrated on the MSRPar task, for which we report results within the state-of-the-art range.

**Conversation Trees: A Grammar Model for Topic Structure in Forums***Annie Louis and Shay B. Cohen*

Online forum discussions proceed differently from face-to-face conversations and any single thread on an online forum contains posts on different subtopics. This work aims to characterize the content of a forum thread as a conversation tree of topics. We present models that jointly perform two tasks: segment a thread into subparts, and assign a topic to each part. Our core idea is a definition of topic structure using probabilistic grammars. By leveraging the flexibility of two grammar formalisms, Context-Free Grammars and Linear Context-Free Rewriting Systems, our models create desirable structures for forum threads: our topic segmentation is hierarchical, links non-adjacent segments on the same topic, and jointly labels the topic during segmentation. We show that our models outperform a number of tree generation baselines.

**Fast, Flexible Models for Discovering Topic Correlation across Weakly-Related Collections***Jingwei Zhang, Aaron Gerow, Jaan Altosaar, James Evans, and Richard Jean So*

Weak topic correlation across document collections with different numbers of topics in individual collections present challenges for existing cross-collection topic models. This paper introduces two probabilistic topic models, Correlated LDA (C-LDA) and Correlated HDP (C-HDP). These address problems that

can arise when analyzing large, asymmetric, and potentially weakly-related collections. Topic correlations in weakly-related collections typically lie in the tail of the topic distribution, where they would be overlooked by models unable to fit large numbers of topics. To efficiently model this long tail for large-scale analysis, our models implement a parallel sampling algorithm based on the Metropolis-Hastings and alias methods (Yuan et al., 2014). The models are first evaluated on synthetic data, generated to simulate various collection-level asymmetries. We then present a case study of modeling over 300k documents in collections of sciences and humanities research from JSTOR.

### **Molding CNNs for text: non-linear, non-consecutive convolutions**

*Tao Lei, Regina Barzilay, and Tommi Jaakkola*

The success of deep learning often derives from well-chosen operational building blocks. In this work, we revise the temporal convolution operation in CNNs to better adapt it to text processing. Instead of concatenating word representations, we appeal to tensor algebra and use low-rank n-gram tensors to directly exploit interactions between words already at the convolution stage. Moreover, we extend the n-gram convolution to non-consecutive words to recognize patterns with intervening words. Through a combination of low-rank tensors, and pattern weighting, we can efficiently evaluate the resulting convolution operation via dynamic programming. We test the resulting architecture on standard sentiment classification and news categorization tasks. Our model achieves state-of-the-art performance both in terms of accuracy and training speed. For instance, we obtain 51.2% accuracy on the fine-grained sentiment classification task.

### **Multi-Perspective Sentence Similarity Modeling with Convolutional Neural Networks**

*Hua He, Kevin Gimpel, and Jimmy Lin*

Modeling sentence similarity is complicated by the ambiguity and variability of linguistic expression. To cope with these challenges, we propose a model for comparing sentences that uses a multiplicity of perspectives. We first model each sentence using a convolutional neural network that extracts features at multiple levels of granularity and uses multiple types of pooling. We then compare our sentence representations at several granularities using multiple similarity metrics. We apply our model to three tasks, including the Microsoft Research paraphrase identification task and two SemEval semantic textual similarity tasks. We obtain strong performance on all tasks, rivaling or exceeding the state-of-the-art without using external resources like WordNet or parsers.

### **Posterior calibration and exploratory analysis for natural language processing models**

*Khanh Nguyen and Brendan O'Connor*

Many models in natural language processing define probabilistic distributions over linguistic structures. We argue that (1) the quality of a model's posterior distribution can and should be directly evaluated, as to whether probabilities correspond to empirical frequencies; and (2) NLP uncertainty can be projected not only to pipeline components, but also to exploratory data analysis, telling a user when to trust and not trust the NLP analysis. We present a method to analyze calibration, and apply it to compare the miscalibration of several commonly used models. We also contribute a coreference sampling algorithm that can create confidence intervals for a political event extraction task.

### **A Generative Word Embedding Model and its Low Rank Positive Semidefinite Solution**

*Shaohua Li, Jun Zhu, and Chunyan Miao*

Most existing word embedding methods can be categorized into Neural Embedding Models and Matrix Factorization (MF)-based methods. However some models are opaque to probabilistic interpretation, and MF-based methods, typically solved using Singular Value Decomposition (SVD), may incur loss of corpus information. In addition, it is desirable to incorporate global latent factors, such as topics, sentiments or writing styles, into the word embedding model. Since generative models provide a principled way to incorporate latent factors, we propose a generative word embedding model, which is easy to interpret, and can serve as a basis of more sophisticated latent factor models. The model inference reduces to a low rank weighted positive semidefinite approximation problem. Its optimization is approached by eigendecomposition on a submatrix, followed by online blockwise regression, which is scalable and avoids the information loss in SVD. In experiments on 7 common benchmark datasets, our vectors are competitive to word2vec, and better than other MF-based methods.

**Reading Documents for Bayesian Online Change Point Detection***Taehoon Kim and Jaesik Choi*

Modeling non-stationary time-series data for making predictions is a challenging but important task. One of the key issues is to identify long-term changes accurately in time-varying data. Bayesian Online Change Point Detection (BO-CPD) algorithms efficiently detect long-term changes without assuming the Markov property which is vulnerable to local signal noise. We propose a Document based BO-CPD (DBO-CPD) model which automatically detects long-term temporal changes of continuous variables based on a novel dynamic Bayesian analysis which combines a non-parametric regression, the Gaussian Process (GP), with generative models of texts such as news articles and posts on social networks. Since texts often include important clues of signal changes, DBO-CPD enables the accurate prediction of long-term changes accurately. We show that our algorithm outperforms existing BO-CPDs in two real-world datasets: stock prices and movie revenues.

**Session 6E: Short Paper Posters**

Lower Level Foyer

Chair: Ido Dagan

15:40–17:20

**Recognizing Textual Entailment Using Probabilistic Inference***Lei Sha, Sujian Li, Baobao Chang, Zhifang Sui, and Tingsong Jiang*

Recognizing Text Entailment (RTE) plays an important role in NLP applications including question answering, information retrieval, etc. In recent work, some research explore “deep” expressions such as discourse commitments or strict logic for representing the text. However, these expressions suffer from the limitation of inference inconvenience or translation loss. To overcome the limitations, in this paper, we propose to use the predicate-argument structures to represent the discourse commitments extracted from text. At the same time, with the help of the YAGO knowledge, we borrow the distant supervision technique to mine the implicit facts from the text. We also construct a probabilistic network for all the facts and conduct inference to judge the confidence of each fact for RTE. The experimental results show that our proposed method achieves a competitive result compared to the previous work.

**Chinese Semantic Role Labeling with Bidirectional Recurrent Neural Networks***Zhen Wang, Tingsong Jiang, Baobao Chang, and Zhifang Sui*

Traditional approaches to Chinese Semantic Role Labeling (SRL) almost rely on feature engineering, which means their performances are highly dependent on a large number of handcrafted features. Even worse, the long-range dependencies in a sentence can hardly be modeled by these methods. In this paper, we introduce bidirectional recurrent neural network (RNN) with long-short-term memory (LSTM) to capture bidirectional and long-range dependencies in a sentence with minimal feature engineering. Experimental results on Chinese Proposition Bank (CPB) show a significant improvement over the state-of-the-art methods. Moreover, our model makes it convenient to introduce heterogeneous resource, which makes a further improvement to our experimental performance.

**Unsupervised Negation Focus Identification with Word-Topic Graph Model***Bowei Zou, Guodong Zhou, and Qiaoming Zhu*

Due to the commonality in natural language, negation focus plays a critical role in deep understanding of context. However, existing studies for negation focus identification major on supervised learning which is time-consuming and expensive due to manual preparation of annotated corpus. To address this problem, we propose an unsupervised word-topic graph model to represent and measure the focus candidates from both lexical and topic perspectives. Moreover, we propose a document-sensitive biased PageRank algorithm to optimize the ranking scores of focus candidates. Evaluation on the \*SEM 2012 shared task corpus shows that our proposed method outperforms the state of the art on negation focus identification.

**Reverse-engineering Language: A Study on the Semantic Compositionality of German Compounds***Corina Dima*

In this paper we analyze the performance of different composition models on a large dataset of German compound nouns. Given a vector space model for the German language, we try to reconstruct the observed representation (the corpus-estimated vector) of a compound by composing the observed representations of its two immediate constituents. We explore the composition models proposed in the literature and also present a new, simple model that achieves the best performance on our dataset.

**Event Detection and Factuality Assessment with Non-Expert Supervision***Kenton Lee, Yoav Artzi, Jejin Choi, and Luke Zettlemoyer*

Events are communicated in natural language with varying degrees of certainty. For example, if you are “hoping for a raise,” it may be somewhat less likely than if you are “expecting” one. To study these distinctions, we present scalable, high-quality annotation schemes for event detection and fine-grained factuality assessment. We find that non-experts, with very little training, can reliably provide judgments about what events are mentioned and the extent to which the author thinks they actually happened. We also show how such data enables the development of regression models for fine-grained scalar factuality predictions that outperform strong baselines.

**Large-Scale Acquisition of Entailment Pattern Pairs by Exploiting Transitivity***Julien Kloetzer, Kentaro Torisawa, Chikara Hashimoto, and Jong-Hoon Oh*

We propose a novel method for acquiring entailment pairs of binary patterns on a large-scale. This method exploits the transitivity of entailment and a self-training scheme to improve the performance of an already strong supervised classifier for entailment, and unlike previous methods that exploit transitivity, it works on a large-scale. With it we acquired 138.1 million pattern pairs with 70% precision with such non-trivial lexical substitution as “use Y to distribute X” -> “X is available on Y” whose extraction is considered difficult. This represents 50.4 million more pattern pairs (a 57.5% increase) than what our supervised baseline extracted at the same precision.

### **Context-Dependent Knowledge Graph Embedding**

*Yuanfei Luo, Quan Wang, Bin Wang, and Li Guo*

We consider the problem of embedding knowledge graphs (KGs) into continuous vector spaces. Existing methods can only deal with explicit relationships within each triple, i.e., local connectivity patterns, but cannot handle implicit relationships across different triples, i.e., contextual connectivity patterns. This paper proposes context-dependent KG embedding, a two-stage scheme that takes into account both types of connectivity patterns and obtains more accurate embeddings. We evaluate our approach on the tasks of link prediction and triple classification, and achieve significant and consistent improvements over state-of-the-art methods.

### **Learning to Identify the Best Contexts for Knowledge-based WSD**

*Evgenia Wasserman Pritsker, William Cohen, and Einat Minkov*

We outline a learning framework that aims at identifying useful contextual cues for knowledge-based word sense disambiguation. The usefulness of individual context words is evaluated based on diverse lexico-statistical and syntactic information, as well as simple word distance. Experiments using two different knowledge-based methods and benchmark datasets show significant improvements due to context modeling, beating the conventional window-based approach.

### **Measuring Prerequisite Relations Among Concepts**

*Chen Liang, Zhaohui Wu, Wenyi Huang, and C. Lee Giles*

A prerequisite relation describes a basic relation among concepts in cognition, education and other areas. However, as a semantic relation, it has not been well studied in computational linguistics. We investigate the problem of measuring prerequisite relations among concepts and propose a simple link-based metric, namely reference distance (RefD), that effectively models the relation by measuring how differently two concepts refer to each other. Evaluations on two datasets that include seven domains show that our single metric based method outperforms existing supervised learning based methods.

### **Adapting Phrase-based Machine Translation to Normalise Medical Terms in Social Media Messages**

*Nut Limsopatham and Nigel Collier*

Previous studies have shown that health reports in social media, such as DailyStrength and Twitter, have potential for monitoring health conditions (e.g. adverse drug reactions, infectious diseases) in particular communities. However, in order for a machine to understand and make inferences on these health conditions, the ability to recognise when laymen’s terms refer to a particular medical concept (i.e. text normalisation) is required. To achieve this, we propose to adapt an existing phrase-based machine translation (MT) technique and a vector representation of words to map between a social media phrase and a medical concept. We evaluate our proposed approach using a collection of phrases from tweets related to adverse drug reactions. Our experimental results show that the combination of a phrase-based MT technique and the similarity between word vector representations outperforms the baselines that apply only either of them by up to 55%.

### **Script Induction as Language Modeling**

*Rachel Rudinger, Pushpendre Rastogi, Francis Ferraro, and Benjamin Van Durme*

The narrative cloze is an evaluation metric commonly used for work on automatic script induction. While prior work in this area has focused on count-based methods from distributional semantics, such as pointwise mutual information, we argue that the narrative cloze can be productively reframed as a language modeling task. By training a discriminative language model for this task, we attain improvements of up to 27 percent over prior methods on standard narrative cloze metrics.

### **Online Learning of Interpretable Word Embeddings**

*Hongyin Luo, Zhiyuan Liu, Huanbo Luan, and Maosong Sun*

Word embeddings encode semantic meanings of words into low-dimension word vectors. In most word embeddings, one cannot interpret the meanings of specific dimensions of those word vectors. Non-negative matrix factorization (NMF) has been proposed to learn interpretable word embeddings via non-negative constraints. However, NMF methods suffer from scale and memory issue because they have to maintain a global matrix for learning. To alleviate this challenge, we propose online learning of interpretable word embeddings from streaming text data. Experiments show that our model consistently outperforms the state-of-the-art word embedding methods in both representation ability and interpretability. The source code of this paper can be obtained from <http://github.com/skTim/OIWE>.

### A Strong Lexical Matching Method for the Machine Comprehension Test

*Ellery Smith, Nicola Greco, Matko Bosnjak, and Andreas Vlachos*

Machine comprehension of text is the overarching goal of a great deal of research in natural language processing. The Machine Comprehension Test (Richardson et al., 2013) was recently proposed to assess methods on an open-domain, extensible, and easy-to-evaluate task consisting of two datasets. In this paper we develop a lexical matching method that takes into account multiple context windows, question types and coreference resolution. We show that the proposed method outperforms the baseline of Richardson et al. (2013), and despite its relative simplicity, is comparable to recent work using machine learning. We hope that our approach will inform future work on this task. Furthermore, we argue that MC500 is harder than MC160 due to the way question answer pairs were created.

## Main Conference: Monday, September 21

### Overview

---

07:30–18:00	<b>Registration</b>	Culturgest		
08:00–09:00	<b>Morning Coffee</b>	Foyers		
09:00–10:00	<b>Session P4: Plenary Session</b>	Main Auditorium		
09:00–09:05	<i>Best Paper Awards (Chris Callison-Burch and Jian Su)</i>			
09:05–09:30	<i>Broad-coverage CCG Semantic Parsing with AMR Yoav Artzi, Kenton Lee, and Luke Zettlemoyer</i>	(p. 119)		
09:30–09:55	<i>Semantically Conditioned LSTM-based Natural Language Generation for Spoken Dialogue Systems Tsung-Hsien Wen, Milica Gasic, Nikola Mrkšić, Pei-Hao Su, David Vandyke, and Steve Young</i>	(p. 119)		
09:55–10:05	<i>A large annotated corpus for learning natural language inference Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning</i>	(p. 119)		
10:05–10:30	<b>Coffee break</b>	Foyers		
10:30–12:10	<b>Session 7</b>			
	Semantics (Long +TACL Papers)	Information Extraction (Long Papers)	Computational Psycholinguistics / Machine Translation (Long Papers)	Long +TACL Paper Posters
	Main Auditorium	Small Auditorium	Room 2	Short Paper Posters <i>Lower Level Foyer</i>
12:10–13:30	<b>Lunch</b>			
13:30–15:15	<b>Session 8</b>			
	Fun and Quirky Topics (Short Papers)	Semantics (Short Papers)	Statistical Models and Machine Learning Methods / Machine Translation (Short Papers)	Long Paper Posters <hr/> Short Paper Posters
	Main Auditorium	Small Auditorium	Room 2	<i>Lower Level Foyer</i>
15:15–15:40	<b>Coffee break</b>	Foyers		
15:40–17:20	<b>Session 9</b>			

Statistical Models and Machine Learn- ing Methods (Long + TACL Papers) <i>Main Auditorium</i>	Text Mining and NLP Applications (Long Papers) <i>Small Auditorium</i>	Spoken Language Processing and Language Model- ing (Long Papers) <i>Room 2</i>	Long Paper Posters <hr/> Short Paper Posters <i>Lower Level Foyer</i>
---	---	--	---

17:30–17:50 **Session P5: Closing Remarks**

*Main Auditorium*

18:30–20:00 **Farewell Drink**

*Culturgest*

## Abstracts: Session P4

---

### Session P4: Plenary Session

Main Auditorium

#### **Broad-coverage CCG Semantic Parsing with AMR**

*Yoav Artzi, Kenton Lee, and Luke Zettlemoyer*

09:05-09:30

We propose a grammar induction technique for AMR semantic parsing. While previous grammar induction techniques were designed to re-learn a new parser for each target application, the recently annotated AMR Bank provides a unique opportunity to induce a single model for understanding broad-coverage newswire text and support a wide range of applications. We present a new model that combines CCG parsing to recover compositional aspects of meaning and a factor graph to model non-compositional phenomena, such as anaphoric dependencies. Our approach achieves 66.2 Smatch F1 score on the AMR bank, significantly outperforming the previous state of the art.

#### **Semantically Conditioned LSTM-based Natural Language Generation for Spoken Dialogue Systems**

*Tsung-Hsien Wen, Milica Gasic, Nikola Mrkšić, Pei-Hao Su, David Vandyke, and Steve Young*

09:30-09:55

Natural language generation (NLG) is a critical component of spoken dialogue and it has a significant impact both on usability and perceived quality. Most NLG systems in common use employ rules and heuristics and tend to generate rigid and stylised responses without the natural variation of human language. They are also not easily scaled to systems covering multiple domains and languages. This paper presents a statistical language generator based on a semantically controlled Long Short-term Memory (LSTM) structure. The LSTM generator can learn from unaligned data by jointly optimising sentence planning and surface realisation using a simple cross entropy training criterion, and language variation can be easily achieved by sampling from output candidates. With fewer heuristics, an objective evaluation in two differing test domains showed the proposed method improved performance compared to previous methods. Human judges scored the LSTM system higher on informativeness and naturalness and overall preferred it to the other systems.

#### **A large annotated corpus for learning natural language inference**

*Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning* 09:55-10:05

Understanding entailment and contradiction is fundamental to understanding natural language, and inference about entailment and contradiction is a valuable testing ground for the development of semantic representations. However, machine learning research in this area has been dramatically limited by the lack of large-scale resources. To address this, we introduce the Stanford Natural Language Inference corpus, a new, freely available collection of labeled sentence pairs, written by humans doing a novel grounded task based on image captioning. At 570K pairs, it is two orders of magnitude larger than all other resources of its type. This increase in scale allows lexicalized classifiers to outperform some sophisticated existing entailment models, and it allows a neural network-based model to perform competitively on natural language inference benchmarks for the first time.

## Session 7 Overview

---

	<b>Track A</b>	<b>Track B</b>	<b>Track C</b>
	Semantics (Long +TACL Papers) Main Auditorium	Information Extraction (Long Papers) Small Auditorium	Computational Psycholinguistics / Machine Translation (Long Papers) Room 2
10:30	Do Multi-Sense Embeddings Improve Natural Language Understanding? <i>Li and Jurafsky</i>	Distant Supervision for Relation Extraction via Piecewise Convolutional Neural Networks <i>Zeng, Liu, Chen, and Zhao</i>	A Computational Cognitive Model of Novel Word Generalization <i>Nematzadeh, Grant, and Stevenson</i>
10:55	Learning Semantic Composition to Detect Non-compositionality of Multiword Expressions <i>Yazdani, Farahmand, and Henderson</i>	CORE: Context-Aware Open Relation Extraction with Factorization Machines <i>Petroni, Del Corro, and Gemulla</i>	Personality Profiling of Fictional Characters using Sense-Level Links between Lexical Resources <i>Flekova and Gurevych</i>
11:20	Solving General Arithmetic Word Problems <i>Roy and Roth</i>	Improved Relation Extraction with Feature-Rich Compositional Embedding Models <i>Gormley, Yu, and Dredze</i>	Leave-one-out Word Alignment without Garbage Collector Effects <i>Wang, Utiyama, Finch, Watanabe, and Sumita</i>
11:45	[TACL] From Paraphrase Database to Compositional Paraphrase Model and Back <i>Wieting, Bansal, Gimpel, Livescu, and Roth</i>	Classifying Relations via Long Short Term Memory Networks along Shortest Dependency Paths <i>Xu, Mou, Li, Chen, Peng, and Jin</i>	Generalized Agreement for Bidirectional Word Alignment <i>Liu, Liu, Sun, Luan, and Yu</i>

**Track D:** Long +TACL Paper Posters

Chair: David McClosky

Lower Level Foyer

10:30–12:10

---

### Poster Cluster 1: Word Segmentation, Tagging and Parsing (P1-6)

- A Transition-based Model for Joint Segmentation, POS-tagging and Normalization  
*T. Qian, Y. Zhang, M. Zhang, Y. Ren, and D. Ji*
- Multilingual discriminative lexicalized phrase structure parsing  
*B. Crabbe*
- Hierarchical Low-Rank Tensors for Multilingual Transfer Parsing  
*Y. Zhang and R. Barzilay*
- Diversity in Spectral Learning for Natural Language Parsing  
*S. Narayan and S. B. Cohen*

- Transition-based Dependency Parsing Using Two Heterogeneous Gated Recursive Neural Networks  
*X. Chen, Y. Zhou, C. Zhu, X. Qiu, and X. Huang*
- [TACL] A Graph-based Lattice Dependency Parser for Joint Morphological Segmentation and Syntactic Analysis  
*W. Seeker and Ö. Çetinoğlu*

**Track E: Short Paper Posters**

Lower Level Foyer

Chair: Jun-Ping Ng

10:30-12:10

---

**Poster Cluster 1: Spoken Language Processing (P1-3)**

- Turn-taking phenomena in incremental dialogue systems  
*H. Khouzaimi, R. Laroche, and F. Lefevre*
- Hierarchical Latent Words Language Models for Robust Modeling to Out-Of Domain Tasks  
*R. Masumura, T. Asami, T. Oba, H. Masataki, S. Sakauchi, and A. Ito*
- A Coarse-Grained Model for Optimal Coupling of ASR and SMT Systems for Speech Translation  
*G. Kumar, G. Blackwood, J. Trmal, D. Povey, and S. Khudanpur*

**Poster Cluster 2: Summarization (P4-18)**

- Abstractive Multi-document Summarization with Semantic Information Extraction  
*W. Li*
- Concept-based Summarization using Integer Linear Programming: From Concept Pruning to Multiple Optimal Solutions  
*F. Boudin, H. Mougar, and B. Favre*
- GhostWriter: Using an LSTM for Automatic Rap Lyric Generation  
*P. Potash, A. Romanov, and A. Rumshisky*
- Better Summarization Evaluation with Word Embeddings for ROUGE  
*J.-P. Ng and V. Abrecht*
- Krimping texts for better summarization  
*M. Litvak, M. Last, and N. Vanetik*
- From the Virtual to the RealWorld: Referring to Objects in Real-World Spatial Scenes  
*D. Gkatzia, V. Rieser, P. Bartie, and W. Mackaness*
- An Unsupervised Bayesian Modelling Approach for Storyline Detection on News Articles  
*D. Zhou, H. Xu, and Y. He*
- Topical Coherence for Graph-based Extractive Summarization  
*D. Parveen, H.-M. Ramsl, and M. Strube*
- Summarizing Student Responses to Reflection Prompts  
*W. Luo and D. Litman*

- Extractive Summarization by Maximizing Semantic Volume  
*D. Yogatama, F. Liu, and N. A. Smith*
- LCSTS: A Large Scale Chinese Short Text Summarization Dataset  
*B. Hu, Q. Chen, and F. Zhu*
- Discourse Planning with an N-gram Model of Relations  
*O. Biran and K. McKeown*
- Experiments with Generative Models for Dependency Tree Linearization  
*R. Futrell and E. Gibson*
- Summarization Based on Embedding Distributions  
*H. Kobayashi, M. Noguchi, and T. Yatsuka*
- Reversibility reconsidered: finite-state factors for efficient probabilistic sampling in parsing and generation  
*M. Dymetman, S. Venkatapathy, and C. Xiao*

## Abstracts: Session 7

---

### Session 7A: Semantics (Long +TACL Papers)

Main Auditorium

Chair: Luke Zettlemoyer

#### **Do Multi-Sense Embeddings Improve Natural Language Understanding?**

*Jiwei Li and Dan Jurafsky*

10:30–10:55

Learning a distinct representation for each sense of an ambiguous word could lead to more powerful and fine-grained models of vector-space representations. Yet while ‘multi-sense’ methods have been proposed and tested on artificial word-similarity tasks, we don’t know if they improve real natural language understanding tasks. In this paper we introduce a multi-sense embedding model based on Chinese Restaurant Processes that achieves state of the art performance on matching human word similarity judgments, and propose a pipelined architecture for incorporating multi-sense embeddings into language understanding. We then test the performance of our model on part-of-speech tagging, named entity recognition, sentiment analysis, semantic relation identification and semantic relatedness, controlling for embedding dimensionality. We find that multi-sense embeddings do improve performance on some tasks (part-of-speech tagging, semantic relation identification, semantic relatedness) but not on others (named entity recognition, various forms of sentiment analysis). We discuss how these differences may be caused by the different role of word sense information in each of the tasks. The results highlight the importance of testing embedding models in real applications.

#### **Learning Semantic Composition to Detect Non-compositionality of Multiword Expressions**

*Majid Yazdani, Meghdad Farahmand, and James Henderson*

10:55–11:20

Non-compositionality of multi word expressions is an intriguing problem that can be the source of error in a variety of NLP tasks such as language generation, machine translation and word sense disambiguation. In this work we present a method of detecting non-compositional English noun compounds by learning a composition function. We explore a range of possible models for semantic composition, empirically evaluate these models and propose an improvement method over the most accurate ones. We show that a complex function such as polynomial projection can learn semantic composition and identify non-compositionality in an unsupervised way, beating all other baselines ranging from simple to complex. We show further improvements by also training a decomposition function, and with a form of EM algorithm over latent compositionality annotations.

#### **Solving General Arithmetic Word Problems**

*Subhro Roy and Dan Roth*

11:20–11:45

This paper presents a novel approach to automatically solving arithmetic word problems. This is the first algorithmic approach that can handle arithmetic problems with multiple steps and operations, without depending on additional annotations or predefined templates. We develop a theory for expression trees that can be used to represent and evaluate the target arithmetic expressions; we use it to uniquely decompose the target arithmetic problem to multiple classification problems; we then compose an equation tree, combining these with world knowledge through a constrained inference framework. Our classifiers gain from the use of quantity schemas that supports better extraction of features. Experimental results show that our method outperforms existing systems, achieving state of the art performance on benchmark datasets of arithmetic word problems.

#### **[TACL] From Paraphrase Database to Compositional Paraphrase Model and Back**

*John Wieting, Mohit Bansal, Kevin Gimpel, Karen Livescu, and Dan Roth*

11:45–12:10

The Paraphrase Database (PPDB; Ganitkevitch et al., 2013) is an extensive semantic resource, consisting of a list of phrase pairs with (heuristic) confidence estimates. However, it is still unclear how it can best be used, due to the heuristic nature of the confidences and its necessarily incomplete coverage. We propose models to leverage the phrase pairs from the PPDB to build parametric paraphrase models that score paraphrase pairs more accurately than the PPDB’s internal scores while simultaneously improving its coverage. They allow for learning phrase embeddings as well as improved word embeddings. Moreover, we introduce two new, manually annotated datasets to evaluate short-phrase paraphrasing models.

Using our paraphrase model trained using PPDB, we achieve state-of-the-art results on standard word and bigram similarity tasks and beat strong baselines on our new short phrase paraphrase tasks.

**Session 7B: Information Extraction (Long Papers)**

Small Auditorium

Chair: Shiqi Zhao

**Distant Supervision for Relation Extraction via Piecewise Convolutional Neural Networks**

Daojian Zeng, Kang Liu, Yubo Chen, and Jun Zhao

10:30–10:55

Two problems arise when using distant supervision for relation extraction. First, in this method, an already existing knowledge base is heuristically aligned to texts, and the alignment results are treated as labeled data. However, the heuristic alignment can fail, resulting in wrong label problem. In addition, in previous approaches, statistical models have typically been applied to ad hoc features. The noise that originates from the feature extraction process can cause poor performance. In this paper, we propose a novel model dubbed the Piecewise Convolutional Neural Networks (PCNNs) with multi-instance learning to address these two problems. To solve the first problem, distant supervised relation extraction is treated as a multi-instance problem in which the uncertainty of instance labels is taken into account. To address the latter problem, we avoid feature engineering and instead adopt convolutional architecture with piecewise max pooling to automatically learn relevant features. Experiments show that our method is effective and outperforms several competitive baseline methods.

**CORE: Context-Aware Open Relation Extraction with Factorization Machines**

Fabio Petroni, Luciano Del Corro, and Rainer Gemulla

10:55–11:20

We propose CORE, a novel matrix factorization model that leverages contextual information for open relation extraction. Our model is based on factorization machines and integrates facts from various sources, such as knowledge bases or open information extractors, as well as the context in which these facts have been observed. We argue that integrating contextual information—such as metadata about extraction sources, lexical context, or type information—significantly improves prediction performance. Open information extractors, for example, may produce extractions that are unspecific or ambiguous when taken out of context. Our experimental study on a large real-world dataset indicates that CORE has significantly better prediction performance than state-of-the-art approaches when contextual information is available.

**Improved Relation Extraction with Feature-Rich Compositional Embedding Models**

Matthew R. Gormley, Mo Yu, and Mark Dredze

11:20–11:45

Compositional embedding models build a representation (or embedding) for a linguistic structure based on its component word embeddings. We propose a Feature-rich Compositional Embedding Model (FCM) for relation extraction that is expressive, generalizes to new domains, and is easy-to-implement. The key idea is to combine both (unlexicalized) hand-crafted features with learned word embeddings. The model is able to directly tackle the difficulties met by traditional compositional embeddings models, such as handling arbitrary types of sentence annotations and utilizing global information for composition. We test the proposed model on two relation extraction tasks, and demonstrate that our model outperforms both previous compositional models and traditional feature rich models on the ACE 2005 relation extraction task, and the SemEval 2010 relation classification task. The combination of our model and a log-linear classifier with hand-crafted features gives state-of-the-art results. We made our implementation available for general use.

**Classifying Relations via Long Short Term Memory Networks along Shortest Dependency Paths**

Yan Xu, Lili Mou, Ge Li, Yunchuan Chen, Hao Peng, and Zhi Jin

11:45–12:10

Relation classification is an important research arena in the field of natural language processing (NLP). In this paper, we present SDP-LSTM, a novel neural network to classify the relation of two entities in a sentence. The neural architecture leverages the shortest dependency path (SDP) between two entities; multichannel recurrent neural networks, with long short term memory (LSTM) units, pick up heterogeneous information along the SDP. Our proposed model has several distinct features: (1) The shortest dependency paths retain most relevant information (to relation classification), while eliminating irrelevant words in the sentence. (2) The multichannel LSTM networks allow effective information integration from heterogeneous sources over the dependency paths. (3) A customized dropout strategy regularizes the neural network to alleviate overfitting. We test our model on the SemEval 2010 relation classification task, and achieve an F1-score of 83.7%, higher than competing methods in the literature.

**Session 7C: Computational Psycholinguistics / Machine Translation (Long Papers)**

Room 2

*Chair: Nizar Habash***A Computational Cognitive Model of Novel Word Generalization***Aida Nematzadeh, Erin Grant, and Suzanne Stevenson*

10:30-10:55

A key challenge in vocabulary acquisition is learning which of the many possible meanings is appropriate for a word. The word generalization problem refers to how children associate a word such as dog with a meaning at the appropriate category level in the taxonomy of objects, such as Dalmatians, dogs, or animals. We present the first computational study of word generalization integrated within a word learning model. The model simulates child and adult patterns of word generalization in a word-learning task. These patterns arise due to the interaction of type and token frequencies in the input data, an influence often observed in people's generalization of linguistic categories.

**Personality Profiling of Fictional Characters using Sense-Level Links between Lexical Resources***Lucie Flekova and Iryna Gurevych*

10:55-11:20

This study focuses on personality prediction of protagonists in novels based on the Five-Factor Model of personality. We present and publish a novel collaboratively built dataset of fictional character personality and design our task as a text classification problem. We incorporate a range of semantic features, including WordNet and VerbNet sense-level information and word vector representations. We evaluate three machine learning models based on the speech, actions and predicatives of the main characters, and show that especially the lexical-semantic features significantly outperform the baselines. Qualitative analysis reveals that the most predictive features correspond to the reported findings in personality psychology and NLP experiments on human personality.

**Leave-one-out Word Alignment without Garbage Collector Effects***Xiaolin Wang, Masao Utiyama, Andrew Finch, Taro Watanabe, and Eiichiro Sumita* 11:20-11:45

Expectation-maximization algorithms, such as those implemented in GIZA++ pervade the field of unsupervised word alignment. However, these algorithms have a problem of over-fitting, leading to "garbage collector effects," where rare words tend to be erroneously aligned to untranslated words. This paper proposes a leave-one-out expectation-maximization algorithm for unsupervised word alignment to address this problem. The proposed method excludes information derived from the alignment of a sentence pair from the alignment models used to align it. This prevents erroneous alignments within a sentence pair from supporting themselves. Experimental results on Chinese-English and Japanese-English corpora show that the F\$<sub>1</sub>\$, precision and recall of alignment were consistently increased by 5.0% - 17.2%, and BLEU scores of end-to-end translation were raised by 0.03 - 1.30. The proposed method also outperformed \$l\_0\$-normalized GIZA++ and Kneser-Ney smoothed GIZA++.

**Generalized Agreement for Bidirectional Word Alignment***Chunyang Liu, Yang Liu, Maosong Sun, Huanbo Luan, and Heng Yu* 11:45-12:10

While agreement-based joint training has proven to deliver state-of-the-art alignment accuracy, the produced word alignments are usually restricted to one-to-one mappings because of the hard constraint on agreement. We propose a general framework to allow for arbitrary loss functions that measure the disagreement between asymmetric alignments. The loss functions can not only be defined between asymmetric alignments but also between alignments and other latent structures such as phrase segmentation. We use a Viterbi EM algorithm to train the joint model since the inference is intractable. Experiments on Chinese-English translation show that joint training with generalized agreement achieves significant improvements over two state-of-the-art alignment methods.

**Session 7D: Long +TACL Paper Posters**

Lower Level Foyer

Chair: David McClosky

10:30-12:10

**A Transition-based Model for Joint Segmentation, POS-tagging and Normalization***Tao Qian, Yue Zhang, Meishan Zhang, Yafeng Ren, and Donghong Ji*

We propose a transition-based model for joint word segmentation, POS tagging and text normalization. Different from previous methods, the model can be trained on standard text corpora, overcoming the lack of annotated microblog corpora. To evaluate our model, we develop an annotated corpus based on microblogs. Experimental results show that our joint model can help improve the performance of word segmentation on microblogs, giving an error reduction in segmentation accuracy of 12.02%, compared to the traditional approach.

**Multilingual discriminative lexicalized phrase structure parsing***Benoit Crabbé*

We provide a generalization of discriminative lexicalized shift reduce parsing techniques for phrase structure grammar to a wide range of morphologically rich languages. The model is efficient and outperforms recent strong baselines on almost all languages considered. It takes advantage of a dependency based modelling of morphology and a shallow modelling of constituency boundaries.

**Hierarchical Low-Rank Tensors for Multilingual Transfer Parsing***Yuan Zhang and Regina Barzilay*

Accurate multilingual transfer parsing typically relies on careful feature engineering. In this paper, we propose a hierarchical tensor-based approach for this task. This approach induces a compact feature representation by combining atomic features. However, unlike traditional tensor models, it enables us to incorporate prior knowledge about desired feature interactions, eliminating invalid feature combinations. To this end, we use a hierarchical structure that uses intermediate embeddings to capture desired feature combinations. Algebraically, this hierarchical tensor is equivalent to the sum of traditional tensors with shared components, and thus can be effectively trained with standard online algorithms. In both unsupervised and semi-supervised transfer scenarios, our hierarchical tensor consistently improves UAS and LAS over state-of-the-art multilingual transfer parsers and the traditional tensor model across 10 different languages.

**Diversity in Spectral Learning for Natural Language Parsing***Shashi Narayan and Shay B. Cohen*

We describe an approach to create a diverse set of predictions with spectral learning of latent-variable PCFGs (L-PCFGs). Our approach works by creating multiple spectral models where noise is added to the underlying features in the training set before the estimation of each model. We describe three ways to decode with multiple models. In addition, we describe a simple variant of the spectral algorithm for L-PCFGs that is fast and leads to compact models. Our experiments for natural language parsing, for English and German, show that we get a significant improvement over baselines comparable to state of the art. For English, we achieve the F1 score of 90.18, and for German we achieve the F1 score of 83.38.

**Transition-based Dependency Parsing Using Two Heterogeneous Gated Recursive Neural Networks***Xinchi Chen, Yaqian Zhou, Chenxi Zhu, Xipeng Qiu, and Xuanjing Huang*

Recently, neural network based dependency parsing has attracted much interest, which can effectively alleviate the problems of data sparsity and feature engineering by using the dense features. However, it is still a challenge problem to sufficiently model the complicated syntactic and semantic compositions of the dense features in neural network based methods. In this paper, we propose two heterogeneous gated recursive neural networks: tree structured gated recursive neural network (Tree-GRNN) and directed acyclic graph structured gated recursive neural network (DAG-GRNN). Then we integrate them to automatically learn the compositions of the dense features for transition-based dependency parsing. Specifically, Tree-GRNN models the feature combinations for the trees in stack, which already have partial dependency structures. DAG-GRNN models the feature combinations of the nodes whose dependency relations have not been built yet. Experiment results on two prevalent benchmark datasets (PTB3 and CTB5) show the effectiveness of our proposed model.

**[TACL] A Graph-based Lattice Dependency Parser for Joint Morphological Segmentation and Syntactic Analysis**

Wolfgang Seeker and Özlem Çetinoğlu

Space-delimited words in Turkish and Hebrew text can be further segmented into meaningful units, but syntactic and semantic context is necessary to predict segmentation. At the same time, predicting correct syntactic structures relies on correct segmentation. We present a graph-based lattice dependency parser that operates on morphological lattices to represent different segmentations and morphological analyses for a given input sentence. The lattice parser predicts a dependency tree over a path in the lattice and thus solves the joint task of segmentation, morphological analysis, and syntactic parsing. We conduct experiments on the Turkish and the Hebrew treebank and show that the joint model outperforms three state-of-the-art pipeline systems on both data sets. Our work corroborates findings from constituency lattice parsing for Hebrew and presents the first results for full lattice parsing on Turkish.

**Session 7E: Short Paper Posters**

Lower Level Foyer

Chair: Jun-Ping Ng

10:30-12:10

**Turn-taking phenomena in incremental dialogue systems***Hatim Khouzaimi, Romain Laroche, and Fabrice Lefevre*

In this paper, a turn-taking phenomenon taxonomy is introduced, organised according to the level of information conveyed. It is aimed to provide a better grasp of the behaviours used by humans while talking to each other, so that they can be methodically replicated in dialogue systems. Five interesting phenomena have been implemented in a simulated environment: the system barge-in because of an unclear, an incoherent or a sufficient message, the feedback and the user barge-in. The aim of the experiment is to illustrate that some phenomena are worth implementing in some cases and others are not.

**Hierarchical Latent Words Language Models for Robust Modeling to Out-Of Domain Tasks***Ryo Masumura, Taichi Asami, Takanobu Oba, Hirokazu Masataki, Sumitaka Sakauchi, and Akinori Ito*

This paper focuses on language modeling with adequate robustness to support different domain tasks. To this end, we propose a hierarchical latent word language model (h-LWLM). The proposed model can be regarded as a generalized form of the standard LWLMs. The key advance is introducing a multiple latent variable space with hierarchical structure. The structure can flexibly take account of linguistic phenomena not present in the training data. This paper details the definition as well as a training method based on layer-wise inference and a practical usage in natural language processing tasks with an approximation technique. Experiments on speech recognition show the effectiveness of h-LWLM in out-of domain tasks.

**A Coarse-Grained Model for Optimal Coupling of ASR and SMT Systems for Speech Translation***Gaurav Kumar, Graeme Blackwood, Jan Trmal, Daniel Povey, and Sanjeev Khudanpur*

Speech translation is conventionally carried out by cascading an automatic speech recognition (ASR) and a statistical machine translation (SMT) system. The hypotheses chosen for translation are based on the ASR system's acoustic and language model scores, and typically optimized for word error rate, ignoring the intended downstream use: automatic translation. In this paper, we present a coarse-to-fine model that uses features from the ASR and SMT systems to optimize this coupling. We demonstrate that several standard features utilized by ASR and SMT systems can be used in such a model at the speech-translation interface, and we provide empirical results on the Fisher Spanish-English speech translation corpus.

**Abstractive Multi-document Summarization with Semantic Information Extraction**  
*Wei Li*

This paper proposes a novel approach to generate abstractive summary for multiple documents by extracting semantic information from texts. The concept of Basic Semantic Unit (BSU) is defined to describe the semantics of an event or action. A semantic link network on BSUs is constructed to capture the semantic information of texts. Summary structure is planned with sentences generated based on the semantic link network. Experiments demonstrate that the approach is effective in generating informative, coherent and compact summary.

**Concept-based Summarization using Integer Linear Programming: From Concept Pruning to Multiple Optimal Solutions***Florian Boudin, Hugo Mougard, and Benoit Favre*

In concept-based summarization, sentence selection is modelled as a budgeted maximum coverage problem. As this problem is NP-hard, pruning low-weight concepts is required for the solver to find optimal solutions efficiently. This work shows that reducing the number of concepts in the model leads to lower ROUGE scores, and more importantly to the presence of multiple optimal solutions. We address these issues by extending the model to provide a single optimal solution, and eliminate the need for concept pruning using an approximation algorithm that achieves comparable performance to exact inference.

**GhostWriter: Using an LSTM for Automatic Rap Lyric Generation***Peter Potash, Alexey Romanov, and Anna Rumshisky*

This paper demonstrates the effectiveness of a Long Short-Term Memory language model in our initial efforts to generate unconstrained rap lyrics. The goal of this model is to generate lyrics that are similar in style to that of a given rapper, but not identical to existing lyrics: this is the task of ghostwriting. Unlike previous work, which defines explicit templates for lyric generation, our model defines its own rhyme scheme, line length, and verse length. Our experiments show that a Long Short-Term Memory language model produces better “ghostwritten” lyrics than a baseline model.

**Better Summarization Evaluation with Word Embeddings for ROUGE***Jun-Ping Ng and Viktoria Abrecht*

ROUGE is a widely adopted, automatic evaluation measure for text summarization. While it has been shown to correlate well with human judgements, it is biased towards surface lexical similarities. This makes it unsuitable for the evaluation of abstractive summarization, or summaries with substantial paraphrasing. We study the effectiveness of word embeddings to overcome this disadvantage of ROUGE. Specifically, instead of measuring lexical overlaps, word embeddings are used to compute the semantic similarity of the words used in summaries instead. Our experimental results show that our proposal is able to achieve better correlations with human judgements when measured with the Spearman and Kendall rank coefficients.

**Krimping texts for better summarization***Marina Litvak, Mark Last, and Natalia Vanetik*

Automated text summarization is aimed at extracting essential information from original text and presenting it in a minimal, often predefined, number of words. In this paper, we introduce a new approach for unsupervised extractive summarization, based on the Minimum Description Length (MDL) principle. The approach represents a text as a transactional dataset, with sentences as transactions, and then describes it by itemsets that stand for frequent sequences of words. The summary is then compiled from sentences that compress (and as such, best describe) the document. The problem of summarization is reduced to the maximal coverage, following the assumption that a summary that best describes the original text, should cover most of describing the document word sequences. We solve it by a greedy algorithm and present the evaluation results.

**From the Virtual to the RealWorld: Referring to Objects in Real-World Spatial Scenes***Dimitra Gkatzia, Verena Rieser, Phil Bartie, and William Mackaness*

Predicting the success of referring expressions (RE) is vital for real-world applications such as navigation systems. Traditionally, research has focused on studying Referring Expression Generation (REG) in virtual, controlled environments. In this paper, we describe a novel study of spatial references from real scenes rather than virtual. First, we investigate how humans describe objects in open, uncontrolled scenarios and compare our findings to those reported in virtual environments. We show that REs in real-world scenarios differ significantly to those in virtual worlds. Second, we propose a novel approach to quantifying image complexity when complete annotations are not present (e.g. due to poor object recognition capabilities), and third, we present a model for success prediction of REs for objects in real scenes. Finally, we discuss implications for Natural Language Generation (NLG) systems and future directions.

**An Unsupervised Bayesian Modelling Approach for Storyline Detection on News Articles***Deyu Zhou, Haiyang Xu, and Yulan He*

Storyline detection from news articles aims at summarizing events described under a certain news topic and revealing how those events evolve over time. It is a difficult task because it requires first the detection of events from news articles published in different time periods and then the construction of storylines by linking events into coherent news stories. Moreover, each storyline has different hierarchical structures which are dependent across epochs. Existing approaches often ignore the dependency of hierarchical structures in storyline generation. In this paper, we propose an unsupervised Bayesian model, called dynamic storyline detection model, to extract structured representations and evolution patterns of storylines. The proposed model is evaluated on a large scale news corpus. Experimental results show that our proposed model outperforms several baseline approaches.

## **Topical Coherence for Graph-based Extractive Summarization**

*Daraksha Parveen, Hans-Martin Ramsler, and Michael Strube*

We present an approach for extractive single-document summarization. Our approach is based on a weighted graphical representation of documents obtained by topic modeling. We optimize importance, coherence and non-redundancy simultaneously using ILP. We compare ROUGE scores of our system with state-of-the-art results on scientific articles from PLOS Medicine and on DUC 2002 data. Human judges evaluate the coherence of summaries generated by our system in comparison to two baselines. Our approach obtains competitive performance.

## **Summarizing Student Responses to Reflection Prompts**

*Wencan Luo and Diane Litman*

We propose to automatically summarize student responses to reflection prompts and introduce a novel summarization algorithm that differs from traditional methods in several ways. First, since the linguistic units of student inputs range from single words to multiple sentences, our summaries are created from extracted phrases rather than from sentences. Second, the phrase summarization algorithm ranks the phrases by the number of students who semantically mention a phrase in a summary. Experimental results show that the proposed phrase summarization approach achieves significantly better summarization performance on an engineering course corpus in terms of ROUGE scores when compared to other summarization methods, including MEAD, LexRank and MMR.

## **Extractive Summarization by Maximizing Semantic Volume**

*Dani Yogatama, Fei Liu, and Noah A. Smith*

The most successful approaches to extractive text summarization seek to maximize bigram coverage subject to a budget constraint. In this work, we propose instead to maximize semantic volume. We embed each sentence in a semantic space and construct a summary by choosing a subset of sentences whose convex hull maximizes volume in that space. We provide a greedy algorithm based on the Gram-Schmidt process to efficiently perform volume maximization. Our method outperforms the state-of-the-art summarization approaches on benchmark datasets.

## **LCSTS: A Large Scale Chinese Short Text Summarization Dataset**

*Baotian Hu, Qingcai Chen, and Fangze Zhu*

Automatic text summarization is widely regarded as the highly difficult problem, partially because of the lack of large text summarization data set. Due to the great challenge of constructing the large scale summaries for full text, in this paper, we introduce a large corpus of Chinese short text summarization dataset constructed from the Chinese microblogging website Sina Weibo, which is released to the public<sup>1</sup>. This corpus consists of over 2 million real Chinese short texts with short summaries given by the author of each text. We also manually tagged the relevance of 10,666 short summaries with their corresponding short texts. Based on the corpus, we introduce recurrent neural network for the summary generation and achieve promising results, which not only shows the usefulness of the proposed corpus for short text summarization research, but also provides a baseline for further research on this topic.

## **Discourse Planning with an N-gram Model of Relations**

*Or Biran and Kathleen McKeown*

While it has been established that transitions between discourse relations are important for coherence, such information has not so far been used to aid in language generation. We introduce an approach to discourse planning for concept-to-text generation systems which simultaneously determines the order of messages and the discourse relations between them. This approach makes it straightforward to use statistical transition models, such as n-gram models of discourse relations learned from an annotated corpus. We show that using such a model significantly improves the quality of the generated text as judged by humans.

## **Experiments with Generative Models for Dependency Tree Linearization**

*Richard Futrell and Edward Gibson*

We present experiments with generative models for linearization of unordered labeled syntactic dependency trees (Belz et al., 2011; Rajkumar and White, 2014). Our linearization models are derived from generative models for dependency structure (Eisner, 1996). We present a series of generative dependency models designed to capture successively more information about ordering constraints among

<sup>1</sup><http://icrc.hitsz.edu.cn/Article/show/139.html>

sister dependents. We give a dynamic programming algorithm for computing the conditional probability of word orders given tree structures under these models. The models are tested on corpora of 11 languages using test-set likelihood, and human ratings for generated forms are collected for English. Our models benefit from representing local order constraints among sisters and from backing off to less sparse distributions, including distributions not conditioned on the head.

**Summarization Based on Embedding Distributions**

*Hayato Kobayashi, Masaki Noguchi, and Taichi Yatsuka*

In this study, we consider a summarization method using the document level similarity based on embeddings, or distributed representations of words, where we assume that an embedding of each word can represent its “meaning.” We formalize our task as the problem of maximizing a submodular function defined by the negative summation of the nearest neighbors’ distances on embedding distributions, each of which represents a set of word embeddings in a document. We proved the submodularity of our objective function and that our problem is asymptotically related to the KL-divergence between the probability density functions that correspond to a document and its summary in a continuous space. An experiment using a real dataset demonstrated that our method performed better than the existing method based on the sentence-level similarity.

**Reversibility reconsidered: finite-state factors for efficient probabilistic sampling in parsing and generation**

*Marc Dymetman, Sriram Venkatapathy, and Chunyang Xiao*

We revisit the classical logical notion of generation/parsing reversibility in terms of feasible probabilistic sampling, and argue for an implementation based on finite-state factors. We propose a modular decomposition that reconciles generation accuracy with parsing robustness and allows the introduction of dynamic contextual factors. (Opinion Piece)

## Session 8 Overview

---

Track A	Track B	Track C	
<i>Fun and Quirky Topics (Short Papers)</i>  Main Auditorium	<i>Semantics (Short Papers)</i>  Small Auditorium	<i>Statistical Models and Machine Learning Methods / Machine Translation (Short Papers)</i>  Room 2	
<i>A quantitative analysis of gender differences in movies using psycholinguistic normatives</i>  <i>Ramakrishna, Malandrakis, Staruk, and Narayanan</i>	<i>Knowledge Base Inference using Bridging Entities</i>  <i>Kotnis, Bansal, and Talukdar</i>	<i>Joint Event Trigger Identification and Event Coreference Resolution with Structured Perceptron</i>  <i>Araki and Mitamura</i>	13:30
<i>EMNLP versus ACL: Analyzing NLP research over time</i>  <i>Gollapalli and Li</i>	<i>Specializing Word Embeddings for Similarity or Relatedness</i>  <i>Kiela, Hill, and Clark</i>	<i>A Joint Dependency Model of Morphological and Syntactic Structure for Statistical Machine Translation</i>  <i>Sennrich and Haddow</i>	13:45
<i>Answering Elementary Science Questions by Constructing Coherent Scenes using Background Knowledge</i>  <i>Li and Clark</i>	<i>Evaluation of Word Vector Representations by Subspace Alignment</i>  <i>Tsvetkov, Faruqui, Ling, Lampe, and Dyer</i>	<i>Variable-Length Word Encodings for Neural Translation Models</i>  <i>Chitnis and DeNero</i>	14:00
<i>WikiQA: A Challenge Dataset for Open-Domain Question Answering</i>  <i>Yang, Yih, and Meek</i>	<i>Higher-order logical inference with compositional semantics</i>  <i>Mineshima, Martínez-Gómez, Miyao, and Bekki</i>	<i>A Binarized Neural Network Joint Model for Machine Translation</i>  <i>Zhang, Utiyama, Sumita, Neubig, and Nakamura</i>	14:15
<i>Personalized Machine Translation: Predicting Translational Preferences</i>  <i>Mirkin and Meunier</i>	<i>Any-language frame-semantic parsing</i>  <i>Johannsen, Martínez Alonso, and Søgaard</i>	<i>Bayesian Optimization of Text Representations</i>  <i>Yogatama, Kong, and Smith</i>	14:30
<i>Talking to the crowd: What do people react to in online discussions?</i>  <i>Jaech, Zayats, Fang, Ostendorf, and Hajishirzi</i>	<i>What's in an Embedding? Analyzing Word Embeddings through Multilingual Evaluation</i>  <i>Köhn</i>	<i>A Comparative Study on Regularization Strategies for Embedding-based Neural Networks</i>  <i>Peng, Mou, Li, Chen, Lu, and Jin</i>	14:45
<i>What Your Username Says About You</i>  <i>Jaech and Ostendorf</i>		<i>Efficient Hyper-parameter Optimization for NLP Applications</i>  <i>Wang, Feng, Zhou, Xiang, and Mahadevan</i>	15:00

**Track D: Long Paper Posters**

Lower Level Foyer

Chair: Tim Baldwin

13:30–15:15

**Poster Cluster 1: NLP for the Web and Social Media, including Computational Social Science (P1-6)**

- Improved Arabic Dialect Classification with Social Media Data  
*F. Huang*
- Exploiting Debate Portals for Semi-Supervised Argumentation Mining in User-Generated Web Discourse  
*I. Habernal and I. Gurevych*
- Confounds and Consequences in Geotagged Twitter Data  
*U. Pavalanathan and J. Eisenstein*
- Modeling Reportable Events as Turning Points in Narrative  
*J. Ouyang and K. McKeown*
- Towards the Extraction of Customer-to-Customer Suggestions from Reviews  
*S. Negi and P. Buitelaar*
- Using Content-level Structures for Summarizing Microblog Repost Trees  
*J. Li, W. Gao, Z. Wei, B. Peng, and K.-F. Wong*

**Poster Cluster 2: Discourse (P7-9)**

- Intra-sentential Zero Anaphora Resolution using Subject Sharing Recognition  
*R. Iida, K. Torisawa, C. Hashimoto, J.-H. Oh, and J. Kloetzer*
- Estimation of Discourse Segmentation Labels from Crowd Data  
*Z. Huang, J. Zhong, and R. J. Passonneau*
- Comparing Word Representations for Implicit Discourse Relation Classification  
*C. Braud and P. Denis*

**Track E: Short Paper Posters**

Lower Level Foyer

Chair: Owen Rambow

13:30–15:15

**Poster Cluster 1: Discourse (P1-9)**

- Better Document-level Sentiment Analysis from RST Discourse Parsing  
*P. Bhartia, Y. Ji, and J. Eisenstein*
- Closing the Gap: Domain Adaptation from Explicit to Implicit Discourse Relations  
*Y. Ji, G. Zhang, and J. Eisenstein*
- Wikification of Concept Mentions within Spoken Dialogues Using Domain Constraints from Wikipedia  
*S. Kim, R. E. Banchs, and H. Li*
- Shallow Convolutional Neural Network for Implicit Discourse Relation Recognition  
*B. Zhang, J. Su, D. Xiong, Y. Lu, H. Duan, and J. Yao*

- On the Role of Discourse Markers for Discriminating Claims and Premises in Argumentative Discourse  
*J. Eckle-Kohler, R. Kluge, and I. Gurevych*
- Fatal or not? Finding errors that lead to dialogue breakdowns in chat-oriented dialogue systems  
*R. Higashinaka, M. Mizukami, K. Funakoshi, M. Araki, H. Tsukahara, and Y. Kobayashi*
- Learning Word Meanings and Grammar for Describing Everyday Activities in Smart Environments  
*M. Attamimi, Y. Ando, T. Nakamura, T. Nagai, D. Mochihashi, I. Kobayashi, and H. Asoh*
- Discourse Element Identification in Student Essays based on Global and Local Cohesion  
*W. Song, R. Fu, L. Liu, and T. Liu*
- Adapting Coreference Resolution for Narrative Processing  
*Q. N. T. Do, S. Bethard, and M.-F. Moens*

**Poster Cluster 2: Phonology, Morphology and Word Segmentation (P10-15)**

- Joint Lemmatization and Morphological Tagging with Lemming  
*T. Müller, R. Cotterell, A. Fraser, and H. Schütze*
- Transducer Disambiguation with Sparse Topological Features  
*G. Iglesias, A. de Gispert, and B. Byrne*
- Arabic Diacritization with Recurrent Neural Networks  
*Y. Belinkov and J. Glass*
- Automatic Diacritics Restoration for Hungarian  
*A. Novák and B. Siklósi*
- Morphological Analysis for Unsegmented Languages using Recurrent Neural Network Language Model  
*H. Morita, D. Kawahara, and S. Kurohashi*
- Can Symbol Grounding Improve Low-Level NLP? Word Segmentation as a Case Study  
*H. Kameko, S. Mori, and Y. Tsuruoka*

## Abstracts: Session 8

### Session 8A: Fun and Quirky Topics (Short Papers)

Main Auditorium

Chair: Yanjun Ma

#### A quantitative analysis of gender differences in movies using psycholinguistic normatives

Anil Ramakrishna, Nikolaos Malandrakis, Elizabeth Staruk, and Shrikanth Narayanan  
13:30–13:45

Direct content analysis reveals important details about movies including those of gender representations and potential biases. We investigate the differences between male and female character depictions in movies, based on patterns of language used. Specifically, we use an automatically generated lexicon of linguistic norms characterizing gender lideness. We use multivariate analysis to investigate gender depictions and correlate them with elements of movie production. The proposed metric differentiates between male and female utterances and exhibits some interesting interactions with movie genres and the screenplay writer gender.

#### EMNLP versus ACL: Analyzing NLP research over time

Sujatha Das Gollapalli and Xiaoli Li  
13:45–14:00

The conferences ACL (Association for Computational Linguistics) and EMNLP (Empirical Methods in Natural Language Processing) rank among the premier venues that track the research developments in Natural Language Processing and Computational Linguistics. In this paper, we present a study on the research papers of approximately two decades from these two NLP conferences. We apply keyphrase extraction and corpus analysis tools to the proceedings from these venues and propose probabilistic and vector-based representations to represent the topics published in a venue for a given year. Next, similarity metrics are studied over pairs of venue representations to capture the progress of the two venues with respect to each other and over time.

#### Answering Elementary Science Questions by Constructing Coherent Scenes using Background Knowledge

Yang Li and Peter Clark  
14:00–14:15

Much of what we understand from text is not explicitly stated. Rather, the reader uses his/her knowledge to fill in gaps and create a coherent, mental picture or “scene” depicting what text appears to convey. The scene constitutes an understanding of the text, and can be used to answer questions that go beyond the text. Our goal is to answer elementary science questions, where this requirement is pervasive; A question will often give a partial description of a scene and ask the student about implicit information. We show that by using a simple “knowledge graph” representation of the question, we can leverage several large-scale linguistic resources to provide missing background knowledge, somewhat alleviating the knowledge bottleneck in previous approaches. The coherence of the best resulting scene, built from a question/answer-candidate pair, reflects the confidence that the answer candidate is correct, and thus can be used to answer multiple choice questions. Our experiments show that this approach significantly outperforms competitive algorithms on several datasets tested. The significance of this work is thus to show that a simple “knowledge graph” representation allows a version of “interpretation as scene construction” to be made viable.

#### WikiQA: A Challenge Dataset for Open-Domain Question Answering

Yi Yang, Wen-tau Yih, and Christopher Meek  
14:15–14:30

We describe the WikiQA, a new publicly available set of question and sentence pairs, collected and annotated for research on open-domain question answering. Most previous work on answer sentence selection focuses on a dataset created using the TREC-QA data, which includes editor-generated questions and candidate answer sentences selected by matching content words in the question. WikiQA constructed using a more natural process and is more than an order of magnitude larger than the previous dataset. In addition, the WikiQA dataset also includes questions for which there are no correct sentences, enabling researchers to work on answer triggering, a critical component in any QA system. We compare several systems on the task of answer sentence selection on both datasets and also describe the performance of a system on the problem of answer triggering using the WikiQA dataset.

**Personalized Machine Translation: Predicting Translational Preferences***Shachar Mirkin and Jean-Luc Meunier*

14:30–14:45

Machine Translation (MT) has advanced in recent years to produce better translations for clients' specific domains, and sophisticated tools allow professional translators to obtain translations according to their prior edits. We suggest that MT should be further personalized to the end-user level – the receiver or the author of the text – as done in other applications. As a step in that direction, we propose a method based on a recommender systems approach where the user's preferred translation is predicted based on preferences of similar users. In our experiments, this method outperforms a set of non-personalized methods, suggesting that user preference information can be employed to provide better-suited translations for each user.

**Talking to the crowd: What do people react to in online discussions?***Aaron Jaech, Victoria Zayats, Hao Fang, Mari Ostendorf, and Hannaneh Hajishirzi* 14:45–15:00

This paper addresses the question of how language use affects community reaction to comments in online discussion forums, and the relative importance of the message vs. the messenger. A new comment ranking task is proposed based on community annotated karma in Reddit discussions, which controls for topic and timing of comments. Experimental work with discussion threads from six subreddits shows that the importance of different types of language features varies with the community of interest.

**What Your Username Says About You***Aaron Jaech and Mari Ostendorf*

15:00–15:15

Usernames are ubiquitous on the Internet, and they are often suggestive of user demographics. This work looks at the degree to which gender and language can be inferred from a username alone by making use of unsupervised morphology induction to decompose usernames into sub-units. Experimental results on the two tasks demonstrate the effectiveness of the proposed morphological features compared to a character n-gram baseline.

**Session 8B: Semantics (Short Papers)**

Small Auditorium

Chair: Marco Baroni

**Knowledge Base Inference using Bridging Entities***Bhushan Kotnis, Pradeep Bansal, and Partha P. Talukdar*

13:30–13:45

Large scale Knowledge Bases (such as NELL, Yago, Freebase, etc) are often sparse, i.e., a large number of valid relations between existing entities are missing. Recent research have addressed this problem by augmenting the KB graph with additional edges mined from a large text corpus while keeping the set of nodes fixed, and then using the Path Ranking Algorithm (PRA) to perform KB inference over this augmented graph. In this paper, we extend this line of work by augmenting the KB graph not only with edges, but also with bridging entities, where both the edges and bridging entities are mined from a 500 million web text corpus. Through experiments on real-world datasets, we demonstrate the value of bridging entities in improving the performance and running times of PRA in the KB inference task.

**Specializing Word Embeddings for Similarity or Relatedness***Douwe Kiela, Felix Hill, and Stephen Clark*

13:45–14:00

We demonstrate the advantage of specializing semantic word embeddings for either similarity or relatedness. We compare two variants of retrofitting and a joint-learning approach, and find that all three yield specialized semantic spaces that capture human intuitions regarding similarity and relatedness better than unspecialized spaces. We also show that using specialized spaces in NLP tasks and applications leads to clear improvements, for document classification and synonym selection, which rely on either similarity or relatedness but not both.

**Evaluation of Word Vector Representations by Subspace Alignment***Yulia Tsvetkov, Manaal Faruqui, Wang Ling, Guillaume Lample, and Chris Dyer* 14:00–14:15

Unsupervisedly learned word vectors have proven to provide exceptionally effective features in many NLP tasks. Most common intrinsic evaluations of vector quality measure correlation with similarity judgments. However, these often correlate poorly with how well the learned representations perform as features in downstream evaluation tasks. We present QVEC—a computationally inexpensive intrinsic evaluation measure of the quality of word embeddings based on alignment to a matrix of features extracted from manually crafted lexical resources—that obtains strong correlation with performance of the vectors in a battery of downstream semantic evaluation tasks.

**Higher-order logical inference with compositional semantics***Koji Mineshima, Pascual Martínez-Gómez, Yusuke Miyao, and Daisuke Bekki* 14:15–14:30

We present a higher-order inference system based on a formal compositional semantics and the wide-coverage CCG parser. We develop an improved method to bridge between the parser and semantic composition. The system is evaluated on the FraCaS test suite. In contrast to the widely held view that higher-order logic is unsuitable for efficient logical inferences, the results show that a system based on a reasonably-sized semantic lexicon and a manageable number of non-first-order axioms enables efficient logical inferences, including those concerned with generalized quantifiers and intensional operators, and outperforms the state-of-the-art first-order inference system.

**Any-language frame-semantic parsing***Anders Johannsen, Héctor Martínez Alonso, and Anders Søgaard*

14:30–14:45

We present a multilingual corpus of Wikipedia and Twitter texts annotated with FRAMENET 1.5 semantic frames in nine different languages, as well as a novel technique for weakly supervised crosslingual frame semantic parsing. Our approach only assumes the existence of linked, comparable source and target language corpora (e.g., Wikipedia) and a bilingual dictionary (e.g., Wiktionary or BABELNET). Our approach uses a truly interlingual representation enabling us to use the same model across all nine languages. We present average error reductions over running a state-of-the-art parser on word-to-word translations of 46% for target identification, 37% for frame identification, and 14% for argument identification.

**What's in an Embedding? Analyzing Word Embeddings through Multilingual Evaluation***Arne Köhn*

15:00–15:15

In the last two years, there has been a surge of word embedding algorithms and research on them. However, evaluation has mostly been carried out on a narrow set of tasks, mainly word similarity/relatedness and word relation similarity and on a single language, namely English. We propose an approach to evaluate embeddings on a variety of languages that also yields insights into the structure of the embedding space by investigating how well word embeddings cluster along different syntactic features. We show that all embedding approaches behave similarly in this task, with dependency-based embeddings performing best. This effect is even more pronounced when generating low dimensional embeddings.

## Session 8C: Statistical Models and Machine Learning Methods / Machine Translation (Short Papers)

Room 2

Chair: Kevin Gimpel

### Joint Event Trigger Identification and Event Coreference Resolution with Structured Perceptron

Jun Araki and Teruko Mitamura

13:30–13:45

Events and their coreference offer useful semantic and discourse resources. We show that the semantic and discourse aspects of events interact with each other. However, traditional approaches addressed event extraction and event coreference resolution either separately or sequentially, which limits their interactions. This paper proposes a document-level structured learning model that simultaneously identifies event triggers and resolves event coreference. We demonstrate that the joint model outperforms a pipelined model by 6.9 BLANC F1 and 1.8 CoNLL F1 points in event coreference resolution using a corpus in the biology domain.

### A Joint Dependency Model of Morphological and Syntactic Structure for Statistical Machine Translation

Rico Sennrich and Barry Haddow

13:45–14:00

When translating between two languages that differ in their degree of morphological synthesis, syntactic structures in one language may be realized as morphological structures in the other, and SMT models need a mechanism to learn such translations. Prior work has used morpheme splitting with flat representations that do not encode the hierarchical structure between morphemes, but this structure is relevant for learning morphosyntactic constraints and selectional preferences. We propose to model syntactic and morphological structure jointly in a dependency translation model, allowing the system to generalize to the level of morphemes. We present a dependency representation of German compounds and particle verbs that results in improvements in translation quality of 1.4–1.8 BLEU in the WMT English-German translation task.

### Variable-Length Word Encodings for Neural Translation Models

Rohan Chitnis and John DeNero

14:00–14:15

Recent work in neural machine translation has shown promising performance, but the most effective architectures do not scale naturally to large vocabulary sizes. We propose and compare three variable-length encoding schemes that represent a large vocabulary corpus using a much smaller vocabulary with no loss in information. Common words are unaffected by our encoding, but rare words are encoded using a sequence of two pseudo-words. Our method is simple and effective: it requires no complete dictionaries, learning procedures, increased training time, changes to the model, or new parameters. Compared to a baseline that replaces all rare words with an unknown word symbol, our best variable-length encoding strategy improves WMT English-French translation performance by up to 1.7 BLEU.

### A Binarized Neural Network Joint Model for Machine Translation

Jingyi Zhang, Masao Utiyama, Eiichiro Sumita, Graham Neubig, and Satoshi Nakamura

14:15–14:30

The neural network joint model (NNJM), which augments the neural network Language model (NNLM) with an m-word source context window, has achieved large gains in machine translation accuracy, but also has problems with high normalization cost when using large vocabularies. Training the NNJM with noise-contrastive estimation (NCE), instead of standard maximum likelihood estimation (MLE), can reduce computation cost. In this paper, we propose an alternative to NCE, the binarized NNJM (BNNJM), which learns a binary classifier that takes both the context and target words as input, and can be efficiently trained using MLE. We compare the BNNJM and NNJM trained by NCE on Chinese-to-English and Japanese-to-English translation tasks.

### Bayesian Optimization of Text Representations

Dani Yogatama, Lingpeng Kong, and Noah A. Smith

14:30–14:45

When applying machine learning to problems in NLP, there are many choices to make about how to represent input texts. They can have a big effect on performance, but they are often uninteresting to researchers or practitioners who simply need a module that performs well. We apply sequential model-based optimization over this space of choices and show that it makes standard linear models competitive

with more sophisticated, expensive state-of-the-art methods based on latent variables or neural networks on various topic classification and sentiment analysis problems. Our approach is a first step towards black-box NLP systems that work with raw text and do not require manual tuning.

### **A Comparative Study on Regularization Strategies for Embedding-based Neural Networks**

*Hao Peng, Lili Mou, Ge Li, Yunchuan Chen, Yangyang Lu, and Zhi Jin*

14:45–15:00

This paper aims to compare different regularization strategies to address a common phenomenon, severe overfitting, in embedding-based neural networks for NLP. We chose two widely studied neural models and tasks as our testbed. We tried several frequently applied or newly proposed regularization strategies, including penalizing weights (embeddings excluded), penalizing embeddings, re-embedding words, and dropout. We also emphasized on incremental hyperparameter tuning, and combining different regularizations. The results in this work provide a picture on tuning hyperparameters for neural NLP.

### **Efficient Hyper-parameter Optimization for NLP Applications**

*Lidan Wang, Minwei Feng, Bowen Zhou, Bing Xiang, and Sridhar Mahadevan* 15:00–15:15

Hyper-parameter optimization is an important problem in natural language processing (NLP) and machine learning. Recently, a group of studies has focused on using sequential Bayesian Optimization to solve this problem, which aims to reduce the number of iterations and trials required during the optimization process. In this paper, we explore this problem from a different angle, and propose a multi-stage hyper-parameter optimization that breaks the problem into multiple stages with increasingly amounts of data. Early stage provides fast estimates of good candidates which are used to initialize later stages for better performance and speed. We demonstrate the utility of this new algorithm by evaluating its speed and accuracy against state-of-the-art Bayesian Optimization algorithms on classification and prediction tasks.

**Session 8D: Long Paper Posters**

Lower Level Foyer

Chair: Tim Baldwin

13:30–15:15

**Improved Arabic Dialect Classification with Social Media Data***Fei Huang*

Arabic dialect classification has been an important and challenging problem for Arabic language processing, especially for social media text analysis and machine translation. In this paper we propose an approach to improving Arabic dialect classification with semi-supervised learning: multiple classifiers are trained with weakly supervised, strongly supervised, and unsupervised data. Their combination yields significant and consistent improvement on two different test sets. The dialect classification accuracy is improved by 5% over the strongly supervised classifier and 20% over the weakly supervised classifier. Furthermore, when applying the improved dialect classifier to build a Modern Standard Arabic (MSA) language model (LM), the new model size is reduced by 70% while the English-Arabic translation quality is improved by 0.6 BLEU point.

**Exploiting Debate Portals for Semi-Supervised Argumentation Mining in User-Generated Web Discourse***Ivan Habernal and Iryna Gurevych*

Analyzing arguments in user-generated Web discourse has recently gained attention in argumentation mining, an evolving field of NLP. Current approaches, which employ fully-supervised machine learning, are usually domain dependent and suffer from the lack of large and diverse annotated corpora. However, annotating arguments in discourse is costly, error-prone, and highly context-dependent. We asked whether leveraging unlabeled data in a semi-supervised manner can boost the performance of argument component identification and to which extent is the approach independent of domain and register. We propose novel features that exploit clustering of unlabeled data from debate portals based on a word embeddings representation. Using these features, we significantly outperform several baselines in the cross-validation, cross-domain, and cross-register evaluation scenarios.

**Confounds and Consequences in Geotagged Twitter Data***Umashanthi Pavalanathan and Jacob Eisenstein*

Twitter is often used in quantitative studies that identify geographically-preferred topics, writing styles, and entities. These studies rely on either GPS coordinates attached to individual messages, or on the user-supplied location field in each profile. In this paper, we compare these data acquisition techniques and quantify the biases that they introduce; we also measure their effects on linguistic analysis and text-based geolocation. GPS-tagging and self-reported locations yield measurably different corpora, and these linguistic differences are partially attributable to differences in dataset composition by age and gender. Using a latent variable model to induce age and gender, we show how these demographic variables interact with geography to affect language use. We also show that the accuracy of text-based geolocation varies with population demographics, giving the best results for men above the age of 40.

**Modeling Reportable Events as Turning Points in Narrative***Jessica Ouyang and Kathleen McKeown*

We present novel experiments in modeling the rise and fall of story characteristics within narrative, leading up to the Most Reportable Event (MRE), the compelling event that is the nucleus of the story. We construct a corpus of personal narratives from the bulletin board website Reddit, using the organization of Reddit content into topic-specific communities to automatically identify narratives. Leveraging the structure of Reddit comment threads, we automatically label a large dataset of narratives. We present a change-based model of narrative that tracks changes in formality, affect, and other characteristics over the course of a story, and we use this model in distant supervision and self-training experiments that achieve significant improvements over the baselines at the task of identifying MREs.

**Towards the Extraction of Customer-to-Customer Suggestions from Reviews***Sapna Negi and Paul Buitelaar*

State of the art in opinion mining mainly focuses on positive and negative sentiment summarisation of online customer reviews. We observe that reviewers tend to provide advice, recommendations and tips to the fellow customers on a variety of points of interest. In this work, we target the automatic detection of suggestion expressing sentences in customer reviews. This is a novel problem, and therefore to

begin with, requires a well formed problem definition and benchmark dataset. This work provides a 3-fold contribution, namely, problem definition, benchmark dataset, and an approach for detection of suggestions for the customers. The problem is framed as a sentence classification problem, and a set of linguistically motivated features are proposed. Analysis of the nature of suggestions, and classification errors, highlight challenges and research opportunities associated with this problem.

### **Using Content-level Structures for Summarizing Microblog Repost Trees**

*Jing Li, Wei Gao, Zhongyu Wei, Baolin Peng, and Kam-Fai Wong*

A microblog repost tree provides strong clues on how an event described therein develops. To help social media users capture the main clues of events on microblogging sites, we propose a novel repost tree summarization framework by effectively differentiating two kinds of messages on repost trees called leaders and followers, which are derived from content-level structure information, i.e., contents of messages and the reposting relations. To this end, Conditional Random Fields (CRF) model is used to detect leaders across repost tree paths. We then present a variant of random-walk-based summarization model to rank and select salient messages based on the result of leader detection. To reduce the error propagation cascaded from leader detection, we improve the framework by enhancing the random walk with adjustment steps for sampling from leader probabilities given all the reposting messages. For evaluation, we construct two annotated corpora, one for leader detection, and the other for repost tree summarization. Experimental results confirm the effectiveness of our method.

### **Intra-sentential Zero Anaphora Resolution using Subject Sharing Recognition**

*Ryu Iida, Kentaro Torisawa, Chikara Hashimoto, Jong-Hoon Oh, and Julien Kloetzer*

In this work, we improve the performance of intra-sentential zero anaphora resolution in Japanese using a novel method of recognizing subject sharing relations. In Japanese, a large portion of intra-sentential zero anaphora can be regarded as subject sharing relations between predicates, that is, the subject of some predicate is also the unrealized subject of other predicates. We develop an accurate recognizer of subject sharing relations for pairs of predicates in a single sentence, and then construct a subject shared predicate network, which is a set of predicates that are linked by the subject sharing relations recognized by our recognizer. We finally combine our zero anaphora resolution method exploiting the subject shared predicate network and a state-of-the-art ILP-based zero anaphora resolution method. Our combined method achieved a significant improvement over the the ILP-based method alone on intra-sentential zero anaphora resolution in Japanese. To the best of our knowledge, this is the first work to explicitly use an independent subject sharing recognizer in zero anaphora resolution.

### **Estimation of Discourse Segmentation Labels from Crowd Data**

*Ziheng Huang, Jialu Zhong, and Rebecca J. Passonneau*

For annotation tasks involving independent judgments, probabilistic models have been used to infer ground truth labels from data where a crowd of many annotators labels the same items. Such models have been shown to produce results superior to taking the majority vote, but have not been applied to sequential data. We present two methods to infer ground truth labels from sequential annotations where we assume judgments are not independent, based on the observation that an annotator's segments all tend to be several utterances long. The data consists of crowd labels for annotation of discourse segment boundaries. The new methods extend Hidden Markov Models to relax the independence assumption. The two methods are distinct, so positive labels proposed by both are taken to be ground truth. In addition, results of the models are checked using metrics that test whether an annotator's accuracy relative to a given model remains consistent across different conversations.

### **Comparing Word Representations for Implicit Discourse Relation Classification**

*Chloé Braud and Pascal Denis*

This paper presents a detailed comparative framework for assessing the usefulness of unsupervised word representations for identifying so-called implicit discourse relations. Specifically, we compare standard one-hot word pair representations against low-dimensional ones based on Brown clusters and word embeddings. We also consider various word vector combination schemes for deriving discourse segment representations from word vectors, and compare representations based either on all words or limited to head words. Our main finding is that denser representations systematically outperform sparser ones and give state-of-the-art performance or above without the need for additional hand-crafted features.

**Session 8E: Short Paper Posters**

Lower Level Foyer

Chair: Owen Rambow

13:30–15:15

**Better Document-level Sentiment Analysis from RST Discourse Parsing***Parminder Bhatia, Yangfeng Ji, and Jacob Eisenstein*

Discourse structure is the hidden link between surface features and document-level properties, such as sentiment polarity. We show that the discourse analyses produced by Rhetorical Structure Theory (RST) parsers can improve document-level sentiment analysis, via composition of local information up the discourse tree. First, we show that reweighting discourse units according to their position in a dependency representation of the rhetorical structure can yield substantial improvements on lexicon-based sentiment analysis. Next, we present a recursive neural network over the RST structure, which offers significant improvements over classification-based methods.

**Closing the Gap: Domain Adaptation from Explicit to Implicit Discourse Relations***Yangfeng Ji, Gongbo Zhang, and Jacob Eisenstein*

Many discourse relations are explicitly marked with discourse connectives, and these examples could potentially serve as a plentiful source of training data for recognizing implicit discourse relations. However, there are important linguistic differences between explicit and implicit discourse relations, which limit the accuracy of such an approach. We account for these differences by applying techniques from domain adaptation, treating implicitly and explicitly-marked discourse relations as separate domains. The distribution of surface features varies across these two domains, so we apply a marginalized denoising autoencoder to induce a dense, domain-general representation. The label distribution is also domain-specific, so we apply a resampling technique that is similar to instance weighting. In combination with a set of automatically-labeled data, these improvements eliminate more than 80% of the transfer loss incurred by training an implicit discourse relation classifier on explicitly-marked discourse relations.

**Wikification of Concept Mentions within Spoken Dialogues Using Domain Constraints from Wikipedia***Seokhwan Kim, Rafael E. Banchs, and Haizhou Li*

While most previous work on Wikification has focused on written texts, this paper presents a Wikification approach for spoken dialogues. A set of analyzers are proposed to learn dialogue-specific properties along with domain knowledge of conversations from Wikipedia. Then, the analyzed properties are used as constraints for generating candidates, and the candidates are ranked to find the appropriate links. The experimental results show that our proposed approach can significantly improve the performances of the task in human-human dialogues.

**Shallow Convolutional Neural Network for Implicit Discourse Relation Recognition***Biao Zhang, Jinsong Su, Deyi Xiong, Yaojie Lu, Hong Duan, and Junfeng Yao*

Implicit discourse relation recognition remains a serious challenge due to the absence of discourse connectives. In this paper, we propose a Shallow Convolutional Neural Network (SCNN) for implicit discourse relation recognition, which contains only one hidden layer but is effective in relation recognition. The shallow structure alleviates the overfitting problem, while the convolution and nonlinear operations help preserve the recognition and generalization ability of our model. Experiments on the benchmark data set show that our model achieves comparable and even better performance when comparing against current state-of-the-art systems.

**On the Role of Discourse Markers for Discriminating Claims and Premises in Argumentative Discourse***Judith Eckle-Kohler, Roland Kluge, and Iryna Gurevych*

This paper presents a study on the role of discourse markers in argumentative discourse. We annotated a German corpus with arguments according to the common claim-premise model of argumentation and performed various statistical analyses regarding the discriminative nature of discourse markers for claims and premises. Our experiments show that particular semantic groups of discourse markers are indicative of either claims or premises and constitute highly predictive features for discriminating between them.

## **Fatal or not? Finding errors that lead to dialogue breakdowns in chat-oriented dialogue systems**

*Ryuichiro Higashinaka, Masahiro Mizukami, Kotaro Funakoshi, Masahiro Araki, Hiroshi Tsukahara, and Yuka Kobayashi*

This paper aims to find errors that lead to dialogue breakdowns in chat-oriented dialogue systems. We collected chat dialogue data, annotated them with dialogue breakdown labels, and collected comments describing the error that led to the breakdown. By mining the comments, we first identified error types. Then, we calculated the correlation between an error type and the degree of dialogue breakdown it incurred, quantifying its impact on dialogue breakdown. This is the first study to quantitatively analyze error types and their effect in chat-oriented dialogue systems.

## **Learning Word Meanings and Grammar for Describing Everyday Activities in Smart Environments**

*Muhammad Attamimi, Yuji Ando, Tomoaki Nakamura, Takayuki Nagai, Daichi Mochihashi, Ichiro Kobayashi, and Hideki Asoh*

If intelligent systems are to interact with humans in a natural manner, the ability to describe daily life activities is important. To achieve this, sensing human activities by capturing multimodal information is necessary. In this study, we consider a smart environment for sensing activities with respect to realistic scenarios. We next propose a sentence generation system from observed multimodal information in a bottom up manner using multilayered multimodal latent Dirichlet allocation and Bayesian hidden Markov models. We evaluate the grammar learning and sentence generation as a complete process within a realistic setting. The experimental result reveals the effectiveness of the proposed method.

## **Discourse Element Identification in Student Essays based on Global and Local Cohesion**

*Wei Song, Ruiji Fu, Lizhen Liu, and Ting Liu*

We present a method of using cohesion to improve discourse element identification for sentences in student essays. New features for each sentence are derived by considering its relations to global and local cohesion, which are created by means of cohesive resources and subtopic coverage. In our experiments, we obtain significant improvements on identifying all discourse elements, especially of +5% F1 score on thesis and main idea. The analysis shows that global cohesion can better capture thesis statements.

## **Adapting Coreference Resolution for Narrative Processing**

*Quynh Ngoc Thi Do, Steven Bethard, and Marie-Francine Moens*

Domain adaptation is a challenge for supervised NLP systems because of expensive and time-consuming manual annotation resources. We present a novel method to adapt a supervised coreference resolution system trained on newswire domain to short narrative stories without retraining the system. The idea is to perform inference via an Integer Linear Programming (ILP) formulation with the features of narratives adopted soft constraints. When testing on the UMIREC and N2 corpora with the-state-of-the-art Berkeley coreference resolution system trained on OntoNotes, our inference substantially outperforms the original inference on the CoNLL 2011 metric.

## **Joint Lemmatization and Morphological Tagging with Lemming**

*Thomas Müller, Ryan Cotterell, Alexander Fraser, and Hinrich Schütze*

We present Lemming, a modular log-linear model that jointly models lemmatization and tagging and supports the integration of arbitrary global features. It is trainable on corpora annotated with gold standard tags and lemmata and does not rely on morphological dictionaries or analyzers. Lemming sets the new state of the art in token-based statistical lemmatization on six languages; e.g., for Czech lemmatization, we reduce the error by 60%, from 4.05 to 1.58. We also give empirical evidence that jointly modeling morphological tags and lemmata is mutually beneficial.

## **Transducer Disambiguation with Sparse Topological Features**

*Gonzalo Iglesias, Adrià de Gispert, and Bill Byrne*

We describe a simple and efficient algorithm to disambiguate non-functional weighted finite state transducers (WFSTs), i.e. to generate a new WFST that contains a unique, best-scoring path for each hypothesis in the input labels along with the best output labels. The algorithm uses topological features combined with a tropical sparse tuple vector semiring. We empirically show that our algorithm is more

efficient than previous work in a PoS-tagging disambiguation task. We use our method to rescore very large translation lattices with a bilingual neural network language model, obtaining gains in line with the literature.

### **Arabic Diacritization with Recurrent Neural Networks**

*Yonatan Belinkov and James Glass*

Arabic, Hebrew, and similar languages are typically written without diacritics, leading to ambiguity and posing a major challenge for core language processing tasks like speech recognition. Previous approaches to automatic diacritization employed a variety of machine learning techniques. However, they typically rely on existing tools like morphological analyzers and therefore cannot be easily extended to new genres and languages. We develop a recurrent neural network with long short-term memory layers for predicting diacritics in Arabic text. Our language-independent approach is trained solely from diacritized text without relying on external tools. We show experimentally that our model can rival state-of-the-art methods that have access to additional resources.

### **Automatic Diacritics Restoration for Hungarian**

*Attila Novák and Borbála Siklósi*

In this paper, we describe a method based on statistical machine translation (SMT) that is able to restore accents in Hungarian texts with high accuracy. Due to the agglutinating characteristic of Hungarian, there are always wordforms unknown to a system trained on a fixed vocabulary. In order to be able to handle such words, we integrated a morphological analyzer into the system that can suggest accented word candidates for unknown words. We evaluated the system in different setups, achieving an accuracy above 99% at the highest.

### **Morphological Analysis for Unsegmented Languages using Recurrent Neural Network Language Model**

*Hajime Morita, Daisuke Kawahara, and Sadao Kurohashi*

We present a new morphological analysis model that considers semantic plausibility of word sequences by using a recurrent neural network language model (RNNLM). In unsegmented languages, since language models are learned from automatically segmented texts and inevitably contain errors, it is not apparent that conventional language models contribute to morphological analysis. To solve this problem, we do not use language models based on raw word sequences but use a semantically generalized language model, RNNLM, in morphological analysis. In our experiments on two Japanese corpora, our proposed model significantly outperformed baseline models. This result indicates the effectiveness of RNNLM in morphological analysis.

### **Can Symbol Grounding Improve Low-Level NLP? Word Segmentation as a Case Study**

*Hirotaka Kameko, Shinsuke Mori, and Yoshimasa Tsuruoka*

We propose a novel framework for improving a word segmenter using information acquired from symbol grounding. We generate a term dictionary in three steps: generating a pseudo-stochastically segmented corpus, building a symbol grounding model to enumerate word candidates, and filtering them according to the grounding scores. We applied our method to game records of Japanese chess with commentaries. The experimental results show that the accuracy of a word segmenter can be improved by incorporating the generated dictionary.

## Session 9 Overview

---

Track A	Track B	Track C	
Statistical Models and Machine Learning Methods (Long + TACL Papers) Main Auditorium	Text Mining and NLP Applications (Long Papers) Small Auditorium	Spoken Language Processing and Language Modeling (Long Papers) Room 2	
When Are Tree Structures Necessary for Deep Learning of Representations? <i>Li, Luong, Jurafsky, and Hovy</i>	Verbal and Nonverbal Clues for Real-life Deception Detection <i>Pérez-Rosas, Abouelenien, Mihalcea, Xiao, Linton, and Burzo</i>	Topic Identification and Discovery on Text and Speech <i>May, Ferraro, McCree, Wintrode, Garcia-Romero, and Van Durme</i>	15:40
Discriminative Neural Sentence Modeling by Tree-Based Convolution <i>Mou, Peng, Li, Xu, Zhang, and Jin</i>	Social Media Text Classification under Negative Covariate Shift <i>Fei and Liu</i>	A Dynamic Programming Algorithm for Computing N-gram Posteriors from Lattices <i>Can and Narayanan</i>	16:05
Multi-Timescale Long Short-Term Memory Neural Network for Modelling Sentences and Documents <i>Liu, Qiu, Chen, Wu, and Huang</i>	Co-Training for Topic Classification of Scholarly Data <i>Caragea, Bulgarov, and Mihalcea</i>	Bilingual Structured Language Models for Statistical Machine Translation <i>Garmash and Monz</i>	16:30
[TACL] Learning Structural Kernels for Natural Language Processing <i>Beck, Cohn, Hardmeier, and Specia</i>	Humor Recognition and Humor Anchor Extraction <i>Yang, Lavie, Dyer, and Hovy</i>	Compact, Efficient and Unlimited Capacity: Language Modeling with Compressed Suffix Trees <i>Shareghi, Petri, Haffari, and Cohn</i>	16:55

**Track D: Long Paper Posters**

Chair: Nathan Schneider

Lower Level Foyer

15:40-17:20

**Poster Cluster 1: Semantics (P1-8)**

- ERSOM: A Structural Ontology Matching Approach Using Automatically Learned Entity Representation  
*C. Xiang, T. Jiang, B. Chang, and Z. Sui*
- A Single Word is not Enough: Ranking Multiword Expressions Using Distributional Semantics  
*M. Riedl and C. Biemann*
- Syntactic Dependencies and Distributed Word Representations for Analogy Detection and Mining  
*L. Qiu, Y. Zhang, and Y. Lu*

- Navigating the Semantic Horizon using Relative Neighborhood Graphs  
*A. Cuba Gyllensten and M. Sahlgren*
- Multi- and Cross-Modal Semantics Beyond Vision: Grounding in Auditory Perception  
*D. Kiela and S. Clark*
- Automatic recognition of habituals: a three-way classification of clausal aspect  
*A. Friedrich and M. Pinkal*
- Distributed Representations for Unsupervised Semantic Role Labeling  
*K. Woodsend and M. Lapata*
- A Tableau Prover for Natural Logic and Language  
*L. Abzianidze*

**Track E: Short Paper Posters**

Chair: Kang Liu

Lower Level Foyer

15:40-17:20

**Poster Cluster 1: Sentiment Analysis and Opinion Mining (P1-9)**

- JEAM: A Novel Model for Cross-Domain Sentiment Classification Based on Emotion Analysis  
*K.-H. Luo, Z.-H. Deng, H. Yu, and L.-C. Wei*
- PhraseRNN: Phrase Recursive Neural Network for Aspect-based Sentiment Analysis  
*T. H. Nguyen and K. Shirai*
- ASTD: Arabic Sentiment Tweets Dataset  
*M. Nabil, M. Aly, and A. Atiya*
- Adjective Intensity and Sentiment Analysis  
*R. Sharma, M. Gupta, A. Agarwal, and P. Bhattacharyya*
- The Rating Game: Sentiment Rating Reproducibility from Text  
*L. Borgholt, P. Simonsen, and D. Hovy*
- A Multi-lingual Annotated Dataset for Aspect-Oriented Opinion Mining  
*S. M. Jiménez-Zafra, G. Berardi, A. Esuli, D. Marcheggiani, M. T. Martín-Valdivia, and A. Moreo Fernández*
- Deep Convolutional Neural Network Textual Features and Multiple Kernel Learning for Utterance-level Multimodal Sentiment Analysis  
*S. Poria, E. Cambria, and A. Gelbukh*
- SLSA: A Sentiment Lexicon for Standard Arabic  
*R. Eskander and O. Rambow*
- Reinforcing the Topic of Embeddings with Theta Pure Dependence for Text Classification  
*N. Xing, Y. Hou, P. Zhang, W. Li, and D. Song*

**Poster Cluster 2: NLP for the Web and Social Media, including Computational Social Science (P10-17)**

- That's So Annoying!!!: A Lexical and Frame-Semantic Embedding Based Data Augmentation Approach to Automatic Categorization of Annoying Behaviors using #petpeeve Tweets  
*W. Y. Wang and D. Yang*
- Detection of Steganographic Techniques on Twitter  
*A. Wilson, P. Blunsom, and A. Ker*
- #SupportTheCause: Identifying Motivations to Participate in Online Health Campaigns  
*D. Nguyen, T. van den Broek, C. Hauff, D. Hiemstra, and M. Ehrenhard*
- An Analysis of Domestic Abuse Discourse on Reddit  
*N. Schrading, C. Ovesdotter Alm, R. Ptucha, and C. Homan*
- Twitter-scale New Event Detection via K-term Hashing  
*D. Wurzer, V. Lavrenko, and M. Osborne*
- Classifying Tweet Level Judgements of Rumours in Social Media  
*M. Lukasik, T. Cohn, and K. Bontcheva*
- Identification and Verification of Simple Claims about Statistical Properties  
*A. Vlachos and S. Riedel*

## Abstracts: Session 9

---

### Session 9A: Statistical Models and Machine Learning Methods (Long + TACL Papers)

Main Auditorium

Chair: Jason Eisner

#### **When Are Tree Structures Necessary for Deep Learning of Representations?**

*Jiwei Li, Thang Luong, Dan Jurafsky, and Eduard Hovy*

15:40–16:05

Recursive neural models, which use syntactic parse trees to recursively generate representations bottom-up, are a popular architecture. However there have not been rigorous evaluations showing for exactly which tasks this syntax-based method is appropriate. In this paper, we benchmark recursive neural models against sequential recurrent neural models, enforcing apples-to-apples comparison as much as possible. We investigate 4 tasks: (1) sentiment classification at the sentence level and phrase level; (2) matching questions to answer-phrases; (3) discourse parsing; (4) semantic relation extraction. Our goal is to understand better when, and why, recursive models can outperform simpler models. We find that recursive models help mainly on tasks (like semantic relation extraction) that require long-distance connection modeling, particularly on very long sequences. We then introduce a method for allowing recurrent models to achieve similar performance: breaking long sentences into clause-like units at punctuation and processing them separately before combining. Our results thus help understand the limitations of both classes of models, and suggest directions for improving recurrent models.

#### **Discriminative Neural Sentence Modeling by Tree-Based Convolution**

*Lili Mou, Hao Peng, Ge Li, Yan Xu, Lu Zhang, and Zhi Jin*

16:05–16:30

This paper proposes a tree-based convolutional neural network (TBCNN) for discriminative sentence modeling. Our model leverages either constituency trees or dependency trees. The tree-based convolution process extracts sentences' structural features, which are then aggregated by max pooling. Such architecture allows short propagation paths between the output layer and underlying feature detectors, enabling effective structural feature learning and extraction. We evaluate our models on two tasks: sentiment analysis and question classification. In both experiments, TBCNN outperforms previous state-of-the-art results, including existing neural networks and dedicated feature/rule engineering. We also make efforts to visualize the tree-based convolution process, shedding some light on how our models work.

#### **Multi-Timescale Long Short-Term Memory Neural Network for Modelling Sentences and Documents**

*Pengfei Liu, Xipeng Qiu, Xinchi Chen, Shiyu Wu, and Xuanjing Huang*

16:30–16:55

Neural network based methods have obtained great progress on a variety of natural language processing tasks. However, it is still a challenge task to model long texts, such as sentences and documents. In this paper, we propose a multi-timescale long short-term memory (MT-LSTM) neural network to model long texts. MT-LSTM partitions the hidden states of the standard LSTM into several groups. Each group is activated at different time period. Thus, MT-LSTM can model very long documents as well as short sentences. Experiments on four benchmark datasets show that our model outperforms the other neural models in text classification task.

#### **[TACL] Learning Structural Kernels for Natural Language Processing**

*Daniel Beck, Trevor Cohn, Christian Hardmeier, and Lucia Specia*

16:55–17:20

Structural kernels are a flexible learning paradigm that has been widely used in Natural Language Processing. However, the problem of model selection in kernel-based methods is usually overlooked. Previous approaches mostly rely on setting default values for kernel hyperparameters or using grid search, which is slow and coarse-grained. In contrast, Bayesian methods allow efficient model selection by maximizing the evidence on the training data through gradient-based methods. In this paper we show how to perform this in the context of structural kernels by using Gaussian Processes. Experimental results on tree kernels show that this procedure results in better prediction performance compared to hyperparameter optimization via grid search. The framework proposed in this paper can be adapted to other structures besides trees, e.g., strings and graphs, thereby extending the utility of kernel-based methods.

**Session 9B: Text Mining and NLP Applications (Long Papers)**

Small Auditorium

Chair: Shuming Shi

**Verbal and Nonverbal Clues for Real-life Deception Detection**

Verónica Pérez-Rosas, Mohamed Abouelenien, Rada Mihalcea, Yao Xiao, CJ Linton, and Mihai Burzo 15:40–16:05

Deception detection has been receiving an increasing amount of attention from the computational linguistics, speech, and multimodal processing communities. One of the major challenges encountered in this task is the availability of data, and most of the research work to date has been conducted on acted or artificially collected data. The generated deception models are thus lacking real-world evidence. In this paper, we explore the use of multimodal real-life data for the task of deception detection. We develop a new deception dataset consisting of videos from real-life scenarios, and build deception tools relying on verbal and nonverbal features. We achieve classification accuracies in the range of 77–82% when using a model that extracts and fuses features from the linguistic and visual modalities. We show that these results outperform the human capability of identifying deceit.

**Social Media Text Classification under Negative Covariate Shift**

Geli Fei and Bing Liu

16:05–16:30

In a typical social media content analysis task, the user is interested in analyzing posts of a particular topic. Identifying such posts is often formulated as a classification problem. However, this problem is challenging. One key issue is covariate shift. That is, the training data is not fully representative of the test data. We observed that the covariate shift mainly occurs in the negative data because topics discussed in social media are highly diverse and numerous, but the user-labeled negative training data may cover only a small number of topics. This paper proposes a novel technique to solve the problem. The key novelty of the technique is the transformation of document representation from the traditional n-gram feature space to a center-based similarity (CBS) space. In the CBS space, the covariate shift problem is significantly mitigated, which enables us to build much better classifiers. Experiment results show that the proposed approach markedly improves classification.

**Co-Training for Topic Classification of Scholarly Data**

Cornelia Caragea, Florin Bulgarov, and Rada Mihalcea

16:30–16:55

With the exponential growth of scholarly data during the past few years, effective methods for topic classification are greatly needed. Current approaches usually require large amounts of expensive labeled data in order to make accurate predictions. In this paper, we posit that, in addition to a research article's textual content, its citation network also contains valuable information. We describe a co-training approach that uses the text and citation information of a research article as two different views to predict the topic of the article. We show that this method improves significantly over the individual classifiers, while also bringing a substantial reduction in the amount of labeled data required for training accurate classifiers.

**Humor Recognition and Humor Anchor Extraction**

Diyi Yang, Alon Lavie, Chris Dyer, and Eduard Hovy

16:55–17:20

Humor is an essential component in personal communication. How to create computational models to discover the structure behind humor, recognize humor and even extract humor anchors remains a challenge. In this work, we first identify several semantic structures behind humor and design sets of features for each theory, and next employ a computational approach to recognize humor. Furthermore, we develop a simple and effective method to extract anchors that enable humor in a sentence. Experiments conducted on two datasets demonstrate that our humor recognizer is effective in automatically distinguishing between humorous and non-humorous texts and our extracted humor anchors correlate quite well with human annotations.

## Session 9C: Spoken Language Processing and Language Modeling (Long Papers)

Room 2

Chair: Isabel Trancoso

### Topic Identification and Discovery on Text and Speech

*Chandler May, Francis Ferraro, Alan McCree, Jonathan Wintrode, Daniel Garcia-Romero, and Benjamin Van Durme* 15:40–16:05

We compare the multinomial i-vector framework from the speech community with LDA, SAGE, and LSA as feature learners for topic ID on multinomial speech and text data. We also compare the learned representations in their ability to discover topics, quantified by distributional similarity to gold-standard topics and by human interpretability. We find that topic ID and topic discovery are competing objectives. We argue that LSA and i-vectors should be more widely considered by the text processing community as pre-processing steps for downstream tasks, and also speculate about speech processing tasks that could benefit from more interpretable representations like SAGE.

### A Dynamic Programming Algorithm for Computing N-gram Posteriors from Lattices

*Dogan Can and Shrikanth Narayanan* 16:05–16:30

Efficient computation of n-gram posterior probabilities from lattices has applications in lattice-based minimum Bayes-risk decoding in statistical machine translation and the estimation of expected document frequencies from spoken corpora. In this paper, we present an algorithm for computing the posterior probabilities of all n-grams in a lattice and constructing a minimal deterministic weighted finite-state automaton associating each n-gram with its posterior for efficient storage and retrieval. Our algorithm builds upon the best known algorithm in literature for computing n-gram posteriors from lattices and leverages the following observations to significantly improve the time and space requirements: i) the n-grams for which the posteriors will be computed typically comprises all n-grams in the lattice up to a certain length, ii) posterior is equivalent to expected count for an n-gram that do not repeat on any path, iii) there are efficient algorithms for computing n-gram expected counts from lattices. We present experimental results comparing our algorithm with the best known algorithm in literature as well as a baseline algorithm based on weighted finite-state automata operations.

### Bilingual Structured Language Models for Statistical Machine Translation

*Ekaterina Garmash and Christof Monz* 16:30–16:55

This paper describes a novel target-side syntactic language model for phrase-based statistical machine translation, bilingual structured language model. Our approach represents a new way to adapt structured language models (Chelba and Jelinek, 2000) to statistical machine translation, and a first attempt to adapt them to phrase-based statistical machine translation. We propose a number of variations of the bilingual structured language model and evaluate them in a series of rescoring experiments. Rescoring of 1000-best translation lists produces statistically significant improvements of up to 0.7 BLEU over a strong baseline for Chinese-English, but does not yield improvements for Arabic-English.

### Compact, Efficient and Unlimited Capacity: Language Modeling with Compressed Suffix Trees

*Ehsan Shareghi, Matthias Petri, Gholamreza Haffari, and Trevor Cohn* 16:55–17:20

Efficient methods for storing and querying language models are critical for scaling to large corpora and high Markov orders. In this paper we propose methods for modeling extremely large corpora without imposing a Markov condition. At its core, our approach uses a succinct index — a compressed suffix tree — which provides near optimal compression while supporting efficient search. We present algorithms for on-the-fly computation of probabilities under a Kneser-Ney language model. Our technique is exact and although slower than leading LM toolkits, it shows promising scaling properties, which we demonstrate through infinite-order modeling over the full Wikipedia collection.

**Session 9D: Long Paper Posters**

Lower Level Foyer

Chair: Nathan Schneider

15:40-17:20

**ERSOM: A Structural Ontology Matching Approach Using Automatically Learned Entity Representation***Chuncheng Xiang, Tingsong Jiang, Baobao Chang, and Zhifang Sui*

As a key representation model of knowledge, ontology has been widely used in a lot of NLP related tasks, such as semantic parsing, information extraction and text mining etc. In this paper, we study the task of ontology matching, which concentrates on finding semantically related entities between different ontologies that describe the same domain, to solve the semantic heterogeneity problem. Previous works exploit different kinds of descriptions of an entity in ontology directly and separately to find the correspondences without considering the higher level correlations between the descriptions. Besides, the structural information of ontology haven't been utilized adequately for ontology matching. We propose in this paper an ontology matching approach, named ERSOM, which mainly includes an unsupervised representation learning method based on the deep neural networks to learn the general representation of the entities and an iterative similarity propagation method that takes advantage of more abundant structure information of the ontology to discover more mappings.

**A Single Word is not Enough: Ranking Multiword Expressions Using Distributional Semantics***Martin Riedl and Chris Biemann*

We present a new unsupervised mechanism, which ranks word n-grams according to their multiwordness. It heavily relies on a new uniqueness measure that computes, based on a distributional thesaurus, how often an n-gram could be replaced in context by a single-worded term. In addition with a down-weighting mechanism for incomplete terms this forms a new measure called DRUID. Results show large improvements on two small test sets over competitive baselines. We demonstrate the scalability of the method to large corpora, and the independence of the measure of shallow syntactic filtering.

**Syntactic Dependencies and Distributed Word Representations for Analogy Detection and Mining***Likun Qiu, Yue Zhang, and Yanan Lu*

Distributed word representations capture relational similarities by means of vector arithmetics, giving high accuracies on analogy detection. We empirically investigate the use of syntactic dependencies on improving analogy detection based on distributed word representations, showing that a dependency-based embeddings does not perform better than an ngram-based embeddings, but dependency structures can be used to improve analogy detection by filtering candidates. In addition, we show that distributed representations of dependency structure can be used for measuring relational similarities, thereby help analogy mining.

**Navigating the Semantic Horizon using Relative Neighborhood Graphs***Amaru Cuba Gyllensten and Magnus Sahlgren*

This paper introduces a novel way to navigate neighborhoods in distributional semantic models. The approach is based on relative neighborhood graphs, which uncover the topological structure of local neighborhoods in semantic space. This has the potential to overcome both the problem with selecting a proper k in k-NN search, and the problem that a ranked list of neighbors may conflate several different senses. We provide both qualitative and quantitative results that support the viability of the proposed method.

**Multi- and Cross-Modal Semantics Beyond Vision: Grounding in Auditory Perception***Douwe Kiela and Stephen Clark*

Multi-modal semantics has relied on feature norms or raw image data for perceptual input. In this paper we examine grounding semantic representations in raw auditory data, using standard evaluations for multi-modal semantics, including measuring conceptual similarity and relatedness. We also evaluate cross-modal mappings, through a zero-shot learning task mapping between linguistic and auditory modalities. In addition, we evaluate multi-modal representations on an unsupervised musical instrument

clustering task. To our knowledge, this is the first work to combine linguistic and auditory information into multi-modal representations.

### **Automatic recognition of habituals: a three-way classification of clausal aspect**

*Annemarie Friedrich and Manfred Pinkal*

This paper provides the first fully automatic approach for classifying clauses with respect to their aspectual properties as habitual, episodic or static. We bring together two strands of previous work, which address only the related tasks of the episodic-habitual and stative-dynamic distinctions, respectively. Our method combines different sources of information found to be useful for these tasks. We are the first to exhaustively classify ALL clauses of a text, achieving up to 80% accuracy (baseline 58%) for the three-way classification task, and up to 85% accuracy for related subtasks (baselines 50% and 60%), outperforming previous work. In addition, we provide a new large corpus of Wikipedia texts labeled according to our linguistically motivated guidelines.

### **Distributed Representations for Unsupervised Semantic Role Labeling**

*Kristian Woodsend and Mirella Lapata*

We present a new approach for unsupervised semantic role labeling that leverages distributed representations. We induce embeddings to represent a predicate, its arguments and their complex interdependence. Argument embeddings are learned from surrounding contexts involving the predicate and neighboring arguments, while predicate embeddings are learned from argument contexts. The induced representations are clustered into roles using a linear programming formulation of hierarchical clustering, where we can model task-specific knowledge. Experiments show improved performance over both previous unsupervised semantic role labeling approaches and other distributed word representation models.

### **A Tableau Prover for Natural Logic and Language**

*Lasha Abzianidze*

Modeling the entailment relation over sentences is one of the generic problems of natural language understanding. In order to account for this problem, we design a theorem prover for Natural Logic, a logic whose terms resemble natural language expressions. The prover is based on an analytic tableau method and employs syntactically and semantically motivated schematic rules. Pairing the prover with a preprocessor, which generates formulas of Natural Logic from linguistic expressions, results in a proof system for natural language. It is shown that the system obtains a comparable accuracy (81%) on the unseen SICK data while achieving the state-of-the-art precision (98%).

**Session 9E: Short Paper Posters**

Lower Level Foyer

Chair: Kang Liu

15:40–17:20

**JEAM: A Novel Model for Cross-Domain Sentiment Classification Based on Emotion Analysis***Kun-Hu Luo, Zhi-Hong Deng, Hongliang Yu, and Liang-Chen Wei*

Cross-domain sentiment classification (CSC) aims at learning a sentiment classifier for unlabeled data in the target domain based on the labeled data from a different source domain. Due to the differences of data distribution of two domains in terms of the raw features, the CSC problem is difficult and challenging. Previous researchers mainly focused on concepts mining by clustering words across data domains, which ignored the importance of authors' emotion contained in data, or the different representations of the emotion between domains. In this paper, we propose a novel framework to solve the CSC problem, by modelling the emotion across domains. We first develop a probabilistic model named JEAM to model author's emotion state when writing. Then, an EM algorithm is introduced to solve the likelihood maximization problem and to obtain the latent emotion distribution of the author. Finally, a supervised learning method is utilized to assign the sentiment polarity to a given review. Extensive experiments show that our approach is effective and outperforms state-of-the-art approaches.

**PhraseRNN: Phrase Recursive Neural Network for Aspect-based Sentiment Analysis***Thien Hai Nguyen and Kyoaki Shirai*

This paper presents a new method to identify sentiment of an aspect of an entity. It is an extension of RNN (Recursive Neural Network) that takes both dependency and constituent trees of a sentence into account. Results of an experiment show that our method significantly outperforms previous methods.

**ASTD: Arabic Sentiment Tweets Dataset***Mahmoud Nabil, Mohamed Aly, and Amir Atiya*

This paper introduces ASTD, an Arabic social sentiment analysis dataset gathered from Twitter. It consists of about 10,000 tweets which are classified as objective, subjective positive, subjective negative, and subjective mixed. We present the properties and the statistics of the dataset, and run experiments using standard partitioning of the dataset. Our experiments provide benchmark results for 4 way sentiment classification on the dataset.

**Adjective Intensity and Sentiment Analysis***Raksha Sharma, Mohit Gupta, Astha Agarwal, and Pushpak Bhattacharyya*

For fine-grained sentiment analysis, we need to go beyond zero-one polarity and find a way to compare adjectives that share a common semantic property. In this paper, we present a semi-supervised approach to assign intensity levels to adjectives, viz. high, medium and low, where adjectives are compared when they belong to the same semantic category. For example, in the semantic category of EXPERTISE, expert, experienced and familiar are respectively of level high, medium and low. We obtain an overall accuracy of 77% for intensity assignment. We show the significance of considering intensity information of adjectives in predicting star-rating of reviews. Our intensity based prediction system results in an accuracy of 59% for a 5-star rated movie review corpus.

**The Rating Game: Sentiment Rating Reproducibility from Text***Lasse Borgholt, Peter Simonsen, and Dirk Hovy*

Sentiment analysis models often use ratings as labels, assuming that these ratings reflect the sentiment of the accompanying text. We investigate (i) whether human readers can infer ratings from review text, (ii) how human performance compares to a regression model, and (iii) whether model performance is affected by the rating "source" (i.e. original author vs. annotator). We collect IMDb movie reviews with author-provided ratings, and have them re-annotated by crowdsourced and trained annotators. Annotators reproduce the original ratings better than a model, but are still far off in more than 5% of the cases. Models trained on annotator-labels outperform those trained on author-labels, questioning the usefulness of author-rated reviews as training data for sentiment analysis.

**A Multi-lingual Annotated Dataset for Aspect-Oriented Opinion Mining***Salud M. Jiménez-Zafra, Giacomo Berardi, Andrea Esuli, Diego Marcheggiani, María Teresa Martín-Valdivia, and Alejandro Moreo Fernández*

We present the Trip-MAML dataset, a Multi-Lingual dataset of hotel reviews that have been manually annotated at the sentence-level with Multi-Aspect sentiment labels. This dataset has been built as an extension of an existent English-only dataset, adding documents written in Italian and Spanish. We detail the dataset construction process, covering the data gathering, selection, and annotation. We present inter-annotator agreement figures and baseline experimental results, comparing the three languages. Trip-MAML is a multi-lingual dataset for aspect-oriented opinion mining that enables researchers (i) to face the problem on languages other than English and (ii) to experiment the application of cross-lingual learning methods to the task.

### **Deep Convolutional Neural Network Textual Features and Multiple Kernel Learning for Utterance-level Multimodal Sentiment Analysis**

*Soujanya Poria, Erik Cambria, and Alexander Gelbukh*

We present a novel way of extracting features from short texts, based on the activation values of an inner layer of a deep convolutional neural network. We use the extracted features in multimodal sentiment analysis of short video clips representing one sentence each. We use the combined feature vectors of textual, visual, and audio modalities to train a classifier based on multiple kernel learning, which is known to be good at heterogeneous data. We obtain 14% performance improvement over the state of the art and present a parallelizable decision-level data fusion method, which is much faster, though slightly less accurate.

### **SLSA: A Sentiment Lexicon for Standard Arabic**

*Ramy Eskander and Owen Rambow*

Sentiment analysis has been a major area of interest, for which the existence of high-quality resources is crucial. In Arabic, there is a reasonable number of sentiment lexicons but with major deficiencies. The paper presents a large-scale Standard Arabic Sentiment Lexicon (SLSA) that is publicly available for free and avoids the deficiencies in the current resources. SLSA has the highest up-to-date reported coverage. The construction of SLSA is based on linking the lexicon of AraMorph with SentiWordNet along with a few heuristics and powerful back-off. SLSA shows a relative improvement of 37.8% over a state-of-the-art lexicon when tested for accuracy. It also outperforms it by an absolute 3.5% of F1-score when tested for sentiment analysis.

### **Reinforcing the Topic of Embeddings with Theta Pure Dependence for Text Classification**

*Ning Xing, Yuxian Hou, Peng Zhang, Wenjie Li, and Dawei Song*

For sentiment classification, it is often recognized that embedding based on distributional hypothesis is weak in capturing sentiment contrast—contrasting words may have similar local context. Based on broader context, we propose to incorporate Theta Pure Dependence (TPD) into the Paragraph Vector method to reinforce topical and sentimental information. TPD has a theoretical guarantee that the word dependency is pure, i.e., the dependence pattern has the integral meaning whose underlying distribution can not be conditionally factorized. Our method outperforms the state-of-the-art performance on text classification tasks.

### **That's So Annoying!!!: A Lexical and Frame-Semantic Embedding Based Data Augmentation Approach to Automatic Categorization of Annoying Behaviors using #petpeeve Tweets**

*William Yang Wang and Diyi Yang*

We propose a novel data augmentation approach to enhance computational behavioral analysis using social media text. In particular, we collect a Twitter corpus of the descriptions of annoying behaviors using the #petpeeve hashtags. In the qualitative analysis, we study the language use in these tweets, with a special focus on the fine-grained categories and the geographic variation of the language. In quantitative analysis, we show that lexical and syntactic features are useful for automatic categorization of annoying behaviors, and frame-semantic features further boost the performance; that leveraging large lexical embeddings to create additional training instances significantly improves the lexical model; and incorporating frame-semantic embedding achieves the best overall performance.

### **Detection of Steganographic Techniques on Twitter**

*Alex Wilson, Phil Blunsom, and Andrew Ker*

We propose a method to detect hidden data in English text. We target a system previously thought secure, which hides messages in tweets. The method brings ideas from image steganalysis into the linguistic domain, including the training of a feature-rich model for detection. To identify Twitter users guilty of steganography, we aggregate evidence; a first, in any domain. We test our system on a set of 1M steganographic tweets, and show it to be effective.

### #SupportTheCause: Identifying Motivations to Participate in Online Health Campaigns

*Dong Nguyen, Tjits van den Broek, Claudia Hauff, Djoerd Hiemstra, and Michel Ehrenhard*  
 We consider the task of automatically identifying participants' motivations in the public health campaign Movember and investigate the impact of the different motivations on the amount of campaign donations raised. Our classification scheme is based on the Social Identity Model of Collective Action (van Zomeren et al., 2008). We find that automatic classification based on Movember profiles is fairly accurate, while automatic classification based on tweets is challenging. Using our classifier, we find a strong relation between types of motivations and donations. Our study is a first step towards scaling-up collective action research methods.

### An Analysis of Domestic Abuse Discourse on Reddit

*Nicolas Schrading, Cecilia Ovesdotter Alm, Ray Ptucha, and Christopher Homan*

Domestic abuse affects people of every race, class, age, and nation. There is significant research on the prevalence and effects of domestic abuse; however, such research typically involves population-based surveys that have high financial costs. This work provides a qualitative analysis of domestic abuse using data collected from the social and news-aggregation website reddit.com. We develop classifiers to detect submissions discussing domestic abuse, achieving accuracies of up to 92%, a substantial error reduction over its baseline. Analysis of the top features used in detecting abuse discourse provides insight into the dynamics of abusive relationships.

### Twitter-scale New Event Detection via K-term Hashing

*Dominik Wurzer, Victor Lavrenko, and Miles Osborne*

First Story Detection is hard because the most accurate systems become progressively slower with each document processed. We present a novel approach to FSD, which operates in constant time/space and scales to very high volume streams. We show that when computing novelty over a large dataset of tweets, our method performs 192 times faster than a state-of-the-art baseline without sacrificing accuracy. Our method is capable of performing FSD on the full Twitter stream on a single core of modest hardware.

### Classifying Tweet Level Judgements of Rumours in Social Media

*Michał Lukasik, Trevor Cohn, and Kalina Bontcheva*

Social media is a rich source of rumours and corresponding community reactions. Rumours reflect different characteristics, some shared and some individual. We formulate the problem of classifying tweet level judgements of rumours as a supervised learning task. Both supervised and unsupervised domain adaptation are considered, in which tweets from a rumour are classified on the basis of other annotated rumours. We demonstrate how multi-task learning helps achieve good results on rumours from the 2011 England riots.

### Identification and Verification of Simple Claims about Statistical Properties

*Andreas Vlachos and Sebastian Riedel*

In this paper we study the identification and verification of simple claims about statistical properties, e.g. claims about the population or the inflation rate of a country. We show that this problem is similar to extracting numerical information from text and following recent work, instead of annotating data for each property of interest in order to learn supervised models, we develop a distantly supervised baseline approach using a knowledge base and raw text. In experiments on 16 statistical properties about countries from Freebase we show that our approach identifies simple statistical claims about properties with 60% precision, while it is able to verify these claims without requiring any explicit supervision for either tasks. Furthermore, we evaluate our approach as a statistical property extractor and we show it achieves 0.11 mean absolute percentage error.



## Index

### A

Abdelali, Ahmed, 28, 100  
Aberton, Maxime, 40  
Abouelenien, Mohamed, 151  
Abrecht, Viktoria, 130  
Abujabal, Abdalghani, 84  
Abzianidze, Lasha, 154  
Afantenos, Stergos, 86  
Agarwal, Astha, 155  
Agirre, Eneko, 36  
Aharoni, Ehud, 61  
Ahlbom, Rickard, 33  
Alberti, Chris, 102  
Alfalahi, Alyaa, 33  
Alfonseca, Enrique, 59  
Alishahi, Afra, 39  
Alkhouli, Tamer, 26, 107  
Allauzen, Alexandre, 25, 26, 51, 89  
Allen, James, 87  
Alm, Cecilia Ovesdotter, 11, 21  
Al-Mannai, Kamla, 100  
Almeida, Inês, 40  
Alonso, Miguel A., 29  
Althobaiti, Maha, 74  
Altosaar, Jaan, 111  
Alves, Ana, 40  
Aly, Mohamed, 155  
Alzate Perez, Carlos, 61  
Amir, Silvio, 102, 111  
Ananiadou, Sophia, 8  
Anderson, Tim, 25  
Ando, Yuji, 145  
Andreas, Jacob, 97

Angeli, Gabor, 71, 117, 119  
Apidianaki, Marianna, 27  
Araki, Jun, 140  
Araki, Masahiro, 145  
Artzi, Yoav, 46, 114, 117, 119  
Asami, Taichi, 129  
Asher, Nicholas, 86  
Asker, Lars, 33  
Asoh, Hideki, 145  
Atiya, Amir, 155  
Atserias, Jordi, 27  
Attamimi, Muhammad, 145  
Augenstein, Isabelle, 74  
Avramidis, Eleftherios, 24  
Aziz, Wilker, 31

### B

B. Pelz, Jeff, 39  
Bai, Shuanhu, 72  
Bai, Tao, 60  
Bakhshandeh, Omid, 87  
Balahur, Alexandra, 29  
Balasubramanian, Niranjan, 72  
Baldwin, Tim, 35, 134, 142  
Ballesteros, Miguel, 58  
Bamman, David, 48  
Banchs, Rafael E., 144  
Bansal, Mohit, 123  
Bansal, Pradeep, 138  
Baral, Chitta, 88  
Barbu, Andrei, 110  
Baroni, Marco, 7, 40, 46, 138  
Barrett, Maria, 37

- Barrón-Cedeño, Alberto, 65  
Bartie, Phil, 130  
Barzilay, Regina, 46, 71, 112, 127  
Baskalayci, Roza, 33  
Basu, Anupam, 30  
Batista, David S., 63  
Batista, Fernando, 8  
Batra, Shashank, 50  
Beck, Daniel, 27, 150  
Beekhuizen, Barend, 38  
Bekki, Daisuke, 138  
Bel, Núria, 27  
Belinkov, Yonatan, 146  
Bellandi, Tommaso, 34  
Belz, Anja, 39, 40  
Bengio, Yoshua, 25, 42, 80  
Benhaloum, Amine, 91  
Benjelloun, Sami, 40  
Berardi, Giacomo, 155  
Berdicevskis, Aleksandrs, 37  
Berg-Kirkpatrick, Taylor, 53  
Bernardi, Raffaella, 36  
Berwick, Robert, 37  
Berzak, Yevgeni, 110  
Bethard, Steven, 33, 87, 145  
Bhatia, Parminder, 144  
Bhattacharyya, Pushpak, 29, 50, 155  
Bicici, Ergun, 24, 27  
Biemann, Chris, 153  
Billingsley, Richard, 57  
Bing, Lidong, 64  
Biran, Or, 131  
Birch, Alexandra, 25  
Bird, Steven, 58  
Bisazza, Arianna, 32  
Black, Alan W, 102, 111  
Blackwood, Graeme, 129  
Blain, Frédéric, 27  
Bloem, Jelke, 37  
Blunsom, Phil, 156  
Bogoychev, Nikolay, 25  
Bojar, Ondřej, 24, 26  
Bojar, Ondrej, 24  
Boleda, Gemma, 36, 46  
Bontcheva, Kalina, 52, 157  
Bordes, Antoine, 53  
Borgholt, Lasse, 155  
Bosch, Antal van den, 38  
Bosnjak, Matko, 116  
Boudin, Florian, 129  
Bougares, Fethi, 27, 90  
Bowen, Victor, 48  
Bowman, Sam, 81, 87  
Bowman, Samuel R., 71, 117, 119  
Boyd-Graber, Jordan, 7, 47, 52, 57  
Brandes, Jasper, 30  
Braud, Chloé, 143  
Braune, Fabienne, 91  
Bresnan, Joan, 38  
Brochhagen, Thomas, 38  
Broek, Tijs van den, 33, 157  
Brun, Caroline, 91  
Bryce, Bill, 26  
Buitelaar, Paul, 142  
Bulgarov, Florin, 151  
Bungum, Lars, 30  
Burchardt, Aljoscha, 24  
Burlot, Franck, 25  
Burzo, Mihai, 151  
Busse, Sebastian, 34  
Byrne, Bill, 145
- C**
- Cagnina, Leticia, 29  
Cai, Dongfeng, 27  
Callin, Jimmy, 32  
Callison-Burch, Chris, 5, 6, 34, 96  
Cambria, Erik, 156  
Can, Dogan, 152  
Candeias, Sara, 40  
Cao, Yixin, 72  
Cap, Fabienne, 25  
Caragea, Cornelia, 151  
Carman, Mark, 29  
Carpuat, Marine, 29, 31  
Carput, Marine, 7  
Carreras, Xavier, 63  
Cassani, Giovanni, 37  
Castilho, Sheila, 37  
Çetinoğlu, Özlem, 128  
Cettolo, Mauro, 31  
Chai, Kian Ming A., 60  
Chambers, Nate, 54, 60  
Chambers, Nathanael, 48  
Chang, Angel, 40  
Chang, Baobao, 78, 114, 153  
Chang, Shih-Fu, 51  
Charniak, Eugene, 58, 102

- Chatterjee, Rajen, 24  
Chaudhari, Sneha, 64  
Chen, Boxing, 27, 93, 99  
Chen, Danqi, 110  
Chen, Daoxu, 60  
Chen, Jinying, 34  
Chen, Liwei, 77  
Chen, Qingcai, 131  
Chen, Wei, 98  
Chen, Xinchi, 76, 98, 127, 150  
Chen, Yadong, 72  
Chen, Yubo, 125  
Chen, Yunchuan, 125, 141  
Chen, Zheng, 52  
Cheng, Hao, 73  
Cheng, Jianpeng, 111  
Cheung, Jackie Chi Kit, 49  
Chiang, David, 85, 90  
Chiticariu, Laura, 11, 15  
Chitnis, Rohan, 140  
Cho, Eunah, 25  
Cho, Kyunghyun, 25  
Choe, Do Kook, 102  
Choi, Jaesik, 113  
Choi, Yejin, 87, 114  
Chopra, Sumit, 59  
Choudhury, Pallavi, 110  
Chrupała, Grzegorz, 39  
Clark, Chris, 35  
Clark, Peter, 72, 136  
Clark, Stephen, 35, 49, 94, 101, 138, 153  
Cocos, Anne, 34  
Cohan, Arman, 59  
Cohen, Shay, 8  
Cohen, Shay B., 111, 127  
Cohen, William, 64, 115  
Coheur, Luisa, 39  
Coheur, Luísa, 8, 40  
Cohn, Trevor, 7, 52, 58, 150, 152, 157  
Collier, Nigel, 115  
Collins, Michael, 51, 58  
Colmenares, Carlos A., 59  
Çöltekin, Çağrı, 37  
Comelles, Elisabet, 27  
Cook, Paul, 58  
Coppola, Greg, 102  
Cordero-Rama, Jose, 39  
Cornudella, Miquel, 37  
Correia, Rui, 35  
Cotterell, Ryan, 85, 145  
Coulmance, Jocelyn, 91  
Crabbé, Benoit, 127  
Cristia, Alex, 38  
Cruz Diaz, Noa P., 33  
Csértő, István, 30  
Cuba Gyllenstein, Amaru, 153  
Cunha, Iria da, 27
- D**
- Da San Martino, Giovanni, 65  
Daelemans, Walter, 37  
Dagan, Ido, 105, 114  
Dai, Shuaixiang, 77  
Dalianis, Hercules, 33, 34  
Dankin, Lena, 61  
Das, Dipanjan, 87  
Daumé III, Hal, 47  
Del Corro, Luciano, 84, 125  
Dellandrea, Emmanuel, 51  
Delli Bovi, Claudio, 73  
Dell'Orletta, Felice, 34  
Delmonte, Rodolfo, 35  
Demberg, Vera, 7  
DeNero, John, 47, 89, 140  
Deng, Lingjia, 50  
Deng, Zhi-Hong, 155  
Denis, Pascal, 86, 143  
Desmet, Bart, 28  
Devlin, Jacob, 51, 52  
Dey, Lipika, 30  
D'hondt, Eva, 33  
Dickinson, Anna, 35  
Dietz, Laura, 39  
Dima, Corina, 114  
Ding, Chenchen, 89  
DO, Quoc-Khanh, 25, 26, 89  
Do, Quoc-Khanh, 25  
Do, Quynh Ngoc Thi, 145  
Dowling, Philipp, 28  
Downey, Doug, 57  
Dredze, Mark, 58, 64, 125  
D'Souza, Jennifer, 74  
Du, Lan, 57  
Duan, Hong, 144  
Duh, Kevin, 32, 44, 51  
Dušek, Ondřej, 25  
Duong, Long, 58  
Dupoux, Emmanuel, 38

Durrani, Nadir, 100  
Dutta, Sourav, 74, 84  
Dwi Prasetyo, Nugroho, 33  
Dyer, Chris, 58, 85, 102, 111, 138, 151  
Dymetman, Marc, 132

**E**

Ebert, Sebastian, 30  
Eckle-Kohler, Judith, 144  
Eger, Steffen, 98  
Ehrenhard, Michel, 157  
Eisenstein, Jacob, 86, 142, 144  
Eisner, Jason, 58, 85, 150  
Elgammal, Ahmed, 39  
Elhoseiny, Mohamed, 39  
Ellebracht, Lily D., 39  
Erdmann, Grant, 25, 28  
Eshghi, Arash, 40  
Eskander, Ramy, 156  
Eskenazi, Maxine, 35  
Espinosa Anke, Luis, 73  
Esplà-Gomis, Miquel, 26, 27  
Esuli, Andrea, 155  
Etzioni, Oren, 72, 110  
Evans, James, 111  
Evers-Vermeul, Jacqueline, 31

**F**

Faloutsos, Christos, 90  
Fang, Hao, 73, 137  
Farahmand, Meghdad, 123  
Faret, Jørgen, 30  
Farhadi, Ali, 110  
Faria, Pablo, 37  
Faruqui, Manaal, 138  
Farzindar, Atefeh, 11, 19  
Favre, Benoit, 129  
Federmann, Christian, 24  
Fei, Geli, 151  
Fei-Fei, Li, 40  
Feng, Minwei, 28, 141  
Feng, Yansong, 64  
Fernandez, Ramon, 102, 111  
Fernandez, Raquel, 36  
Ferrari, Vittorio, 39  
Ferraro, Francis, 51, 71, 115, 152  
Filice, Simone, 65  
Filippova, Katja, 59  
Finch, Andrew, 90, 126

Firat, Orhan, 25  
Fishel, Mark, 31  
FitzGerald, Nicholas, 87  
Flekova, Lucie, 30, 126  
Fomicheva, Marina, 27  
Forcada, Mikel, 27  
Foster, Jennifer, 102  
Foulds, James, 87  
Fox, Peter, 73  
Frank, Anette, 35  
Fraser, Alexander, 25, 91, 145  
Freitag, Markus, 28  
Friedrich, Annemarie, 35, 154  
Friedrich, Stefanie, 34  
Fu, Ruiji, 145  
Fu, Xiao, 90  
Fujii, Atsushi, 70  
Fukumoto, Fumiyo, 76  
Fülöp, Éva, 30  
Funaki, Ruka, 66  
Funakoshi, Kotaro, 145  
Futrell, Richard, 131

**G**

Gaizauskas, Robert, 39, 51  
Galley, Michel, 51  
Gambäck, Björn, 30  
Gamon, Michael, 110  
Ganchev, Kuzman, 87  
Gao, Wei, 143  
Garcia-Duran, Alberto, 53  
Garcia-Romero, Daniel, 152  
Gardner, Matt, 90, 110  
Garmash, Ekaterina, 152  
Gasic, Milica, 117, 119  
Geigle, Chase, 26  
Gelbukh, Alexander, 156  
Gemulla, Rainer, 84, 125  
Genabith, Josef van, 24, 25, 27, 89  
Genco, Ethan, 48  
Gerow, Aaron, 111  
Gertz, Michael, 64  
Getoor, Lise, 87  
Ghazvininejad, Marjan, 85  
Ghosh, Debanjan, 88  
Gibson, Edward, 131  
Giles, C. Lee, 61, 115  
Gillian, Edward, 61  
Gillis, Steven, 37

Gimpel, Kevin, 112, 123, 140  
Ginter, Filip, 25  
Gispert, Adrià de, 145  
Gkatzia, Dimitra, 130  
Glass, James, 146  
Goharian, Nazli, 59  
Goldberg, Yoav, 102  
Gollapalli, Sujatha Das, 136  
Gomes, Luís, 25  
Gómez-Rodríguez, Carlos, 29  
Gong, Yeyun, 60  
Gong, Zhengxian, 31  
Goodman, Noah, 37  
Goot, Erik van der, 29  
Gormley, Matthew R., 58, 125  
Graça, João, 6, 8  
Graca, Miguel, 26  
Graham, Yvette, 49, 59  
Grant, Erin, 126  
Greco, Nicola, 116  
Green, Spence, 89  
Gribkoff, Eric, 72  
Griffon, Nicolas, 33  
Grimm, Robert, 37  
Grishman, Ralph, 35  
Grissom II, Alvin, 47  
Grönroos, Stig-Arne, 25, 27  
Grouin, Cyril, 33  
Grundkiewicz, Roman, 61  
Grycner, Adam, 87  
Gu, Qing, 60  
Guckelsberger, Christian, 60  
Guillou, Liane, 31, 32  
Gunn, Martin, 33  
Guo, Hongyu, 27  
Guo, Li, 115  
Guo, Weiwei, 88  
Guo, Xiaofei, 72  
Gupta, Abhijeet, 46  
Gupta, Mohit, 155  
Gupta, Rohit, 27, 89  
Gurevych, Iryna, 53, 126, 142, 144  
Guta, Andreas, 26, 107  
Guu, Kelvin, 57  
Guzmán, Fanciscog, 6  
Guzmán, Francisco, 28  
Gwinnup, Jeremy, 25, 28

**H**  
Ha, Thanh-Le, 25  
Habash, Nizar, 76, 101, 126  
Habernal, Ivan, 142  
Haddow, Barry, 24, 25, 140  
Haffari, Gholamreza, 152  
Hajishirzi, Hannaneh, 110, 137  
Hamon, Thierry, 33  
Han, Xianpei, 99  
Hanbury, Allan, 70  
Harari, Daniel, 110  
Hardmeier, Christian, 31, 32, 150  
Harihara, Ganesh, 48  
Hasan, Saša, 61  
Hashimoto, Chikara, 114, 143  
Hauff, Claudia, 33, 157  
He, He, 47  
He, Hua, 112  
He, Luheng, 71, 109  
He, Xiaodong, 24  
He, Yifan, 35  
He, Yulan, 130  
Hegde, Manjunath, 64  
Heger, Carmen, 61  
Henderson, James, 52, 63, 123  
Henriksson, Aron, 33, 34  
Herbelot, Aurélie, 46  
Hermjakob, Ulf, 97  
Herrmann, Teresa, 25  
Hessel, Jack, 39  
Hiemstra, Djoerd, 33, 157  
Higashinaka, Ryuichiro, 145  
Hilderman, Robert, 30  
Hill, Felix, 138  
Hirvonen, Maija, 39  
Hockenmaier, Julia, 7  
Hoek, Jet, 31  
Hokamp, Chris, 24, 27  
Homan, Christopher, 157  
Hong, Kai, 49  
Hong, Yu, 72  
Hannibal, Matthew, 102  
Hoste, Veronique, 28  
Hou, Yuexian, 156  
Hovy, Dirk, 30, 155  
Hovy, Eduard, 101, 150, 151  
Hu, Baotian, 131  
Hu, Linmei, 76  
Hu, Yifan, 61

- Huang, Fei, 142  
Huang, Kejun, 90  
Huang, Liang, 100  
Huang, Shujian, 99  
Huang, Songfang, 64  
Huang, Ting-Hao, 51  
Huang, Wenyi, 115  
Huang, Xaojiang, 84  
Huang, Xuanjing, 60, 76, 98, 127, 150  
Huang, Ziheng, 143  
Huck, Matthias, 24–26
- I**
- Iglesias, Gonzalo, 145  
Iida, Ryu, 143  
Inkpen, Diana, 11, 19, 30  
Inui, Kentaro, 7  
Inui, Takashi, 29  
Isozaki, Hideki, 28  
Ito, Akinori, 129  
Ittycheriah, Abraham, 32  
Ive, Julia, 25
- J**
- Jaakkola, Tommi, 112  
Jaech, Aaron, 137  
Jagannatha, Abhyuday, 34  
Jain, Sarthak, 50  
Jamison, Emily, 53  
Jäschke, Robert, 48  
Jayanth, Jayanth, 50  
Jean So, Richard, 111  
Jean, Sébastien, 25  
Jermsurawong, Jermsak, 76  
Ji, Donghong, 127  
Ji, Duo, 27  
Ji, Heng, 7, 51, 72, 73, 84  
Ji, Yangfeng, 86, 144  
Jiang, Jing, 7  
Jiang, Tingsong, 114, 153  
Jiang, Zhiwei, 60  
Jiménez-Zafra, Salud M., 155  
Jin, Lifeng, 91  
Jin, Zhi, 125, 141, 150  
Joachims, Thorsten, 57  
Johannsen, Anders, 36, 138  
Johnson, Mark, 57, 102  
Joshi, Aditya, 29  
Joshi, Aravind, 35
- Joty, Shafiq, 65, 100, 109  
Junczys-Dowmunt, Marcin, 61  
Jurafsky, Dan, 123, 150  
Jurgens, David, 11, 12, 71, 76
- K**
- Kádár, Ákos, 39  
Kaeshammer, Miriam, 26  
Kaiser, Lukasz, 59  
Kaljahi, Rasoul, 102  
Kameko, Hirotaka, 146  
Kamigaito, Hidetaka, 99  
Kamiran, Faisal, 77  
Kamran, Amir, 26  
Kanerva, Jenna, 25  
Karim, Asim, 77  
Kartsaklis, Dimitri, 111  
Katz, Boris, 110  
Kawahara, Daisuke, 146  
Kazi, Michaeel, 25  
Ker, Andrew, 156  
Kerren, Andreas, 30, 33  
Kessler, Wiltrud, 29  
Khalifa, Salam, 101  
Khapra, Mitesh M., 61  
Khattri, Anupam, 29  
Khazaei, Taraneh, 35  
Khot, Tushar, 72  
Khouzaimi, Hatim, 129  
Khudanpur, Sanjeev, 129  
Kiddon, Chloé, 87  
Kiela, Douwe, 49, 138, 153  
Kiesel, Johannes, 70  
Kim, Jonghoon, 76  
Kim, Jung-jae, 88  
Kim, Seokhwan, 144  
Kim, Sun, 77  
Kim, Taehoon, 113  
Kim, Young-Bum, 100  
Kim, Yunsu, 26  
Kiperwasser, Eliyahu, 102  
Klassen, Prescott, 33  
Klein, Dan, 53, 97  
Klementiev, Alex, 57  
Klenner, Manfred, 30  
Klerke, Sigrid, 37  
Klinger, Roman, 29  
Kloetzer, Julien, 114, 143  
Kluge, Roland, 144

- Knight, Kevin, 85, 97  
Kobayashi, Hayato, 132  
Kobayashi, Ichiro, 145  
Kobayashi, Yuka, 145  
Kober, Thomas, 29  
Koehn, Philipp, 24–26  
Köhn, Arne, 138  
Kolachina, Prasanth, 25  
Kondrak, Greg, 7  
Kong, Lingpeng, 140  
Kordoni, Valia, 11, 20  
Korhonen, Anna, 37  
Kotnis, Bhushan, 138  
Kouchi, Natsume, 28  
Kővágó, Pál, 30  
Kow, Eric, 86  
Kozareva, Zornitsa, 6, 29  
Kreutzer, Julia, 27  
Krishna, Ranjay, 40  
Krishnamurthy, Jayant, 39  
Kruschwitz, Udo, 74  
Kruszewski, Germán, 46  
Kuhn, Jonas, 29  
Kuhn, Roland, 27  
Kulkarni, Tejas, 46  
Kumar, Gaurav, 129  
Kummerfeld, Jonathan K., 53  
Kurimo, Mikko, 25, 39  
Kurohashi, Sadao, 146  
Kwan, Sharon, 33
- L**  
Laaksonen, Jorma, 39  
Labeau, Matthieu, 25, 51  
Labutov, Igor, 57  
Lample, Guillaume, 138  
Langlois, David, 27  
Lapata, Mirella, 96, 97, 154  
Lapshinova-Koltunski, Ekaterina, 32  
Laroche, Romain, 129  
Last, Mark, 130  
Lautenbacher, Olli Philippe, 39  
Lavergne, Thomas, 25  
Lavie, Alon, 151  
Lavrenko, Victor, 157  
Lazaridou, Angeliki, 40  
Le Roux, Joseph, 102  
Le, Phong, 97  
Lee, Alan, 35
- Lee, Kenton, 114, 117, 119  
Lee, Mark, 30  
Lee, Sungjin, 61, 65  
Lefevre, Fabrice, 129  
Lei, Tao, 112  
Lemon, Oliver, 7, 40  
Lenci, Alessandro, 37  
Levinboim, Tomer, 90  
Lewis, Mike, 71, 109  
Li, Ge, 125, 141, 150  
Li, Haizhou, 144  
Li, Hang, 6  
Li, Hongzhi, 51  
Li, Jing, 143  
Li, Jiwei, 123, 150  
Li, Juanzi, 72, 76  
Li, Junhui, 109  
Li, Junyi Jessy, 100  
Li, Li, 78  
Li, Liangyou, 28, 47  
Li, Maggie, 6  
Li, Mu, 85  
Li, Shaohua, 112  
Li, Sheng, 85  
Li, Sujian, 77, 114  
Li, Wei, 129  
Li, Wenjie, 77, 156  
Li, Xiaoli, 76, 136  
Li, Yang, 136  
Li, Yanran, 77  
Li, Yunyao, 11, 15  
Liang, Chen, 115  
Liang, Percy, 7, 57  
Liao, Lejian, 65  
Limsopatham, Nut, 115  
Lin, Chin-Yew, 65, 84, 96  
Lin, Chu-Cheng, 102  
Lin, Jimmy, 112  
Lin, Rui, 85  
Lin, Yankai, 73  
Ling, Wang, 102, 111, 138  
Linton, CJ, 151  
Linzen, Tal, 92  
Litman, Diane, 131  
Litvak, Marina, 130  
Liu and Dan Bikel, Bing, 109  
Liu, Bing, 151  
Liu, Chunyang, 126  
Liu, Fei, 131

- Liu, Jiangming, 59  
Liu, Kang, 125, 148, 155  
Liu, Lizhen, 145  
Liu, Pengfei, 98, 109, 150  
Liu, Qun, 24, 27, 28, 47  
Liu, Rong, 72  
Liu, Shujie, 85  
Liu, Song, 73  
Liu, Ting, 109, 145  
Liu, Xiaojiang, 96  
Liu, Yang, 99, 126  
Liu, Yisi, 26  
Liu, Zhiyuan, 73, 115  
Livescu, Karen, 123  
Ljubešić, Nikola, 26  
Lo, Chi-ku, 28  
Loáiciga, Sharid, 32  
Logacheva, Varvara, 24, 27  
Lopez de Lacalle, Oier, 36  
Lopez, Adam, 32  
Löser, Kevin, 25, 51  
Louis, Annie, 35, 96, 111  
Lu, Wei, 55, 63, 84, 109  
Lu, Yanan, 153  
Lu, Yangyang, 141  
Lu, Yaojie, 144  
Lu, Ying, 51  
Luan, Huanbo, 73, 99, 115, 126  
Ludusan, Bogdan, 38  
Luis, Tiago, 111  
Lukasik, Michal, 52, 157  
Lun-Wei, Ku, 7  
Luo, Gang, 84  
Luo, Hongyin, 115  
Luo, Kangqi, 65  
Luo, Kun-Hu, 155  
Luo, Tianyi, 72  
Luo, Wencan, 131  
Luo, Xusheng, 65  
Luo, Yuanfei, 115  
Luong, Ngoc Quang, 32  
Luong, Thang, 37, 107, 150  
Luu Anh, Tuan, 88  
Lynn, Teresa, 102
- M**  
Ma, Qingsong, 27  
Ma, Wei-Yun, 89  
Ma, Xiaogang, 73  
Ma, Xuezhe, 101  
Ma, Yanjun, 136  
Machado, Penousal, 40  
Mackaness, William, 130  
Macken, Lieve, 28  
Madhyastha, Pranava Swaroop, 39  
Mahadevan, Sridhar, 141  
Mahmud, Jalal, 48  
Maiya, Arun, 78  
Malandrakis, Nikolaos, 136  
Malcolm, Clint, 110  
Maletti, Andreas, 26  
Malinovskiy, Anton, 27  
Mamede, Nuno, 35  
Maña López, Manuel, 33  
Manning, Christopher D., 40, 71, 107, 117, 119  
Mansour, Saab, 61  
Maqsud, Umar, 30  
Marcheggiani, Diego, 155  
Marcu, Daniel, 97  
Marcus, Mitchell, 49  
Marie, Benjamin, 25, 27, 89  
Marneffe, Marie-Catherine de, 91  
Màrquez, Lluís, 3, 6, 65  
Marsi, Erwin, 63  
Martínez Alonso, Héctor, 36, 138  
Martínez, Paloma, 33  
Martínez-Gómez, Pascual, 138  
Martins, André, 6, 8  
Martins, Bruno, 8, 63  
Martín-Valdivia, María Teresa, 155  
Marton, Yuval, 6  
Martschat, Sebastian, 86  
Marty, Jean-Marc, 91  
Marujo, Luis, 111  
Masataki, Hirokazu, 129  
Mascarell, Laura, 31  
Masino, Aaron, 34  
Massung, Sean, 26  
Masumura, Ryo, 129  
Matsumoto, Yuji, 32  
Mavridou, Kleio-Isidora, 35  
Max, Aurélien, 89  
May, Chandler, 152  
May, Christina, 25  
May, Jonathan, 97  
Maynard, Diana, 74  
McCarthy, Lucas, 33

- McClosky, David, 102, 120, 127  
McCree, Alan, 152  
McKeown, Kathleen, 89, 131, 142  
Mediani, Mohammed, 25  
Meek, Christopher, 136  
Mehdizadeh Seraj, Ramtin, 107  
Memisevic, Roland, 25  
Menezes, Arul, 52  
Meng, Helen, 109  
Mercer, Robert, 35  
Meunier, Jean-Luc, 137  
Miao, Chunyan, 112  
Miculicich Werlen, Lesly, 32  
Mihalcea, Rada, 91, 151  
Miháltz, Márton, 30  
Miller, John, 57  
Mimno, David, 57  
Min, Zhang, 7  
Mineshima, Koji, 138  
Minkov, Einat, 67, 73, 115  
Mirkin, Shachar, 91, 137  
Mitamura, Teruko, 140  
Mitchell and Partha Talukdar, Meg, 110  
Mitchell, Jeff, 33  
Mitchell, Margaret, 51  
Mitchell, Meg, 8  
Mitchell, Tom, 63, 90, 110  
Mitra, Arindam, 88  
Mitra, Pabitra, 30  
Miyao, Yusuke, 8, 138  
Mizukami, Masahiro, 28, 145  
Mochihashi, Daichi, 145  
Moeenuddin, Muhammad, 77  
Moens, Marie-Francine, 34, 39, 96, 145  
Mohammad, Saif, 11, 21, 30, 43, 49  
Moniz, Helena, 8  
Montemagni, Simonetta, 34  
Montoyo, Andres, 29  
Monz, Christof, 24, 32, 152  
Moore, Robert, 101  
Moreno-Noguer, Francesc, 39, 51  
Moreo Fernández, Alejandro, 155  
Morgan, John, 47  
Mori, Shinsuke, 98, 146  
Morita, Hajime, 146  
Moschitti, Alessandro, 65  
Mostafazadeh, Nasrin, 51  
Mou, Lili, 125, 141, 150  
Mougard, Hugo, 129  
Mrkšić, Nikola, 117, 119  
Müller, Thomas, 145  
Murakami, Yuki, 35  
Muresan, Smaranda, 88  
Murray, Kenton, 85  
Muscat, Adrian, 40
- N**
- Nabil, Mahmoud, 155  
Nadejde, Maria, 26  
Nagai, Takayuki, 145  
Nagata, Masaaki, 101  
Nakamura, Satoshi, 28, 140  
Nakamura, Tomoaki, 145  
Nakashole, Ndapandula, 63  
Nakayama, Hideki, 66  
Nakayama, Yuki, 70  
Nakov, Preslav, 11, 17, 31, 65  
Napoles, Courtney, 96  
Narasimhan, Karthik, 46  
Narayan, Shashi, 127  
Narayanan, Shrikanth, 136, 152  
Naskar, Sudip, 25  
Naskar, Sudip Kumar, 24  
Nastase Pilehvar, Vivi, 17  
Nastase, Vivi, 11, 17  
Navigli, Roberto, 71, 73  
Negi, Sapna, 142  
Negri, Matteo, 24  
Nematzadeh, Aida, 126  
Nenkova, Ani, 49, 100  
Neubig, Graham, 28, 140  
Neuman, Yair, 11, 14  
Neveol, Aurelie, 63  
Névéol, Aurélie, 33  
Ney, Hermann, 25, 26, 28, 107  
Ng, Hwee Tou, 60  
Ng, Jun-Ping, 121, 129, 130  
Ng, Raymond W. M., 90  
Ng, See Kiong, 88  
Ng, Vincent, 74  
Nguyen, Dat Quoc, 57  
Nguyen, Dong, 33, 157  
Nguyen, Khanh, 112  
Nguyen, Thien Hai, 155  
Nie, Zaiqing, 84  
Niehues, Jan, 25, 26  
Noguchi, Masaki, 132  
Noord, Gertjan van, 31

Nopp, Clemens, 70

Novák, Attila, 146

Novák, Michal, 25, 31

Nowson, Scott, 91

## O

O. Alm, Cecilia, 39

Ó Séaghdha, Diarmuid, 17

Oba, Takanobu, 129

O'Connor, Brendan, 7, 48, 112

Ödling, David, 51

O'Donnell, Timothy, 37, 92

Oele, Dieke, 31

Oh, Alice, 7

Oh, Jong-Hoon, 114, 143

Ohkuma, Tomoko, 40

Okumura, Manabu, 99

Orasan, Constantin, 27, 89

Ordan, Noam, 24

Ororbia II, Alexander, 61

Ortiz Rojas, Sergio, 26

Osborne, Miles, 157

Ostendorf, Mari, 7, 24, 73, 137

Österlund, Arvid, 51

Ouyang, Jessica, 142

Ovesdotter Alm, Cecilia, 21, 157

Öztürk, Pinar, 63

## P

Padó, Sebastian, 46

Padró, Lluís, 6

Paetzold, Gustavo, 27

Pal, Santanu, 24, 25

Palmer, Alexis, 35

Palmero Aprosio, Alessio, 77

Pan, Shimei, 48

Pan, Yiqiao, 72

Pantel, Patrick, 110

Papalexakis, Evangelos, 90

Papavassiliou, Vassilis, 26

Paperno, Denis, 36, 46

Paradis, Carita, 30, 33

Parveen, Daraksha, 131

Pasca, Marius, 11, 16

Passonneau, Rebecca J., 143

Pastra, Katerina, 39

Pavalanathan, Umashanthi, 142

Pavlick, Ellie, 34

Payne, Tom, 33

Peate Sørensen, Melissa, 35

Pécheux, Nicolas, 24, 25

Pecina, Pavel, 24

Pedersen, Ted, 33

Peitz, Stephan, 28

Peldszus, Andreas, 86

Pellicer, Elena, 33

Peng, Baolin, 143

Peng, Hao, 125, 141, 150

Peng, Haoruo, 26

Peng, Nanyun, 64, 85

Perez, Julien, 91

Pérez-Rosas, Verónica, 91, 151

Perret, Jérémie, 86

Pershina, Maria, 35

Peter Jan-Thorsten, 25, 28, 107

Petri, Matthias, 152

Petroni, Fabio, 125

Petrov, Slav, 102

Petruck, Miriam R. L., 11, 20

Pham, Hieu, 107

Pham, Ngoc-Quan, 32

Phandi, Peter, 60

Pighin, Daniele, 6

Pilehvar, Mohammad Taher, 11, 12

Pinkal, Manfred, 35, 154

Piper, Andrew, 76

Pirinen, Tommi, 26

Plank, Barbara, 6, 30

Plas, Lonneke van der, 32

Poesio, Massimo, 74

Poibeau, Thierry, 37

Polajnar, Tamara, 35

Polisciuc, Evgheni, 40

Pólya, Tibor, 30

Ponnuraj, Ganesa Thandavam, 87

Ponzetto, Simone Paolo, 39

Poon, Hoifung, 110

Popel, Martin, 25

Popescu-Belis, Andrei, 31, 32

Popović, Maja, 24, 27

Popovic, Maja, 82, 89

Poria, Soujanya, 156

Post, Matt, 24

Potash, Peter, 130

Potts, Christopher, 71, 117, 119

Povey, Daniel, 129

Pradhan, Sameer, 35

Prasad, Rashmi, 35

Preoțiuc-Pietro, Daniel, 30  
 Prokopidis, Prokopis, 26  
 Prud'hommeaux, Emily, 39  
 Ptucha, Ray, 157  
 Pujara, Jay, 87  
 Pust, Michael, 97

**Q**

Qadir, Ashequl, 50  
 Qayyum, Abdul, 77  
 Qian, Tao, 127  
 Qian, Ting, 34  
 Qin, Bing, 109  
 Qiu, Likun, 153  
 Qiu, Xipeng, 76, 98, 127, 150  
 Quattoni, Ariadna, 39  
 Quernheim, Daniel, 25  
 Quirk, Chris, 52

**R**

R. Haake, Anne, 39  
 Rabinovich, Ella, 96  
 Radford, Will, 63  
 Rafae, Abdul, 77  
 Raja, Vignesh, 26  
 Ramakrishna, Anil, 136  
 Rambow, Owen, 134, 144, 156  
 Ramisa, Arnau, 39, 51  
 Ramm, Anita, 25  
 Ramsl, Hans-Martin, 131  
 Ranta, Aarne, 25  
 Rao, Siwei, 73  
 Rasooli, Mohammad Sadegh, 58  
 Rastogi, Pushpendre, 115  
 Rawlins, Kyle, 71  
 Rei, Marek, 52  
 Reisinger, Drew, 71  
 Reiss, Frederick, 11, 15  
 Reitan, Johan, 30  
 Reitter, David, 61  
 Ren, Yafeng, 127  
 Resnik, Philip, 24, 52  
 Ribeyre, Corentin, 102  
 Riedel, Sebastian, 7, 157  
 Riedl, Martin, 153  
 Rieser, Verena, 130  
 Rietig, Felix, 26  
 Riezler, Stefan, 27  
 Riloff, Ellen, 50

Rimell, Laura, 35  
 Rinott, Ruty, 61  
 Rohde, Hannah, 35  
 Roller, Roland, 34  
 Romanov, Alexey, 130  
 Rosa, Rudolf, 25  
 Rosso, Paolo, 29  
 Roth, Dan, 84, 123  
 Roth, Michael, 35, 97  
 Roturier, Johann, 102  
 Rousseau, Francois, 76  
 Routledge, Bryan, 111  
 Roy, Subhro, 26, 123  
 Rubino, Raphael, 26  
 Rudinger, Rachel, 71, 115  
 Rui, Yong, 96  
 Rumshisky, Anna, 130  
 Ruppenhofer, Josef, 30  
 Ruppert, Eugen, 30  
 Rush, Alexander M., 59  
 Ruths, Derek, 76

**S**

Sabharwal, Ashish, 72  
 Sadeque, Farig, 33  
 Sahlgren, Magnus, 30, 51, 153  
 Sajjad, Hassan, 28, 77, 100  
 Sakaki, Shigeyuki, 40  
 Sakauchi, Sumitaka, 129  
 Sakti, Sakriani, 28  
 Salesky, Elizabeth, 25  
 Sánchez-Martínez, Felipe, 27  
 Sanders, Ted J.M., 31  
 Sarikaya, Ruhi, 100  
 Sarkar, Anoop, 107  
 Savva, Nicolas, 39  
 Scarton, Carolina, 24, 27  
 Scarton, Caroline, 24  
 Schamoni, Shigehiko, 27  
 Schamp-Bjerede, Teri, 30  
 Schmidt, Benedikt, 60  
 Schmit, Wernard, 30  
 Schnabel, Tobias, 57, 101  
 Schneider, Nathan, 147, 153  
 Schrading, Nicolas, 157  
 Schulder, Marc, 30  
 Schulz, Axel, 60  
 Schuster, Sebastian, 40  
 Schütze, Hinrich, 30, 53, 73, 101, 145

- Schwartz, Lane, 26  
Séaghdha, Diarmuid Ó, 11, 17  
Seeker, Wolfgang, 128  
Seemann, Nina, 26, 91  
Segura-Bedmar, Isabel, 33  
Seidl, Amanda, 38  
Sekine, Satoshi, 29  
Sennrich, Rico, 26, 47, 140  
Seo, Minjoon, 110  
Sergienya, Irina, 53  
Sha, Lei, 78, 114  
Shah, Kashif, 27, 90  
Shahrour, Anas, 101  
Shang, Liugang, 27  
Shao, Chao, 76  
Shareghi, Ehsan, 152  
Sharma, Raksha, 155  
Shen, Shiqi, 99  
Shi, Shuming, 96, 151  
Shigenaka, Ryosuke, 40  
Shirai, Kyoaki, 155  
Shrestha, Prasha, 33  
Si, Luo, 65  
Siahbani, Maryam, 107  
Sidhaye, Priya, 49  
Sidiropoulos, Nikos, 90  
Siklósi, Borbála, 146  
Silva, Mário J., 63  
Sim Smith, Karin, 25, 31  
Sim, Yanchuan, 111  
Sima'an, Khalil, 6, 27, 47  
Simeon, Credell, 30  
Simion, Andrei, 51  
Simonsen, Peter, 155  
Sinha, Priyanka, 30  
Skeppstedt, Maria, 30, 33  
Slonim, Noam, 61  
Smith, Ellery, 116  
Smith, Noah, 104, 111  
Smith, Noah A., 6, 48, 58, 111, 131, 140  
Smith, Phillip, 30  
Snyder, Benjamin, 100  
Søgaard, Anders, 37, 138  
Solorio, Thamar, 33  
Song, Dandan, 65  
Song, Dawei, 156  
Song, Wei, 145  
Specia, Lucia, 7, 24, 25, 27, 31, 90, 107,  
    150  
                Srijith, P. K., 52  
                Srinivasan, Padmini, 48  
                Stanojević, Miloš, 26, 27, 47  
                Staruk, Elizabeth, 136  
                Stede, Manfred, 86  
                Steedman, Mark, 33  
                Steele, David, 25  
                Stein, Benno, 70  
                Stein, Cliff, 51  
                Stevenson, Mark, 34  
                Stevenson, Suzanne, 38, 126  
                Strötgen, Jannik, 64  
                Strube, Michael, 86, 131  
                Stymne, Sara, 31  
                Su, Jian, 5, 6  
                Su, Jinsong, 99, 144  
                Su, Pei-Hao, 117, 119  
                Suárez-Paniagua, Víctor, 33  
                Sugiyama, Kyoshiro, 28  
                Sui, Zhifang, 114, 153  
                Sultan, Md Arafat, 87  
                Sumanth, Chiraag, 30  
                Sumita, Eiichiro, 89, 90, 99, 126, 140  
                Sumner, Tamara, 87  
                Sun, Fei, 77  
                Sun, Gang, 60  
                Sun, Maosong, 73, 99, 115, 126  
                Sun, Xu, 7, 78  
                Sundararaj, Jayaprakash, 50  
                Suzuki, Yoshimi, 76  
                Szpakowicz, Stan, 11, 17
- T**
- Täckström, Oscar, 87  
Takahasi, Fumihiko, 98  
Takamura, Hiroya, 99  
Takeno, Shunsuke, 101  
Talukdar, Partha P., 64, 90, 138  
Tamchyna, Aleš, 24  
Tan, Liling, 27  
Tang, Duyu, 109  
Tang, Jie, 72  
Taniguchi, Tomoki, 40  
Tannier, Xavier, 33, 63  
Tapi Nzali, Mike Donald, 63  
Teka Hadgu, Asmelash, 48  
Temnikova, Irina, 28  
Tezcan, Arda, 28  
Thompson, Brian, 25

Tian, Xisen, 48

Tiedemann, Jörg, 6, 25, 31, 32

Tien Nguyen, Dat, 40

Tiittula, Liisa, 39

Ting, Liu, 7

Toda, Tomoki, 28

Tonelli, Sara, 77

Toral, Antonio, 26

Torisawa, Kentaro, 114, 143

Toutanova, Kristina, 110

Toutounchi, Farzad, 25

Tran, Nam Khanh, 48

Tran, Tuan, 48

Trancoso, Isabel, 8, 102, 111, 152

Tripodi, Rocco, 35

Trmal, Jan, 129

Tsuboshita, Yukihiro, 40

Tsukahara, Hiroshi, 145

Tsunoda, Takaaki, 29

Tsuruoka, Yoshimasa, 35, 146

Tsvetkov, Yulia, 102, 138

Turchi, Marco, 24

Twitto, Naama, 24

## U

Ullman, Shimon, 110

Upadhyay, Shyam, 26

Usunier, Nicolas, 53

Utiyama, Masao, 89, 90, 126, 140

## V

Vaidyanathan, Preethi, 39

Vala, Hardik, 76

Van Durme, Benjamin, 71, 97, 115, 152

Vanderwende, Lucy, 51

Vandyke, David, 117, 119

Vanetik, Natalia, 130

Váradi, Tamás, 30

Vazirgjannis, Michalis, 76

Vecchi, Eva Maria, 46

Vela, Mihaela, 24, 27, 32

Venkatapathy, Sriram, 132

Venturi, Giulia, 34

Versley, Yannick, 31

Versloot, Arjen, 37

Vilares, David, 29

Villavicencio, Aline, 37

Vinyals, Oriol, 59

Virpioja, Sami, 25, 27

Vlachos, Andreas, 74, 116, 157

Vo, Duy Tin, 70

Vogel, Stephan, 28, 100

Volk, Martin, 31

Vosoughi, Soroush, 29

Vossen, Piek, 29

Vu, Ngoc Thang, 30

Vulić, Ivan, 39, 49

## W

Wachsmuth, Henning, 70

Waibel, Alex, 25, 26

Walker, Marilyn, 50

Wan, Hai, 52

Wan, Xiaojun, 49

Wang, Bin, 115

Wang, Dong, 72

Wang, Han, 73

Wang, Houfeng, 78

Wang, Jian, 72

Wang, Jingang, 65

Wang, Josiah, 39, 51

Wang, Lidan, 141

Wang, Quan, 115

Wang, Richard, 64

Wang, William Yang, 156

Wang, Xiaobin, 72

Wang, Xiaolin, 90, 126

Wang, Xing, 107

Wang, Xuzhong, 76

Wang, Yuehui, 96

Wang, Zhen, 52, 114

Wasserman Pritsker, Evgenia, 115

Watanabe, Taro, 99, 126

Way, Andy, 24, 27, 47

Webber, Bonnie, 31, 32, 35

Weegar, Rebecka, 33

Weerman, Fred, 37

Wees, Marlies van der, 32

Wehrli, Eric, 32

Wei, Furu, 8

Wei, Liang-Chen, 155

Wei, Zhongyu, 143

Weikum, Gerhard, 74, 84, 87

Weiland, Lydia, 39

Weir, David, 29

Weiss, David, 102

Weller, Marion, 25

Wen, Tsung-Hsien, 117, 119

- Wen-tau Yih, Scott, 72  
 Wenzek, Guillaume, 91  
 Weston, Jason, 59  
 Wetzel, Dominikus, 32  
 Wiebe, Janyce, 50, 70  
 Wiegand, Michael, 30  
 Wieting, John, 123  
 Wijaya, Derry Tanti, 63  
 Wilber, Michael, 39  
 Wilbur, W John, 77  
 Williams, Philip, 26  
 Wilson, Alex, 156  
 Wintner, Shuly, 24, 96  
 Wintrode, Jonathan, 152  
 Wisniewski, Guillaume, 24  
 Wong, Kam-Fai, 143  
 Woodsend, Kristian, 154  
 Wu, Dekai, 28  
 Wu, Shiyu, 76, 150  
 Wu, Xiaofeng, 27  
 Wu, Zhaohui, 115  
 Wubben, Sander, 30  
 Ruebker, Joern, 25, 26, 89, 107  
 Wurzer, Dominik, 157
- X**  
 Xiang, Bing, 141  
 Xiang, Chuncheng, 153  
 Xiao, Chunyang, 132  
 Xiao, Jianguo, 49  
 Xiao, Lu, 35  
 Xiao, Yao, 151  
 Xiaoyan, Zhu, 7  
 Xing, Ning, 156  
 Xiong, Deyi, 99, 107, 144  
 Xu, Bo, 98  
 Xu, Haiyang, 130  
 Xu, Kun, 64  
 Xu, Wei, 68, 76, 96  
 Xu, Yan, 125, 150
- Y**  
 Yaghoobzadeh, Yadollah, 73  
 Yamamoto, Kazuhide, 101  
 Yang, Chao, 48  
 Yang, Diyi, 151, 156  
 Yang, Eugene, 48  
 Yang, Huahai, 48  
 Yang, Muyun, 85
- Yang, Weiwei, 52  
 Yang, Yi, 57, 136  
 Yao, Jin-ge, 49  
 Yao, Junfeng, 99, 144  
 Yatsuka, Taichi, 132  
 Yazdani, Majid, 52, 123  
 Yeganova, Lana, 77  
 Yetisgen, Meliha, 33  
 Yih, Wen-tau, 136  
 Yim, Wen-wai, 33  
 Yin, Wenpeng, 101  
 Yogatama, Dani, 131, 140  
 Yoshino, Koichiro, 28  
 Young, Eric, 48  
 Young, Katherine, 25  
 Young, Steve, 117, 119  
 Yu, Heng, 126  
 Yu, Hong, 34, 65  
 Yu, Hongliang, 155  
 Yu, Hui, 27, 28  
 Yu, Mo, 125  
 Yu, Yanchao, 40  
 Yulan, He, 7  
 Yung, Frances, 32  
 Yvon, Fran ois, 24, 25, 89
- Z**  
 Zayats, Victoria, 137  
 Zeng, Daojian, 125  
 Zettlemoyer, Luke, 71, 87, 109, 114, 117, 119, 123  
 Zhai, Feifei, 100  
 Zhang, Biao, 99, 144  
 Zhang, Gongbo, 144  
 Zhang, Jianwen, 52  
 Zhang, Jingwei, 111  
 Zhang, Jingyi, 140  
 Zhang, Lu, 150  
 Zhang, Meishan, 70, 101, 127  
 Zhang, Min, 31, 99, 107  
 Zhang, Peng, 156  
 Zhang, Qi, 60  
 Zhang, Rong, 32  
 Zhang, Tongtao, 51, 72  
 Zhang, Yuan, 127  
 Zhang, Yue, 8, 59, 70, 98, 101, 127, 153  
 Zhang, Yuqi, 25  
 Zhang, Zhiwei, 65  
 Zhao, Dongyan, 64

- Zhao, Jun, 125  
Zhao, Kai, 100  
Zhao, Shi, 78  
Zhao, Shiqi, 7, 125  
Zheng, Jiaping, 34, 65  
Zheng, Jin Guang, 73  
Zhong, Huaping, 52  
Zhong, Jialu, 143  
Zhou, Bowen, 141  
Zhou, Deyu, 130  
Zhou, Guodong, 31, 109, 114  
Zhou, Helen, 29  
Zhou, Lipu, 77  
Zhou, Mengfei, 35  
Zhou, Ming, 85  
Zhou, Yaqian, 127  
Zhu, Chenxi, 76, 98, 127  
Zhu, Fangze, 131  
Zhu, Jun, 112  
Zhu, Kenny, 65  
Zhu, Muhua, 109  
Zhu, Qiaoming, 114  
Zoph, Barret, 85  
Zou, Bowei, 114  
Zuidema, Willem, 97  
Zweigenbaum, Pierre, 33



**WHAT WE**

**INVENT**

CALCULATE

DISPLAY

DEVELOP

REPORT

PROVIDE

OFFER

SUPPLY

ANSWER

WRITE

FASHION

CONCEIVE

ORGANIZE

DESIGN

STREAMLINE

UNCOVER

ENGINEER

CHART

SEE

CLARIFY

EMPOWERS

ENHANCES

SPEEDS

IMPACTS

FUELS

SPARKS

CHANGES

BOOSTS

SPARKS

IMPROVES

ENLIVENS

HELPS

ADVANCES

ENLIGHTENS

FACILITATES

ENHANCES

SOLUTIONS

ANALYSIS

DECISIONS

INVESTMENTS

TRADES

IDEAS

FINANCE

PRODUCTIVITY

EFFICIENCY

**MARKETS**

PRESENTATIONS

COMMUNICATION

LEADERS

COLLABORATION

OPINIONS

OPINIONS

OPINIONS

OPINIONS

**Make  
your  
mark.**

We're looking for smart thinkers in our New York City, San Francisco and London offices! For passionate, talented, ready-for-any challenge professionals, Bloomberg is the place to be.

[jobs.bloomberg.com](http://jobs.bloomberg.com)

**Bloomberg**

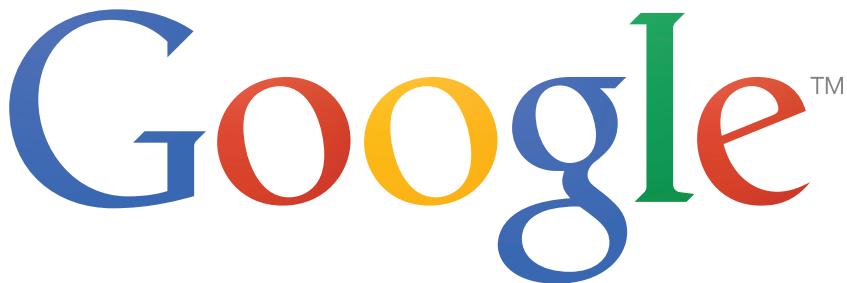


# YOUR SEARCH ENDS HERE

 百度一下

————— 全球最大的中文搜索引擎 ————

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



## Research at Facebook

Giving people the power to share and connect requires constant innovation. At Facebook, we solve technical problems that no one else has seen because no one else has built a social network of this size. Research permeates everything that we do. We believe the most interesting research questions are derived from real world problems. Facebook engineers work on cutting-edge research with a practical focus and push product boundaries every day. With more than 20% of our engineers having PhDs, Facebook has a culture that encourages research activity and values engagement with academia. We offer graduate fellowships to support the research of promising students, establish meaningful research collaborations, and publish at top conferences.



### We are Hiring!

For more information, please visit:  
[research.facebook.com](http://research.facebook.com)

Our engineers work to solve real-world problems that impact millions of people in areas such as:

- Artificial Intelligence
- Compilers
- Computer Vision
- Data Mining
- Databases
- Distributed Systems
- Human-Computer Interaction
- Internet Economics
- Machine Learning
- Natural Language Processing
- Networking
- Operating Systems
- Programming Languages
- Security/Privacy
- Social Computing

**facebook**



## Our mission is simple:

Connect the world's professionals to make them more productive and successful

Learn more at: [linkedin.com/careers](https://linkedin.com/careers)



# RECRUIT

Recruit Technologies Co.,Ltd.



Inventing a more human conversation with technology.

[research.nuance.com](http://research.nuance.com)



The world's brightest technology minds come to [Amazon.com](https://www.amazon.com) to research and develop technology that improves the lives of shoppers, sellers and developers around the world. At Amazon, our Machine Learning team is comprised of technical leaders who develop planet-scale platforms for machine learning on the cloud, assist in the benchmarking and future development of existing machine learning applications across Amazon, and help develop novel and infinitely-scalable applications.



World-class AI Research

Wonderful Sunny Environment

40 Research Staff and Faculty

20 PhD Students

Undergrad & Graduate Interns

Short- and Long-term Visitors

[ai.isi.edu](http://ai.isi.edu)

**Join Us!**

**VoiceBox**  
TECHNOLOGIES

>We Are a Sea  
of Opportunity

Come help advance our  
patented NLU technology  
in growth markets.

voicebox.com  
@voicebox\_tech  
in



CHANGING THE WAY WE TALK TO MACHINES

VocalIQ is building the world's first self-learning dialogue API - putting real, natural conversation between people and their devices.

We're hiring!

Would you like to live in Cambridge, England?  
Do any of these excite you?

DEEP LEARNING   NLP   DIALOGUE SYSTEMS  
SCALABLE SYSTEMS   REINFORCEMENT LEARNING

We'd love to hear from you!  
Visit [vocaliq.com](http://vocaliq.com) for more details.





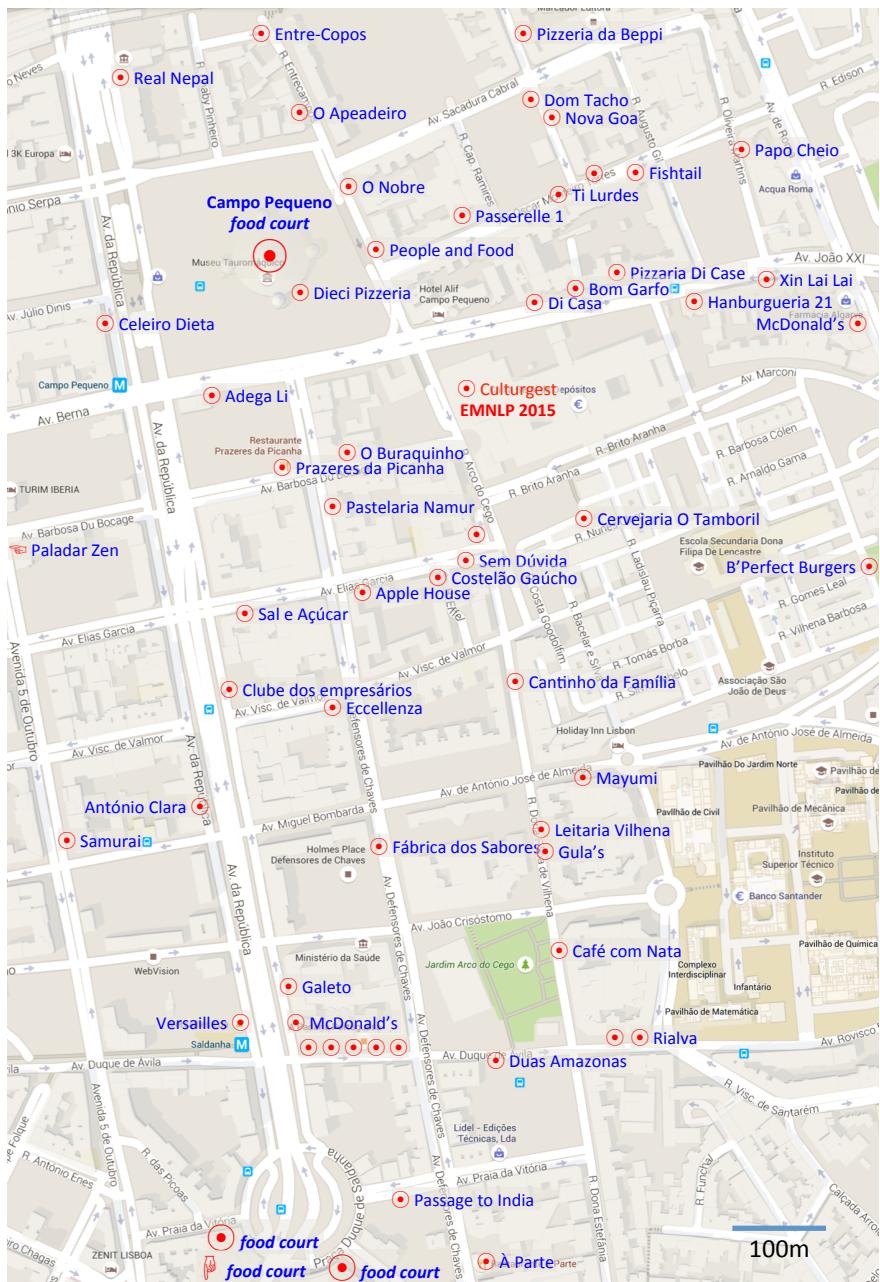
 prüberam

The logo for Prüberam consists of a stylized icon on the left and the brand name on the right. The icon is composed of a series of pinkish-purple dots arranged in a grid-like pattern, with some dots missing, creating a sense of depth or a digital interface. To the right of the icon, the word "prüberam" is written in a large, bold, dark gray sans-serif font. The letters are lowercase, except for the first letter which is capitalized.

## Food Quick Reference

A considerable number of restaurants can be found around the conference venue. The following table shows a small selection of those restaurants and the map that follows presents a more complete selection. *Food courts* can accomodate a large number of people and can provide a rapid service. The closest one is the Campo Pequeno food court. Three additional food courts are located in Saldanha, near the subway. They are located at 15 minutes walking distance or one subway stop away (see bottom of the map).

Type	Establishment	type	Price
<i>Food courts</i>			
	Campo Pequeno	Campo Pequeno	€
	Monumental	Praça Duque Saldanha	€
	Atrium Saldanha	Praça Duque Saldanha	€
	Saldanha Residence	Av. Fontes Pereira de Melo, 42E	€
<i>Vegetarian and vegan</i>			
	Celeiro Dieta	Av. da República, 83	€
	Paladar Zen	Av. Barbosa du Bocage, 107C	€€
<i>Italian</i>			
	Italy Caffé Ristorante	Av. Duque d'Ávila, 26	€€
	Di Casa	Avenida João XXI, 64	€€
<i>Indian</i>			
	Passage to India	Av. Praia da Vitoria, 45	€€
<i>Snack</i>			
	Padaria Portuguesa	Av. Duque de d'Ávila, 24	€
<i>Portuguese</i>			
	À Parte	Av. Defensores de Chaves, 14C	€€
	Vilhena	Rua Dona Filipa de Vilhena, 18	€€
	Charrua do Lavrador	Av. Duque d'Ávila 13	€€
<i>Japanese</i>			
	Mayumi	Av. António José de Almeida 5C	€€
	King Sushi	Av. 5 de Outubro, 124	€€
<i>Fast food</i>			
	McDonald's	Av. Da República, 10F	€
	McDonald's	Av. de Roma, 3	€



EMNLP 2015 gratefully acknowledges the following sponsors for their support:

Platinum

Bloomberg

Baidu 百度

Gold

Google™ facebook LinkedIn™

Silver

RECRUIT  
Recruit Technologies Co.,Ltd.

NUANCE

amazon

AI @ ISI

VoiceBox  
TECHNOLOGIES

DigitalGenius  
Artificial Intelligence for Brands

Bronze

Best Student Paper

priberam

vocaliq  
The dialogue company

IBM Watson

Supporters

Rakuten  
Institute of Technology

Unbabel

LISBOA  
CÂMARA MUNICIPAL



Institutional partners

inescid  
lisboa

TÉCNICO  
LISBOA

instituto de  
telecomunicações

TAP  
TAP PORTUGAL  
de braços abertos. with arms wide open  
A STAR ALLIANCE MEMBER