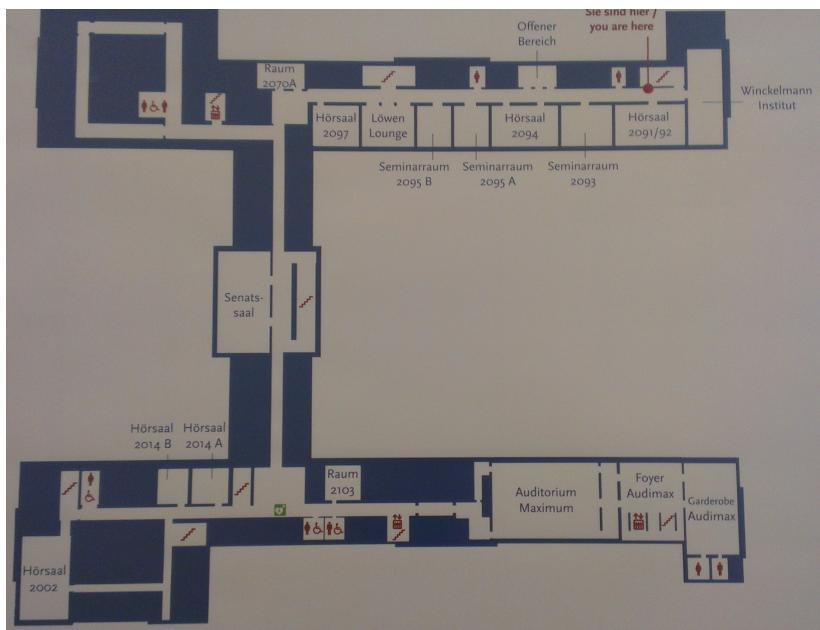


The 14th Conference of the
**North American Chapter of the Association for
Computational Linguistics:**
Human Language Technologies

CONFERENCE HANDBOOK



Cover design by ...
Handbook assembled by Florian Kunneman and Matt Post
Printing by Omnipress of Madison, Wisconsin

Contents

Table of Contents	i
1 Conference Information	1
Message from the General Chair	1
Message from the Program Committee Co-Chairs	2
Organizing Committee	4
2 Tutorials: Sunday, August 7	5
T1: Multimodal Learning and Reasoning	6
T2: NLP Approaches to Computational Argumentation	7
T3: Computer Aided Translation: Advances and Challenges	9
T4: Semantic Representations of Word Senses and Concepts	10
T5: Neural Machine Translation	12
T6: Game Theory and Natural Language: Origin, Evolution and Processing	14
T7: Understanding Short Texts	16
T8: MetaNet: Repository, Identification System, and Applications	18
3 Main Conference: Monday, August 8	19
Session 1	20
Recruitment Lunch	27
Session 2	28
Session 3	36
Poster and dinner session I (includes Demonstrations A and SRW)	47
4 Main Conference: Tuesday, August 9	75
Session 4	76
Session 5	83
Session 6	89
Poster and dinner session II (includes Demonstrations B)	99
Social Event: Boat trip	113

5 Main Conference: Wednesday, August 10	115
Session 7	116
Session 8	120
6 *SEM: Thursday–Friday, August 11–12	127
*SEM 2016: The 5th Joint Conference on Lexical and Computational Semantics	129
7 CoNLL: Thursday–Friday, August 11–12	131
CoNLL 2016: The SIGNLL Conference on Computational Natural Language Learning	132
8 Workshops: Thursday–Friday, August 11–12	135
W1: 1st Conference on Machine Translation	136
W2: 1st Workshop on Representation Learning for NLP	142
W3: 10th Linguistic Annotation Workshop	144
W4: 12th Workshop on Multiword Expressions	146
W5: 7th Workshop on Cognitive Aspects of Computational Language Learning	148
W6: 14th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology	150
W7: 10th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities	152
W8: 5th Workshop on Vision and Language	154
W9: 15th Workshop on Biomedical Natural Language Processing	155
W10: SIGFSM Workshop on Statistical NLP and Weighted Automata	157
W11: 1st Workshop on Evaluating Vector-Space Representations for NLP	158
W12: 10th Web as Corpus Workshop	160
W13: 6th NEWS Named Entities Workshop	161
W14: 3rd Workshop on Argument Mining	162
Author Index	165
9 Local Guide	165
Transportation	165
Airports	165
Transport from airports to city	165
Train Transport	166
City transport	166
Bicycles	167
Food	167
Weather	168
Safety	168

Conference Information

Message from the General Chair

Welcome to Berlin, welcome to ACL-2016!

A special welcome if this is your first ACL - I hope it fulfills your expectations and remains in your memory as a great start. Its magnitude may be a bit overwhelming - our field is on an expanding trajectory, and even a selection of the best work fills a great number of parallel sessions over a number of days; plus, there are the workshops to quench your topical thirst. Just dive in and drink it up.

If it is your first ACL, then take a good look occasionally at the people in your sessions, listening to your talk or grilling you at your poster; many of these people you will get to know over the years as you grow with them and go to the same meetings in the future. The ACL is not only a breeding ground for project ideas and future collaborations, but is also a hunting ground where job seekers find potential employers. Our new Recruitment Lunch was set up to foster this. Last but not least, ACLs have proven to be the starting moment for lasting friendships. Make the most of it!

To all others, welcome back to another great edition of ACL. The programme chairs and the chair teams for the demos, tutorials, and workshops have put together an exciting programme; the local chairs have fitted this all masterfully in Humboldt University's buildings at Unter den Linden in the very heart of Berlin. Cast in a quote in the marble entry hall, Karl Marx encourages us to use our insights and technologies to make the world a better place. This tantalizing encouragement is worth pondering over.

On behalf of the entire ACL-2016 organizing team, welcome to Berlin; I wish you a fruitful conference!

Antal van den Bosch
General Chair

Message from the Program Committee Co-Chairs

Welcome to the 54th Annual Meeting of the Association for Computational Linguistics! This year, ACL received 825 long paper submissions (a new record) and 463 short paper submissions.¹ Of the long papers, 231 were accepted for presentation at ACL—116 as oral presentations and 115 as poster presentations. 97 short papers were accepted—50 as oral and 47 as poster presentations. In addition, ACL also features 25 presentations of papers accepted in the *Transactions of the Association for Computational Linguistics* (TACL). With 353 papers being presented, this is the largest ACL program to date.

In keeping with the tremendous growth of our field, we introduced some changes to the conference. First, oral presentations were shortened to fifteen (twelve) minutes for long (short) papers, plus time for questions. While this places a greater demand on speakers to be concise, we believe it is worth the effort, allowing far more work to be presented orally. Second, we took advantage of the many halls available at Humboldt University and expanded the number of parallel talks during some conference sessions.

We introduced a category of outstanding papers to help recognize the highest quality work in the community this year. The 11 outstanding papers (9 long, 2 short, 0.85% of submissions) represent a broad spectrum of exciting contributions; they are recognized by especially prominent placement in the program. From these, a best paper and an IBM-sponsored best student paper have been selected; those will be announced in the awards session on Wednesday afternoon.

Following other recent ACL conferences, submissions were reviewed under different categories and using different review forms for empirical/data-driven, theoretical, applications/tools, resources/evaluation, and survey papers. We introduced special fields in the paper submission form for authors to explicitly note the release of open-source implementations to enable reproducibility, and to note freely available datasets. We also allowed authors to submit appendices of arbitrary length for details that would enable replication; reviewers were not expected to read this material.

Another innovation we explored during the review period was the scheduling of short paper review before long paper review. While this was planned to make the entire review period more compact (fitting between the constraints of NAACL 2016 and EMNLP 2016 at either end), we found that reviewing short papers first eliminated many of the surprises for the long paper review process.

We sought to follow recently-evolved best practices in planning the poster sessions, so that the many high-quality works presented in that format will be visible and authors and attendees benefit from the interactions during the two poster sessions.

ACL 2016 will have two distinguished invited speakers: Amber Boydston (Associate Professor of Political Science at the University of California, Davis) and Mark Steedman (Professor of Cognitive Science at the University of Edinburgh). We are grateful that they accepted our invitation and look forward to their presentations.

There are many individuals we wish to thank for their contributions to ACL 2016, some multiple times:

- The 38 area chairs who recruited reviewers, led the discussion about each paper, carefully assessed each submission, and authored meta-reviews to guide final decisions: Miguel Ballesteros, David Bamman, Steven Bethard, Jonathan Berant, Gemma Boleda, Ming-Wei Chang, Wanxiang Che, Chris Dyer, Ed Grefenstette, Hannaneh Hajishirzi, Minlie Huang, Mans Hulden, Heng Ji, Jing Jiang, Zornitsa Kozareva, Marco Kuhlmann, Yang Liu, Annie Louis, Wei Lu, Marie-Catherine de Marneffe, Gerard de Melo, David Mimno, Meg Mitchell, Daichi Mochihashi, Graham Neubig, Naoaki Okazaki, Simone Ponzetto,

¹These numbers exclude papers that were not reviewed due to formatting violations or that were withdrawn prior to review.

Matthew Purver, David Reitter, Nathan Schneider, Hinrich Schuetze, Thamar Solorio, Lucia Specia, Partha Talukdar, Ivan Titov, Lu Wang, Nianwen Xue, and Grace Yang.

- Our full program committee of 884 hard-working individuals who reviewed the conference's 1,288 submissions (including secondary reviewers).
- The ACL coordinating committee members, especially Yejin Choi, Graeme Hirst, Chris Manning, and Shiqi Zhao, who answered many questions as they arose during the year.
- TACL editors-in-chief Mark Johnson, Lillian Lee, and Kristina Toutanova, for coordinating with us on TACL presentations at ACL.
- Ani Nenkova and Owen Rambow, program co-chairs of NAACL 2016, and Michael Strube, program co-chair of ACL 2015, who were generous with advice.
- Yusuke Miyao, Yannick Versley, and Hai Zhao, our well-organized publication chairs, and the responsive team at Softconf led by Rich Gerber.
- Valia Kordoni and the local organization team, especially webmaster Kostadin Cholakov.
- Antal van den Bosch, our general chair, who kept us coordinated with the rest of the ACL 2016 team and offered guidance whenever we needed it.
- Antal van den Bosch, Claire Cardie, Pascale Fung, Ray Mooney, and Joakim Nivre, who carefully reviewed papers under consideration for outstanding and best paper recognition.
- Priscilla Rasmussen, who knows everything about how to make ACL a success.

We hope that you enjoy ACL 2016 in Berlin!

Katrin Erk, University of Texas
Noah A. Smith, University of Washington

Program Committee Co-Chairs

Organizing Committee

General Chair

Antal van den Bosch, Radboud University Nijmegen

Program Committee Co-chairs

Katrin Erk, University of Texas at Austin

Noah Smith, University of Washington

Local Organizing Committee

Valia Kordonis (local chair), Humboldt University

Markus Egg (local co-chair), Humboldt University

Kostadin Cholakov, Humboldt University

Maja Popović, Humboldt University

Manfred Stede, University of Potsdam

Anke Lüdeling, Humboldt University

Manfred Krifka, Humboldt University

Ulf Leser, Humboldt University

Workshop Co-chairs

Jun Zhao, Chinese Academy of Sciences

Sabine Schulte im Walde, IMS Stuttgart

Tutorial Co-chairs

Alexandra Birch, University of Edinburgh

Jelle Zuidema, University of Amsterdam

Publication Co-chairs

Yannick Versley, Ruprecht-Karls-Universität Heidelberg

Hai Zhao, Shanghai Jiao Tong University

Yusuke Miyao, National Institute of Informatics, Japan

Demo Co-chairs

Sameer Pradhan, Boulder Learning

Marianna Apidianaki, LIMSI-CNRS

Student Research Workshop Co-chairs

Will Roberts, Humboldt-Universität zu Berlin

Tao Lei, Massachusetts Institute of Technology

He He, University of Maryland

Faculty Advisors to the Student Research Workshop

Yang Liu, Tsinghua University

Chris Biemann, Technische Universität Darmstadt

Gosse Bouma, University of Groningen

Student Volunteer Coordinator

Yulia Grishina, University of Potsdam

Publicity Chair

Barbara Plank, University of Groningen

Business Manager

Priscilla Rasmussen

Conference Handbook Editor

Florian Kunneman, Radboud University Nijmegen

Professional Conference Organiser

Con gressa GmbH, Berlin

2

Tutorials: Sunday, August 7

Overview

7:30 – 6:00	Registration	(TBD)
7:30 – 9:00	Breakfast	(TBD)
9:00 – 12:30	Morning Tutorials	
	tutorial-001	(Governor's Square 15)
	tutorial-001	
	tutorial-002	(Governor's Square 17)
	tutorial-002	
	tutorial-003	(Governor's Square 16)
	tutorial-003	
	tutorial-004	(Governor's Square 17)
	tutorial-004	
12:30 – 2:00	Lunch break	
2:00 – 5:30	Afternoon Tutorials	
	tutorial-005	(Governor's Square 15)
	tutorial-005	
	tutorial-006	(Governor's Square 15)
	tutorial-006	
	tutorial-007	(Governor's Square 15)
	tutorial-007	
	tutorial-008	(Governor's Square 15)
	tutorial-008	

Tutorial 1

tutorial-001

tutorial-001

Sunday, August 7, 2016, 9:00–12:30pm

Governor's Square 15

Natural Language Processing has broadened in scope to tackle more and more challenging language understanding and reasoning tasks. The core NLP tasks remain predominantly unimodal, focusing on linguistic input, despite the fact that we, humans, acquire and use language while communicating in perceptually rich environments. Moving towards human-level AI will require the integration and modeling of multiple modalities beyond language. With this tutorial, our aim is to introduce researchers to the areas of NLP that have dealt with multimodal signals. The key advantage of using multimodal signals in NLP tasks is the complementarity of the data in different modalities. For example, we are less likely to find descriptions of yellow bananas or wooden chairs in text corpora, but these visual attributes can be readily extracted directly from images. Multimodal signals, such as visual, auditory or olfactory data, have proven useful for models of word similarity and relatedness, automatic image and video description, and even predicting the associated smells of words. Finally, multimodality offers a practical opportunity to study and apply multitask learning, a general machine learning paradigm that improves generalization performance of a task by using training signals of other related tasks.

Desmond Elliott is a postdoc at the Institute for Logic, Language and Computation in the University of Amsterdam (The Netherlands). His main research interests are models and evaluation methods for automatic image description. He delivered a tutorial on Datasets and Evaluation Methods for Image Description at the 2015 Integrating Vision and Language Summer School, and is co-organising a shared task on Multimodal Machine Translation at the 2016 Workshop on Machine Translation.

Douwe Kiela is a final year PhD student at the University of Cambridge's Computer Laboratory, supervised by Stephen Clark. He is interested in trying to enrich NLP with additional resources, primarily through grounding representations in perceptual modalities including vision, but also auditory and even olfactory modalities. He is a student board member of EACL and has published 8 top-tier conference papers over the three years of his PhD.

Angeliki Lazaridou is a final year PhD student, supervised by Marco Baroni at the Center for Mind/Brain Sciences of the University of Trento (Italy). Her primary research interests are in the area of multimodal semantics, i.e., making purely text-based models of meaning interact with other modalities, such as visual and sensorimotor. She has focused on learning models with multimodal signals and using those for multimodal inference, work that has appeared at related venues (ACL, NAACL, TACL, EMNLP).

Tutorial 2

tutorial-002

tutorial-002

Sunday, August 7, 2016, 9:00–12:30pm

Governor's Square 17

Argumentation and debating represent primary intellectual activities of the human mind. People in all societies argue and debate, not only to convince others of their own opinions but also in order to explore the differences between multiple perspectives and conceptualizations, and to learn from this exploration. The process of reaching a resolution on controversial topics typically does not follow a simple sequence of purely logical steps. Rather it involves a wide variety of complex and interwoven actions. Presumably, pros and cons are identified, considered, and weighed, via cognitive processes that often involve persuasion and emotions, which are inherently harder to formalize from a computational perspective.

This wide range of conceptual capabilities and activities, have only in part been studied in fields like CL and NLP, and typically within relatively small sub-communities that overlap the ACL audience. The new field of Computational Argumentation has very recently seen significant expansion within the CL and NLP community as new techniques and datasets start to become available, allowing for the first time investigation of the computational aspects of human argumentation in a holistic manner.

The main goal of this tutorial would be to introduce this rapidly evolving field to the CL community. Specifically, we will aim to review recent advances in the field and to outline the challenging research questions - that are most relevant to the ACL audience - that naturally arise when trying to model human argumentation.

We will further emphasize the practical value of this line of study, by considering real-world CL and NLP applications that are expected to emerge from this research, and to impact various industries, including legal, finance, healthcare, media, and education, to name just a few examples.

The first part of the tutorial will provide introduction to the basics of argumentation and rhetoric. Next, we will cover fundamental analysis tasks in Computational Argumentation, including argument mining, revealing argument relations, assessing arguments quality, stance classification, polarity analysis, and more. After the coffee break, we will first review existing resources and recently introduced benchmark data. In the following part we will cover basic synthesis tasks in Computational Argumentation, including the relation to NLG and dialogue systems, and the evolving area of Debate Technologies, defined as technologies developed directly to enhance, support, and engage with human debating. Finally, we will present relevant demos, review potential applications, and discuss the future of this emerging field.

Iryna Gurevych is Full Professor of Natural Language Processing in the Computer Science Department of the Technische Universität (TU) Darmstadt. She heads the Ubiquitous Knowledge Processing (UKP) Lab and the Research Training Group ‘Adaptive Information Preparation from Heterogeneous Sources’ at the TU Darmstadt and Univ. of Heidelberg. Iryna has been working in the areas of lexical-semantic processing, discourse analysis and text mining with applications in social sciences and humanities for about 15 years. Her research interests have recently shifted towards NLP approaches to argumentation analysis. She is supervising several students working in this area and has published on this topic in the recent EMNLP, COLING and other conferences. Iryna has been co-organizer of the NAACL 2015 workshop on Argumentation Mining and the Dagstuhl seminar “Debating Technologies” in December 2015, she will co-organize the ACL 2016 workshop on Argument Mining in Berlin and co-chair the Debating Technologies track.

Chris Reed is Full Professor of Computer Science and Philosophy at the University of Dundee in Scotland, where he heads the Centre for Argument Technology. Chris has been working at the overlap between argumentation theory and artificial intelligence for over twenty years, has won over \$5.6m of funding from government and commercial sources and has over 140 peer-reviewed papers in the area including five books. He has also been instrumental in the development of the Argument Interchange Format, an international standard for computational work in the area; he is spear-heading the major engineering effort behind the Argument Web; and he is a founding editor of the Journal of Argument & Computation. He was co-organiser of COMMA 2014, of the first ACL workshop on Argumentation Mining in 2014, will be chair of the third workshop on Argument Mining with ACL in 2016, and has recently won funding for a \$1m project on the topic in collaboration with IBM.

Noam Slonim is a Senior Technical Staff Member (STSM) at IBM Research. He serves as the IBM Research Technical Lead of topics related to debate technologies, leading a team of more than 40 researchers from several different IBM Research Labs around the world. His main research interest is in developing innovative applications that can enhance, support, and engage with human debating. Correspondingly, he is actively pursuing specific research questions around that area, mainly in the context of developing advanced text analysis applications. From a theoretical perspective, his research over the years has led to the development of various Machine Learning techniques for the analysis of textual and genomic data, that often stem from Information Theoretic concepts and algorithms. Noam has recently published at ACL, EMNLP, and COLING.

Benno Stein is chair of the Web Technology and Information Systems Group at the Bauhaus-Universität Weimar. His research focuses on theory and algorithms for information retrieval, information extraction, and data mining. Since 2008 his group applies Big Data technology for research related to retrieval and extraction tasks. Among others, the group has developed effective and efficient methods for cross-language text reuse, document clustering, paraphrasing, and sentiment analysis. Benno is chair of the international workshop series TIR on Text-based Information Retrieval, initiator and co-organizer of PAN, a research network dedicated to digital text forensics. He is co-founder and spokesman of the Digital Bauhaus Lab, a forthcoming interdisciplinary research center with advanced working environments for research in Computer Science, Engineering, and Media Art and Design.

Tutorial 3

tutorial-003

tutorial-003

Sunday, August 7, 2016, 9:00–12:30pm

Governor's Square 16

Moving beyond post-editing machine translation, a number of recent research efforts have advanced computer aided translation methods that allow for more interactivity, richer information such as confidence scores, and the completed feedback loop of instant adaptation of machine translation models to user translations.

This tutorial will explain the main techniques for several aspects of computer aided translation:

- confidence measures
- interactive machine translation (interactive translation prediction)
- bilingual concordancers
- translation option display
- paraphrasing (alternative translation suggestions)
- visualization of word alignment
- online adaptation
- automatic reviewing
- integration of translation memory
- eye tracking, logging, and cognitive user models

For each of these, the state of the art and open challenges are presented. The tutorial will also look under the hood of the open source CASMACAT toolkit that is based on MATECAT, and available as a “Home Edition” to be installed on a desktop machine. The target audience of this tutorials are researchers interested in computer aided machine translation and practitioners who want to use or deploy advanced CAT technology.

Philip Koehn (John Hopkins University) is a leading researcher in statistical machine translation and co-ordinated the EU-funded CASMACAT project on computer aided translation 2011–2014.

Tutorial 4

tutorial-004

tutorial-004

Sunday, August 7, 2016, 9:00–12:30pm

Governor's Square 17

Representing the semantics of linguistic items in a machine-interpretable form has been a major goal of Natural Language Processing since its earliest days. Among the range of different linguistic items, words have attracted the most research attention. However, word representations have an important limitation: they conflate different meanings of a word into a single vector. Representations of word senses have the potential to overcome this inherent limitation. Indeed, the representation of individual word senses and concepts has recently gained in popularity with several experimental results showing that a considerable performance improvement can be achieved across different NLP applications upon moving from word level to the deeper sense and concept levels. Another interesting point regarding the representation of concepts and word senses is that these models can be seamlessly applied to other linguistic items, such as words, phrases, sentences, etc.

This tutorial will first provide a brief overview of the recent literature concerning word representation (both count based and neural network based). It will then describe the advantages of moving from the word level to the deeper level of word senses and concepts, providing an extensive review of state-of-the-art systems. Approaches covered will not only include those which draw upon knowledge resources such as WordNet, Wikipedia, BabelNet or FreeBase as reference, but also the so-called multi-prototype approaches which learn sense distinctions by using different clustering techniques. Our tutorial will discuss the advantages and potential limitations of all approaches, showing their most successful applications to date. We will conclude by presenting current open problems and lines of future work.

José Camacho Collados is a Google Doctoral Fellow and PhD student at the Sapienza University of Rome, working under the supervision of Prof. Roberto Navigli. His research focuses on Natural Language Processing and on the area of lexical semantics in particular. He has developed NASARI, a novel semantic vector representation for concepts and named entities, which led to two publications in NAACL and ACL 2015. José will co-organize a SemEval 2017 task on multilingual semantic similarity. His background education includes an Erasmus Mundus Master in Natural Language Processing and Human Language Technology and a 5-year BSc degree in Mathematics.

Ignacio Iacobacci is a PhD student at the Sapienza University of Rome, working under the supervision of Prof. Roberto Navigli. His research interests lie in the fields of Machine Learning, Natural Language Processing, Neural Networks. He is currently working on Word Sense Disambiguation and Distributional Semantics. Ignacio presented SensEmbed at ACL 2015, a novel approach for word and relational similarity built from exploiting semantic knowledge for modeling arbitrary word senses in a large sense inventory. His background includes a MSc. in Computer Science and 8 years as a developer including 4 years as a Machine Learning - NLP specialist.

Roberto Navigli is an Associate Professor in the Department of Computer Science at La Sapienza University of Rome and a member of the Linguistic Computing Laboratory. His research interests lie in the field of Natural Language Processing, including: Word Sense Disambiguation and Induction, Ontology Learning, Knowledge Representation and Acquisition, and multilinguality. In 2007 he received a Ph.D. in Computer Science from La Sapienza and he was awarded the Marco Cadoli 2007 AI*IA national prize for the Best Ph.D. Thesis in Artificial Intelligence. In 2013 he received the Marco Somalvico AI*IA prize, awarded every two years to the best young Italian researcher in Artificial Intelligence. He is the creator and founder of BabelNet, both a multilingual encyclopedic dictionary and a semantic network, and its related project Babelfy, a state-of-the-art multilingual disambiguation and entity linking system. He is also the Principal Investigator of MultiJEDI, a 1.3M euro 5-year Starting Grant funded by the European Research Council and the responsible person of the Sapienza unit in LIDER, an EU project on content analytics and language technologies. He is also the Co-PI of “Language Understanding cum Knowledge Yield” (LUcKY), a Google Focused Research Award on Natural Language Understanding.

Mohammad Taher Pilehvar is a Research Associate in the Language Technology Lab of the University of Cambridge where he is currently working on NLP in the biomedical domain. Taher completed his PhD in 2015 under the supervision of Prof. Roberto Navigli. Taher's research lies in lexical semantics, mainly focusing on semantic representation, semantic similarity, and Word Sense Disambiguation. He has co-organized two semeval tasks and has authored multiple conference and journal papers on semantic representation and similarity in top tier venues. He is the first author of a paper on semantic similarity that was nominated for the best paper award at ACL 2013.

Tutorial 5

tutorial-005

tutorial-005

Sunday, August 7, 2016, 2:00–5:30pm

Governor's Square 15

Neural Machine Translation (NMT) is a simple new architecture for getting machines to learn to translate. Despite being relatively new (Kalchbrenner and Blunsom, 2013; Cho et al., 2014; Sutskever et al., 2014), NMT has already shown promising results, achieving state-of-the-art performances for various language pairs (Luong et al., 2015a; Jean et al., 2015; Luong et al., 2015b; Sennrich et al., 2016; Luong and Manning, 2016). While many of these NMT papers were presented to the ACL community, research and practice of NMT are only at their beginning stage. This tutorial would be a great opportunity for the whole community of machine translation and natural language processing to learn more about a very promising new approach to MT. This tutorial has four parts.

In the first part, we start with an overview of MT approaches, including: (a) traditional methods that have been dominant over the past twenty years and (b) recent hybrid models with the use of neural network components. From these, we motivate why an end-to-end approach like neural machine translation is needed. The second part introduces a basic instance of NMT. We start out with a discussion of recurrent neural networks, including the back-propagation-through-time algorithm and stochastic gradient descent optimizers, as these are the foundation on which NMT builds. We then describe in detail the basic sequence-to-sequence architecture of NMT (Cho et al., 2014; Sutskever et al., 2014), the maximum likelihood training approach, and a simple beam-search decoder to produce translations.

The third part of our tutorial describes techniques to build state-of-the-art NMT. We start with approaches to extend the vocabulary coverage of NMT (Luong et al., 2015a; Jean et al., 2015; Chitnis and DeNero, 2015). We then introduce the idea of jointly learning both translations and alignments through an attention mechanism (Bahdanau et al., 2015); other variants of attention (Luong et al., 2015b; Tu et al., 2016) are discussed too. We describe a recent trend in NMT, that is to translate at the sub-word level (Chung et al., 2016; Luong and Manning, 2016; Sennrich et al., 2016), so that language variations can be effectively handled. We then give tips on training and testing NMT systems such as batching and ensembling. In the final part of the tutorial, we briefly describe promising approaches, such as (a) how to combine multiple tasks to help translation (Dong et al., 2015; Luong et al., 2016; Firat et al., 2016; Zoph and Knight, 2016) and (b) how to utilize monolingual corpora (Sennrich et al., 2016). Lastly, we conclude with challenges remained to be solved for future NMT.

PS: we would also like to acknowledge the very first paper by Forcada and Neco (1997) on sequence-to-sequence models for translation!

Thang Luong is currently a 5th-year PhD student in the Stanford NLP group under Prof. Christopher Manning. In the past, he has published papers on various different NLP-related areas such as digital library, machine translation, speech recognition, parsing, psycholinguistics, and word embedding learning. Recently, his main interest shifts towards the area of deep learning using sequence to sequence models to tackle various NLP problems, especially neural machine translation. He has built state-of-the-art (academically) neural machine translation systems both at Google and at Stanford.

Kyunghyun Cho is an assistant professor in the Department of Computer Science and the Center for Data Science at New York University. He has worked in deep learning for natural language processing, language translation, image captioning, and a variety of other subjects including the core methods of deep learning. He completed a post-doctoral fellowship at the Montreal Institute for Learning Algorithms. He earned a Ph.D and an M.Sc (with distinction) from the Aalto University School of Science, and a B.Sc in Computer Science from the Korea Advanced Institute of Science and Technology.

Christopher Manning is a professor of computer science and linguistics at Stanford University. He works on software that can intelligently process, understand, and generate human language material. He is a leader in applying Deep Learning to Natural Language Processing, including exploring Tree Recursive Neural Networks, sentiment analysis, neural network dependency parsing, the GloVe model of word vectors, neural machine translation, and deep language understanding. Manning is an ACM Fellow, a AAAI Fellow, and an ACL Fellow, and has coauthored leading textbooks on statistical natural language processing and information retrieval. He is a member of the Stanford NLP group (@stanfordnlp).

Tutorial 6

tutorial-006

tutorial-006

Sunday, August 7, 2016, 2:00–5:30pm

Governor's Square 16

The development of game theory in the early 1940's by John von Neumann was a reaction against the then dominant view that problems in economic theory can be formulated using standard methods from optimization theory. Indeed, most real-world economic problems involve conflicting interactions among decision-making agents that cannot be adequately captured by a single (global) objective function. The main idea behind game theory is to shift the emphasis from optimality criteria to equilibrium conditions. Game theory provides a framework to model complex scenarios, with applications in economics and social science but also in different fields of information technology. With the recent development of algorithmic game theory, it has been used to solve problems in computer vision, pattern recognition, machine learning and natural language processing.

Game-theoretic frameworks have been used in different ways to study language origin and evolution. Furthermore, the so-called game metaphor has been used by philosophers and linguists to explain how language evolved and how it works. Ludwig Wittgenstein, for example, famously introduced the concept of a language game to explain the conventional nature of language, and put forward the idea of the spontaneous formation of a common language that gradually emerges from the interactions among the speakers within a population.

This concept opens the way to the interpretation of language as a complex adaptive system composed of linguistic units and their interactions, which gives rise to the emergence of structural properties. It is the core part of many computational models of language that are based on classical game theory and evolutionary game theory. With the former it is possible to model how speakers form a signaling system in which the ambiguity of the symbols is minimized; with the latter it is possible to model how speakers coordinate their linguistic choices according to the satisfaction that they have about the outcome of a communication act, converging to a common language. In the same vein, many other attempts have been proposed to explain how other characteristics of language follow similar dynamics.

Game theory, and in particular evolutionary game theory, thanks to their ability to model interactive situations and to integrate information from multiple sources, have also been used to solve specific problems in natural language processing and information retrieval, such as language generation, word sense disambiguation and document and text clustering.

The goal of this tutorial is to offer an introduction to the basic concepts of game theory and to show its main applications in the study of language, from different perspectives. We shall assume no pre-existing knowledge of game theory by the audience, thereby making the tutorial self-contained and understandable by a non-expert.

Marcello Pelillo is Full Professor of Computer Science at Ca' Foscari University, where he directs the European Centre for Living Technology (ECLT). He held visiting research positions at Yale University, McGill University, the University of Vienna, York University (UK), the University College London, and the National ICT Australia (NICTA). He has published more than 200 technical papers in refereed journals, handbooks, and conference proceedings in the areas of machine learning, pattern recognition and computer vision. He has initiated several conference series, including EMMCVPR in 1997, IWCV in 2008, SIMBAD in 2011, and he chairs the EMMCVPR and SIMBAD steering committees. He serves (has served) on the Editorial Boards of the journals IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), IET Computer Vision, Pattern Recognition, Brain Informatics, and he serves on the Advisory Board of the International Journal of Machine Learning and Cybernetics. He has served (serves) as Guest Editor for various special issues of IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Pattern Recognition, Pattern Recognition Letters. He is (or has been) scientific coordinator of several research projects, including SIMBAD, an EU-FP7 project devoted to similarity-based pattern analysis and recognition whose activity is described in a recently published Springer book, and he has recently won an award from the Samsung Global Research Outreach (GRO) program. Prof. Pelillo is a Fellow of the IEEE and a Fellow of the IAPR. He has recently been appointed IEEE SMC Distinguished Lecturer.

Rocco Tripodi is Post-doctoral researcher at European Centre for Living Technology (Ca' Foscari University of Venice), where he is working on learning models based on game theoretic principles. He completed his PhD in Computer Science at Ca' Foscari University in 2015 with a thesis titled "Evolutionary Game Theoretic Models for Natural Language Processing". He was Research Assistant and Adjunct Professor at Ca' Foscari University in 2010, where he worked on ontological semantics and taught Corpus Linguistics and Natural Language Processing for Online Applications. In 2015, he was visiting researcher at Staffordshire University (UK). His research interests are in the areas of machine learning and natural language processing. He is particularly interested in classification problems on networked data, from a semantic perspective and on the design, learning and evolution of linguistic communication systems. At present, he is working on game theoretic learning models, with applications to the analysis of multi-modal signals from interacting agents and on group dynamics.

Tutorial 7

tutorial-007

tutorial-007

Sunday, August 7, 2016, 2:00–5:30pm

Billions of short texts are produced every day, in the form of search queries, ad keywords, tags, tweets, messenger conversations, social network posts, etc. Unlike documents, short texts have some unique characteristics which make them difficult to handle. First, short texts, especially search queries, do not always observe the syntax of a written language. This means traditional NLP techniques, such as syntactic parsing, do not always apply to short texts. Second, short texts contain limited context. The majority of search queries contain less than 5 words, and tweets can have no more than 140 characters. Because of the above reasons, short texts give rise to a significant amount of ambiguity, which makes them extremely difficult to handle. On the other hand, many applications, including search engines, ads, automatic question answering, online advertising, recommendation systems, etc., rely on short text understanding. In all these applications, the necessary first step is to transform an input text into a machine-interpretable representation, namely to “understand” the short text. A growing number of approaches leverage external knowledge to address the issue of inadequate contextual information that accompanies the short texts. These approaches can be classified into two categories: Explicit Representation Model (ERM) and Implicit Representation Model (IRM). In this tutorial, we will present a comprehensive overview of short text understanding based on explicit semantics (knowledge graph representation, acquisition, and reasoning) and implicit semantics (embedding and deep learning). Specifically, we will go over various techniques in knowledge acquisition, representation, and inferencing has been proposed for text understanding, and we will describe massive structured and semi-structured data that have been made available in the recent decade that directly or indirectly encode human knowledge, turning the knowledge representation problems into a computational grand challenge with feasible solutions insight.

Zhongyuan Wang is a Researcher at Microsoft Research Asia (MSRA). He leads two projects at MSRA: Enterprise Dictionary (knowledge mining from Enterprise) and Probbase (knowledge mining from Web). He got his Ph.D. degree in computer science from Renmin University of China, and his PhD thesis is ‘Short Text Understanding’. Zhongyuan Wang has published 20+ papers (including ICDE 2015 Best Paper Award on short text understanding) in the leading international conferences, such as VLDB, ICDE, IJCAI, CIKM, etc. He is also the co-author of the book ‘Web Data Management: Concepts and Techniques’, published in 2014. His research interests include knowledge base, natural language processing, semantic network, machine learning, and web data mining.

Haixun Wang is a research scientist / Engineering manager at Facebook. Before Facebook, he was with Google Research, working on natural language processing. He led research in semantic search, graph data processing systems, and distributed query processing at Microsoft Research Asia. He had been a research staff member at IBM T. J. Watson Research Center from 2000 - 2009. He was Technical Assistant to Stuart Feldman (Vice President of Computer Science of IBM Research) from 2006 to 2007, and Technical Assistant to Mark Wegman (Head of Computer Science of IBM Research) from 2007 to 2009. He received the Ph.D. degree in computer science from the University of California, Los Angeles in 2000. He has published more than 150 research papers in refereed international journals and conference proceedings. He served as PC Chair of conferences such as CIKM12 and he is on the editorial board of IEEE Transactions of Knowledge and Data Engineering (TKDE), and Journal of Computer Science and Technology (JCST). He won the best paper award in ICDE 2015, 10 year best paper award in ICDM 2013, and best paper award of ER 2009.

Tutorial 8

tutorial-008

tutorial-008

Sunday, August 7, 2016, 2:00–5:30pm

Governor's Square 16

The ubiquity of metaphor in language (Lakoff and Johnson 1980) has served as impetus for cognitive linguistic approaches to the study of language, mind, and the study of mind (e.g. Thibodeau & Boroditsky 2011). While native speakers use metaphor naturally and easily, the treatment and interpretation of metaphor in computational systems remains challenging because such systems have not succeeded in developing ways to recognize the semantic elements that define metaphor. This tutorial demonstrates MetaNet's frame-based semantic analyses, and their informing of MetaNet's automatic metaphor identification system. Participants will gain a complete understanding of the theoretical basis and the practical workings of MetaNet, and acquire relevant information about the Frame Semantics basis of that knowledge base and the way that FrameNet handles the widespread phenomenon of metaphor in language. The tutorial is geared to researchers and practitioners of language technology, not necessarily experts in metaphor analysis or knowledgeable about either FrameNet or MetaNet, but who are interested in natural language processing tasks that involve automatic metaphor processing, or could benefit from exposure to tools and resources that support frame-based deep semantic, analyses of language, including metaphor as a widespread phenomenon in human language.

Miriam R. L. Petrucc 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.

Ellen K. Dodge received her PhD in Linguistics from the University of California, Berkeley. Since 2000, she has worked in ICSI's AI Group, first as part of the Neural Theory of Language project, and also FrameNet. She is the primary linguist continuing to develop Embodied Construction Grammar. Since 2012, she has worked on MetaNet, developing formal representations of frame and metaphor networks, as well as automatic methods to identify and analyze metaphors in text.

3

Main Conference: Monday, August 8

Overview

8:45–9:00	Opening session	<i>Plaza Ballroom A & B</i>				
9:00–10:10	Invited talk I: Amber Boydston	<i>Plaza Ballroom A, B, & C</i>				
10:10–10:40	Coffee break	<i>Plaza Ballroom D & E</i>				
Session 1						
10:40–12:00	Semantic parsing I	Information extraction	Machine translation I	Word meaning I	Parsing I	Non-compositionality
12:00–1:40	Lunch break					
12:30–2:30	Recruitment Lunch					
Session 2						
1:40–3:00	Word vectors I	Events and schemas	Sentiment analysis	Parsing II	Information retrieval	Phonology and morphology
3:00–3:30	Coffee break					
Session 3						
3:30–5:10	Question answering I	Sentence vectors	Parsing III	Dialog	Generation	Entities and coreference
6:00–9:00	Poster and dinner session I (includes Demonstrations A and SRW)					
	<i>Plaza Ballroom A, B, & C</i>					

Session 1 Overview – Monday, August 8, 2016

Track A <i>Semantic parsing I</i>	Track B <i>Information extraction</i>	Track C <i>Machine translation I</i>	Track D <i>Word meaning I</i>	Track E <i>Parsing I</i>	Track F <i>Non- compositionality</i>
Plaza Ballroom	Plaza Ballroom	Plaza Ballroom	Plaza Ballroom	Plaza Ballroom	Plaza Ballroom F
papers-443 papers-443	papers-008 papers-008	papers-543 papers-543	papers-520 papers-520	short-241 short-241	papers-122 papers-122
papers-586 papers-586	tacl-901 tacl-901	papers-304 papers-304	papers-154 papers-154	papers-500 papers-500	papers-504 papers-504
papers-624 papers-624	papers-266 papers-266	papers-212 papers-212	papers-697 papers-697	papers-431 papers-431	papers-197 papers-197
papers-186 papers-186	tacl-913 tacl-913	papers-357 papers-357	papers-741 papers-741	papers-318 papers-318	papers-828 papers-828

11:40 11:20 11:00 10:40

Parallel Session 1

1A: Semantic parsing I

Plaza Ballroom F

*Chair:***papers-443****papers-443**

10:40–11:00

Semantic parsers map natural language statements into meaning representations, and must abstract over syntactic phenomena, resolve anaphora, and identify word senses to eliminate ambiguous interpretations. Abstract meaning representation (AMR) is a recent example of one such semantic formalism which, similar to a dependency parse, utilizes a graph to represent relationships between concepts (Banarescu et al., 2013). As with dependency parsing, transition-based approaches are a common approach to this problem. However, when trained in the traditional manner these systems are susceptible to the accumulation of errors when they find undesirable states during greedy decoding. Imitation learning algorithms have been shown to help these systems recover from such errors. To effectively use these methods for AMR parsing we find it highly beneficial to introduce two novel extensions: noise reduction and targeted exploration. The former mitigates the noise in the feature representation, a result of the complexity of the task. The latter targets the exploration steps of imitation learning towards areas which are likely to provide the most information in the context of a large action-space. We achieve state-of-the-art results, and improve upon standard transition-based parsing by 4.7 F1 points.

papers-586**papers-586**

11:00–11:20

Modeling crisp logical regularities is crucial in semantic parsing, making it difficult for neural models with no task-specific prior knowledge to achieve good results. In this paper, we introduce data recombination, a novel framework for injecting such prior knowledge into a model. From the training data, we induce a high-precision synchronous context-free grammar, which captures important conditional independence properties commonly found in semantic parsing. We then train a sequence-to-sequence recurrent network (RNN) model with a novel attention-based copying mechanism on datapoints sampled from this grammar, thereby teaching the model about these structural properties. Data recombination improves the accuracy of our RNN model on three semantic parsing datasets, leading to new state-of-the-art performance on the standard GeoQuery dataset for models with comparable supervision.

papers-624**papers-624**

11:20–11:40

A core problem in learning semantic parsers from denotations is picking out consistent logical forms—those that yield the correct denotation—from a combinatorially large space. To control the search space, previous work relied on restricted set of rules, which limits expressivity. In this paper, we consider a much more expressive class of logical forms, and show how to use dynamic programming to efficiently represent the complete set of consistent logical forms. Expressivity also introduces many more spurious logical forms which are consistent with the correct denotation but do not represent the meaning of the utterance. To address this, we generate fictitious worlds and use crowdsourced denotations on these worlds to filter out spurious logical forms. On the WikiTableQuestions dataset, we increase the coverage of answerable questions from 53.5% to 76%, and the additional crowdsourced supervision lets us rule out 92.1% of spurious logical forms.

papers-186**papers-186**

11:40–12:00

Semantic parsing aims at mapping natural language to machine interpretable meaning representations. Traditional approaches rely on high-quality lexicons, manually-built templates, and linguistic features which are either domain- or representation-specific. In this paper we present a general method based on an attention-enhanced encoder-decoder model. We encode input utterances into vector representations, and generate their logical forms by conditioning the output sequences or trees on the encoding vectors. Experimental results on four datasets show that our approach performs competitively without using hand-engineered features and is easy to adapt across domains and meaning representations.

1B: Information extraction

Plaza Ballroom F

*Chair:***papers-008****papers-008**

10:40–11:00

Slot filling aims to extract the values (slot fillers) of specific attributes (slots types) for a given entity (query) from a large-scale corpus. Slot filling remains very challenging over the past seven years. We propose a simple yet effective unsupervised approach to extract slot fillers based on the following two observations: (1) a trigger is usually a salient node relative to the query and filler nodes in the dependency graph of a context sentence; (2) a relation is likely to exist if the query and candidate filler nodes are strongly connected by a relation-specific trigger. Thus we design a graph-based algorithm to automatically identify triggers based on personalized PageRank and Affinity Propagation for a given (query, filler) pair and then label the slot type based on the identified triggers. Our approach achieves 11.6%–25% higher F-score over state-of-the-art English slot filling methods. Our experiments also demonstrate that as long as a few trigger seeds, name tagging and dependency parsing capabilities exist, this approach can be quickly adapted to any language and new slot types. Our promising results on Chinese slot filling can serve as a new benchmark.

tacl-901**tacl-901**

11:00–11:20

We present DefIE, an approach to large-scale Information Extraction (IE) based on a syntactic-semantic analysis of textual definitions. Given a large corpus of definitions we leverage syntactic dependencies to reduce data sparsity, then disambiguate the arguments and content words of the relation strings, and finally exploit the resulting information to organize the acquired relations hierarchically. The output of DefIE is a high-quality knowledge base consisting of several million automatically acquired semantic relations.

papers-266**papers-266**

11:20–11:40

When a large-scale incident or disaster occurs, there is often a great demand for rapidly developing a system to extract detailed and new information from low-resource languages (LLs). We propose a novel approach to discover comparable documents in high-resource languages (HLs), and project Entity Discovery and Linking results from HLs documents back to LLs. We leverage a wide variety of language-independent forms from multiple data modalities, including image processing (image-to-image retrieval, visual similarity and face recognition) and sound matching. We also propose novel methods to learn entity priors from a large-scale HL corpus and knowledge base. Using Hausa and Chinese as the LLs and English as the HL, experiments show that our approach achieves 36.1% higher Hausa name tagging F-score over a costly supervised model, and 9.4% higher Chinese-to-English Entity Linking accuracy over state-of-the-art.

tacl-913**tacl-913**

11:40–12:00

Named entity recognition is a challenging task that has traditionally required large amounts of knowledge in the form of feature engineering and lexicons to achieve high performance. In this paper, we present a novel neural network architecture that automatically detects word- and character-level features using a hybrid bidirectional LSTM and CNN architecture, eliminating the need for most feature engineering. We also propose a novel method of encoding partial lexicon matches in neural networks and compare it to existing approaches. Extensive evaluation shows that, given only tokenized text and publicly available word embeddings, our system is competitive on the CoNLL-2003 dataset and surpasses the previously reported state of the art on the OntoNotes 5.0 dataset by 2.13 F1 points. By using two lexicons from public sources, we establish new states of the art with an F1 score of 91.62 on CoNLL-2003 and 86.28 on OntoNotes, surpassing systems that employ heavy feature engineering, proprietary lexicons, and rich entity linking information.

1C: Machine translation I

Plaza Ballroom F

*Chair:***papers-543****papers-543**

10:40–11:00

We apply phrase-based and neural models to a core task in interactive machine translation: suggesting how to complete a partial translation. For the phrase-based system, we demonstrate improvements in suggestion quality using novel objective functions, learning techniques, and inference algorithms tailored to this task. Our contributions include new tunable metrics, an improved beam search strategy, an n-best extraction method that increases suggestion diversity, and a tuning procedure for a hierarchical joint model of alignment and translation. The combination of these techniques improves next-word suggestion accuracy dramatically from 28.5% to 41.2% in a large-scale English-German experiment. Our recurrent neural translation system increases accuracy to 53.0%, but inference is two orders of magnitude slower. Manual error analysis shows the strengths and weaknesses of both approaches.

papers-304**papers-304**

11:00–11:20

Attention mechanism has enhanced state-of-the-art Neural Machine Translation (NMT) by jointly learning to align and translate. It tends to ignore past alignment information, however, which often leads to over-translation and under-translation. To address this problem, we propose coverage-based NMT in this paper. We maintain a coverage vector to keep track of the attention history. The coverage vector is fed to the attention model to help adjust future attention, which lets NMT system to consider more about untranslated source words. Experiments show that the proposed approach significantly improves both translation quality and alignment quality over standard attention-based NMT.

papers-212**papers-212**

11:20–11:40

Neural Machine Translation (NMT) has obtained state-of-the-art performance for several language pairs, while only using parallel data for training. Target-side monolingual data plays an important role in boosting fluency for phrase-based statistical machine translation, and we investigate the use of monolingual data for NMT. In contrast to previous work, which combines NMT models with separately trained language models, we note that encoder-decoder NMT architectures already have the capacity to learn the same information as a language model, and we explore strategies to train with monolingual data without changing the neural network architecture. By pairing monolingual training data with an automatic back-translation, we can treat it as additional parallel training data, and we obtain substantial improvements on the WMT 15 task English->German (+2.8–3.7 BLEU), and for the low-resourced IWSLT 14 task Turkish->English (+2.1–3.4 BLEU), obtaining new state-of-the-art results. We also show that fine-tuning on in-domain monolingual and parallel data gives substantial improvements for the IWSLT 15 task English-German.

papers-357**papers-357**

11:40–12:00

One major drawback of phrase-based translation is that it segments an input sentence into continuous phrases. To support linguistically informed source discontinuity, in this paper we construct graphs which combine bigram and dependency relations and propose a graph-based translation model. The model segments an input graph into connected subgraphs, each of which may cover a discontinuous phrase. We use beam search to combine translations of each subgraph left-to-right to produce a complete translation. Experiments on Chinese–English and German–English tasks show that our system is significantly better than the phrase-based model by up to +1.5/+0.5 BLEU scores. By explicitly modeling the graph segmentation, our system obtains further improvement, especially on German–English.

1D: Word meaning I

Plaza Ballroom F

Chair:

papers-520

papers-520

10:40–11:00

As a new generation of cognitive robots start to enter our lives, it is important to enable robots to follow human commands and to learn new actions from human language instructions. To address this issue, this paper presents an approach that explicitly represents verb semantics through hypothesis spaces of fluents and automatically acquires these hypothesis spaces by interacting with humans. The learned hypothesis spaces can be used to automatically plan for lower-level primitive actions towards physical world interaction. Our empirical results have shown that the representation of a hypothesis space of fluents, combined with the learned hypothesis selection algorithm, outperforms a previous baseline. In addition, our approach applies incremental learning which can contribute to life-long learning from humans in the future.

papers-154

papers-154

11:00–11:20

We propose a framework for lexical substitution that is able to perform transfer learning across languages. Datasets for this task are available in at least three languages (English, Italian, and German). Previous work has addressed each of these tasks in isolation. In contrast, we regard the union of three shared tasks as a combined multilingual dataset. We show that a supervised system can be trained effectively, even if training and evaluation data are from different languages. Successful transfer learning between languages suggests that the learned model is in fact independent of the underlying language. We combine state-of-the-art unsupervised features obtained from syntactic word embeddings and distributional thesauri in a supervised delexicalized ranking system. Our system improves over state of the art in the full lexical substitution task in all three languages.

papers-697

papers-697

11:20–11:40

We use Bayesian optimization to learn curricula for word representation learning, optimizing performance on downstream tasks that depend on the learned representations as features. The curricula are modeled by a linear ranking function which is the scalar product of a learned weight vector and an engineered feature vector that characterizes the different aspects of the complexity of each instance in the training corpus. We show that learning the curriculum improves performance on a variety of downstream tasks over random orders and in comparison to the natural corpus order.

papers-741

papers-741

11:40–12:00

The problem of rare and unknown words is an important issue that can potentially effect the performance of many NLP systems, including traditional count-based and deep learning models. We propose a novel way to deal with the rare and unseen words for the neural network models using attention. Our model uses two softmax layers in order to predict the next word in conditional language models: one predicts the location of a word in the source sentence, and the other predicts a word in the shortlist vocabulary. At each timestep, the decision of which softmax layer to use is adaptively made by an MLP which is conditioned on the context.~We motivate this work from a psychological evidence that humans naturally have a tendency to point towards objects in the context or the environment when the name of an object is not known.~Using our proposed model, we observe improvements on two tasks, neural machine translation on the Europarl English to French parallel corpora and text summarization on the Gigaword dataset.

1E: Parsing I

Plaza Ballroom F

*Chair:***short-241****short-241**

The topological field model is commonly used to describe the regularities in German word order. In this work, we show that topological fields can be predicted reliably using sequence labeling and that the predicted field labels can improve transition-based dependency parsing.

papers-500**papers-500**

10:40–11:00

In this paper, we present a generalized transition-based parsing framework where parsers are instantiated in terms of a set of control parameters that constrain transitions between parser states. This generalization provides a unified framework to describe and compare various transition-based parsing approaches from both a theoretical and empirical perspective. This includes well-known transition-systems, but also previously un-studied systems.

papers-431**papers-431**

11:00–11:20

We present a transition-based system that jointly predicts the syntactic structure and lexical units of a sentence by building two structures over the input words: a syntactic dependency tree and a forest of lexical units including multiword expressions (MWEs). This combined representation allows us to capture both the syntactic and semantic structure of MWEs, which in turn enables deeper downstream semantic analysis, especially for semi-compositional MWEs. The proposed system extends the arc-standard transition system for dependency parsing with transitions for building complex lexical units. Experiments on two different data sets show that the approach significantly improves MWE identification accuracy (and sometimes syntactic accuracy) compared to existing joint approaches.

papers-318**papers-318**

11:20–11:40

Dynamic oracle training has shown substantial improvements for dependency parsing in various settings, but has not been explored for constituent parsing. The present article introduces a dynamic oracle for transition-based constituent parsing. Experiments on the 9 languages of the SPMRL dataset show that a neural greedy parser with morphological features, trained with a dynamic oracle, leads to accuracies comparable with the best non-reranking and non-ensemble parsers.

1F: Noncompositionality

Plaza Ballroom F

Chair:

papers-122

papers-122

10:40–11:00

Metaphorical expressions are pervasive in natural language and pose a substantial challenge for computational semantics. The inherent compositionality of metaphor makes it an important test case for compositional distributional semantic models (CDSMs). This paper is the first to investigate whether metaphorical composition warrants a distinct treatment in the CDSM framework. We propose a method to learn metaphors as linear transformations in a vector space and find that, across a variety of semantic domains, explicitly modeling metaphor improves the resulting semantic representations. We then use these representations in a metaphor identification task, achieving a high performance of 0.82 in terms of F-score.

papers-504

papers-504

11:00–11:20

Idiom token classification is the task of deciding for a set of potentially idiomatic phrases whether each occurrence of a phrase is a literal or idiomatic usage of the phrase. In this work we explore the use of Skip-Thought Vectors to create distributed representations that encode features that are predictive with respect to idiom token classification. We show that classifiers using these representations have competitive performance compared with the state of the art in idiom token classification. Importantly, however, our models use only the sentence containing the target phrase as input and are thus less dependent on a potentially inaccurate or incomplete model of discourse context. We further demonstrate the feasibility of using these representations to train a competitive general idiom token classifier.

papers-197

papers-197

11:20–11:40

We present a novel method for jointly learning compositional and non-compositional phrase embeddings by adaptively weighting both types of embeddings using a compositionality scoring function. The scoring function is used to quantify the level of compositionality of each phrase, and the parameters of the function are jointly optimized with the objective for learning phrase embeddings. In experiments, we apply the adaptive joint learning method to the task of learning embeddings of transitive verb phrases, and show that the compositionality scores have strong correlation with human ratings for verb-object compositionality, substantially outperforming the previous state of the art. Moreover, our embeddings improve upon the previous best model on a transitive verb disambiguation task. We also show that a simple ensemble technique further improves the results for both tasks.

papers-828

papers-828

11:40–12:00

Metaphor is a common linguistic tool in communication, making its detection in discourse a crucial task for natural language understanding. One popular approach to this challenge is to capture semantic incohesion between a metaphor and the dominant topic of the surrounding text. While these methods are effective, they tend to overclassify target words as metaphorical when they deviate in meaning from its context. We present a new approach that (1) distinguishes literal and non-literal use of target words by examining sentence-level topic transitions and (2) captures the motivation of speakers to express emotions and abstract concepts metaphorically. Experiments on an online breast cancer discussion forum dataset demonstrate a significant improvement in metaphor detection over the state-of-the-art. These experimental results also reveal a tendency toward metaphor usage in personal topics and certain emotional contexts.

Recruitment Lunch

Monday, August 8, 2016, 12:30–2:30pm

Maritim proArte Hotel, Friedrichstraße 151 (opposite to the Friedrichstraße train and subway station)

The ACL-2016 Recruitment Lunch, sponsored by various companies and institutions, is meant to benefit students and other ACL participants. The aim of the lunch is to give both parties the opportunity to meet and discuss employment possibilities.

The Recruitment Lunch will be held at the Maritim Hotel, Berlin, on Monday 8 August, 12.30–14.30, in advance of the ACL-2016 poster session in the same location. Participation of the lunch will be based on registration and availability, and will be capped at 400 participants.

While enjoying a free lunch offered at food stations, participants will have the opportunity to meet the sponsors at dedicated poster boards and learn about employment opportunities.

Participating Recruiters (as of date of printing):

- Nuance
- Google
- Sony
- Amazon
- Dublin City University
- University of Copenhagen
- Ocado Technology
- Textkernel
- Zalando SE

Session 2 Overview – Monday, August 8, 2016

Track A <i>Word vectors I</i>	Track B <i>Events and schemas</i>	Track C <i>Sentiment analysis</i>	Track D <i>Parsing II</i>	Track E <i>Information retrieval</i>	Track F <i>Phonology and morphology</i>
Plaza Ballroom	Plaza Ballroom	Plaza Ballroom	Plaza Ballroom	Plaza Ballroom	Plaza Ballroom F
tacl-916	papers-272	papers-674	papers-100	papers-683	papers-742
tacl-916	papers-272	papers-674	papers-100	papers-683	papers-742
papers-804	papers-275	papers-797	tacl-903	papers-810	tacl-905
papers-804	papers-275	papers-797	tacl-903	papers-810	tacl-905
papers-093	papers-511	papers-071	papers-468	papers-870	papers-320
papers-093	papers-511	papers-071	papers-468	papers-870	papers-320
papers-332	papers-331	papers-719		short-528	papers-651
papers-332	papers-331	papers-719		short-528	papers-651

02:40 02:20 02:00 01:40

Parallel Session 2

2A: Word vectors I

Plaza Ballroom F

Chair:

tacl-916
tacl-916

13:40–14:00

Continuous word representations have been remarkably useful across NLP tasks but remain poorly understood. We ground word embeddings in semantic spaces studied in the cognitive-psychometric literature, taking these spaces as the primary objects to recover. To this end, we relate log co-occurrences of words in large corpora to semantic similarity assessments and show that co-occurrences are indeed consistent with an Euclidean semantic space hypothesis. Framing word embedding as metric recovery of a semantic space unifies existing word embedding algorithms, ties them to manifold learning, and demonstrates that existing algorithms are consistent metric recovery methods given co-occurrence counts from random walks. Furthermore, we propose a simple, principled, direct metric recovery algorithm that performs on par with the state-of-the-art word embedding and manifold learning methods. Finally, we complement recent focus on analogies by constructing two new inductive reasoning datasets—series completion and classification—and demonstrate that word embeddings can be used to solve them as well.

papers-804
papers-804

14:00–14:20

Neural networks are among the state-of-the-art techniques for language modeling. Existing neural language models typically map discrete words to distributed, dense vector representations. After information processing of the preceding context words by hidden layers, an output layer estimates the probability of the next word. Such approaches are time- and memory-intensive because of the large numbers of parameters for word embeddings and the output layer. In this paper, we propose to compress neural language models by sparse word representations. In the experiments, the number of parameters in our model increases very slowly with the growth of the vocabulary size, which is almost imperceptible. Moreover, our approach not only reduces the parameter space to a large extent, but also improves the performance in terms of the perplexity measure.

papers-093
papers-093

14:20–14:40

We introduce a new methodology for intrinsic evaluation of word representations. Specifically, we identify four fundamental criteria based on the characteristics of natural language that pose difficulties to NLP systems; and develop tests that directly show whether or not representations contain the subspaces necessary to satisfy these criteria. Current intrinsic evaluations are mostly based on the overall similarity or full-space similarity of words and thus view vector representations as points. We show the limits of these point-based intrinsic evaluations. We apply our evaluation methodology to the comparison of a count vector model and several neural network models and demonstrate important properties of these models.

papers-332
papers-332

14:40–15:00

A shared bilingual word embedding space (SBWES) is an indispensable resource in a variety of cross-language NLP and IR tasks. A common approach to the SBWES induction is to learn a mapping function between monolingual semantic spaces, where the mapping critically relies on a seed word lexicon used in the learning process. In this work, we analyze the importance and properties of seed lexicons for the SBWES induction across different dimensions (i.e., lexicon source, lexicon size, translation method, translation pair reliability). On the basis of our analysis, we propose a simple but effective hybrid bilingual word embedding (BWE) model. This model (HYBWE) learns the mapping between two monolingual embedding spaces using only highly reliable symmetric translation pairs from a seed document-level embedding space. We perform bilingual lexicon learning (BLL) with 3 language pairs and show that by carefully selecting reliable translation pairs our new HYBWE model outperforms benchmarking BWE learning models, all of which use more expensive bilingual signals. Effectively, we demonstrate that a SBWES may be induced by leveraging only a very weak bilingual signal (document alignments) along with monolingual data.

2B: Events and schemas

Plaza Ballroom F

*Chair:***papers-272****papers-272**

13:40–14:00

We propose a brand new “Liberal” Event Extraction paradigm to extract events and discover event schemas from any input corpus simultaneously. We incorporate symbolic (e.g., Abstract Meaning Representation) and distributional semantics to detect and represent event structures and adopt a joint typing framework to simultaneously extract event types and argument roles and discover an event schema. Experiments on general and specific domains demonstrate that this framework can construct high-quality schemas with many event and argument role types, covering a high proportion of event types and argument roles in manually defined schemas. We show that extraction performance using discovered schemas is comparable to supervised models trained from a large amount of data labeled according to pre-defined event types. The extraction quality of new event types is also promising.

papers-275**papers-275**

14:00–14:20

Event extraction from texts aims to detect structured information such as what has happened, to whom, where and when. Event extraction and visualization are typically considered as two different tasks. In this paper, we propose a novel approach based on probabilistic modelling to jointly extract and visualize events from tweets where both tasks benefit from each other. We model each event as a joint distribution over named entities, a date, a location and event-related keywords. Moreover, both tweets and event instances are associated with coordinates in the visualization space. The manifold assumption that the intrinsic geometry of tweets is a low-rank, non-linear manifold within the high-dimensional space is incorporated into the learning framework using a regularization. Experimental results show that the proposed approach can effectively deal with both event extraction and visualization and performs remarkably better than both the state-of-the-art event extraction method and a pipeline approach for event extraction and visualization.

papers-511**papers-511**

14:20–14:40

There is a small but growing body of research on statistical scripts, models of event sequences that allow probabilistic inference of implicit events from documents. These systems operate on structured verb-argument events produced by an NLP pipeline. We compare these systems with recent Recurrent Neural Net models that directly operate on raw tokens to predict sentences, finding the latter to be roughly comparable to the former in terms of predicting missing events in documents.

papers-331**papers-331**

14:40–15:00

Natural language understanding often requires deep semantic knowledge. Expanding on previous proposals, we suggest that some important aspects of semantic knowledge can be modeled as a language model if done at an appropriate level of abstraction. We develop two distinct models that capture semantic frame chains and discourse information while abstracting over the specific mentions of predicates and entities. For each model, we investigate four implementations: a “standard” N-gram language model and three discriminatively trained “neural” language models that generate embeddings for semantic frames. The quality of the semantic language models (SemLM) is evaluated both intrinsically, using perplexity and a narrative cloze test and extrinsically – we show that our SemLM helps improve performance on semantic natural language processing tasks such as co-reference resolution and discourse parsing.

2C: Sentiment analysis

Plaza Ballroom F

*Chair:***papers-674****papers-674**

13:40–14:00

Domain adaptation is an important research topic in sentiment analysis area. Existing domain adaptation methods usually transfer sentiment knowledge from only one source domain to target domain. In this paper, we propose a new domain adaptation approach which can exploit sentiment knowledge from multiple source domains. We first extract both global and domain-specific sentiment knowledge from the data of multiple source domains using multi-task learning. Then we transfer them to target domain with the help of words' sentiment polarity relations extracted from the unlabeled target domain data. The similarities between target domain and different source domains are also incorporated into the adaptation process. Experimental results on benchmark dataset show the effectiveness of our approach in improving cross-domain sentiment classification performance.

papers-797**papers-797**

14:00–14:20

Through a particular choice of a predicate (e.g., “x violated y”), a writer can subtly connote a range of implied sentiment and presupposed facts about the entities x and y: (1) writer’s perspective: projecting x as an “antagonist” and y as a “victim”, (2) entities’ perspective: y probably dislikes x, (3) effect: something bad happened to y, (4) value: y is something valuable, and (5) mental state: y is distressed by the event. We introduce connotation frames as a representation formalism to organize these rich dimensions of connotation using typed relations. First, we investigate the feasibility of obtaining connotative labels through crowdsourcing experiments. We then present models for predicting the connotation frames of verb predicates based on their distributional word representations and the interplay between different types of connotative relations. Empirical results confirm that connotation frames can be induced from various data sources that reflect how language is used in context. We conclude with analytical results that show the potential use of connotation frames for analyzing subtle biases in online news media.

papers-071**papers-071**

14:20–14:40

Sentiment classification aims to automatically predict sentiment polarity (e.g., positive or negative) of user generated sentiment data (e.g., reviews, blogs). Due to the mismatch among different domains, a sentiment classifier trained in one domain may not work well when directly applied to other domains. Thus, domain adaptation for sentiment classification algorithms are highly desirable to reduce the domain discrepancy and manual labeling costs. To address the above challenge, we propose a novel domain adaptation method, called Bi-Transferring Deep Neural Networks (BTDNNs). The proposed BTDNNs attempts to transfer the source domain examples to the target domain, and also transfer the target domain examples to the source domain. The linear transformation of BTDNNs ensures the feasibility of transferring between domains, and the distribution consistency between the transferred domain and the desirable domain is constrained with a linear data reconstruction manner. As a result, the transferred source domain is supervised and follows similar distribution as the target domain. Therefore, any supervised method can be used on the transferred source domain to train a classifier for sentiment classification in a target domain. We conduct experiments on a benchmark composed of reviews of 4 types of Amazon products. Experimental results show that our proposed approach significantly outperforms the several baseline methods, and achieves an accuracy which is competitive with the state-of-the-art method for domain adaptation.

papers-719**papers-719**

14:40–15:00

We present a new approach for document-level sentiment inference, where the goal is to predict directed opinions (who feels positively or negatively towards whom) for all entities mentioned in a text. To encourage more complete and consistent predictions, we introduce an ILP that jointly models (1) sentence- and discourse-level sentiment cues, (2) factual evidence about entity factions, and (3) global constraints based on social science theories such as homophily, social balance, and reciprocity. Together, these cues allow for rich inference across groups of entities, including for example that CEOs and the companies they lead are likely to have similar sentiment towards others. We evaluate performance on new, densely labeled data that provides supervision for all pairs, complementing previous work that only labeled pairs mentioned in the same sentence. Experi-

iments demonstrate that the global model outperforms sentence-level baselines, by providing more coherent predictions across sets of related entities.

2D: Parsing II

Plaza Ballroom F

*Chair:***papers-100****papers-100**

13:40–14:00

Different from traditional active learning based on sentence-wise full annotation (FA), this paper proposes active learning with dependency-wise partial annotation (PA) as a finer-grained unit for dependency parsing. At each iteration, we select a few most uncertain words from an unlabeled data pool, manually annotate their syntactic heads, and add the partial trees into labeled data for parser retraining. Compared with sentence-wise FA, dependency-wise PA gives us more flexibility in task selection and avoids wasting time on annotating trivial tasks in a sentence. Our work makes the following contributions. First, we are the first to apply a probabilistic model to active learning for dependency parsing, which can 1) provide tree probabilities and dependency marginal probabilities as principled uncertainty metrics, and 2) directly learn parameters from PA based on a forest-based training objective. Second, we propose and compare several uncertainty metrics through simulation experiments on both Chinese and English. Finally, we conduct human annotation experiments to compare FA and PA on real annotation time and quality.

tacl-903**tacl-903**

14:00–14:20

We study the generalization of maximum spanning tree dependency parsing to maximum acyclic subgraphs. Because the underlying optimization problem is intractable even under an arc-factored model, we consider the restriction to noncrossing dependency graphs. Our main contribution is a cubic-time exact inference algorithm for this class. We extend this algorithm into a practical parser and evaluate its performance on four linguistic data sets used in semantic dependency parsing. We also explore a generalization of our parsing framework to dependency graphs with pagenumber at most k and show that the resulting optimization problem is NP-hard for $k \geq 2$.

papers-468**papers-468**

14:20–14:40

We present a novel dependency parsing method which enforces two structural properties on dependency trees: bounded block degree and well-nestedness. These properties are useful to better represent the set of admissible dependency structures in treebanks and connect dependency parsing to context-sensitive grammatical formalisms. We cast this problem as an Integer Linear Program that we solve with Lagrangian Relaxation from which we derive a heuristic and an exact method based on a Branch-and-Bound search. Experimentally, we see that these methods are efficient and competitive compared to a baseline unconstrained parser, while enforcing structural properties in all cases.

2E: Information retrieval

Plaza Ballroom F

*Chair:***papers-683****papers-683**

13:40–14:00

Continuous space word embeddings have received a great deal of attention in the natural language processing and machine learning communities for their ability to model term similarity and other relationships. We study the use of term relatedness in the context of query expansion for ad hoc information retrieval. We demonstrate that word embeddings such as word2vec and GloVe, when trained globally, underperform corpus and query specific embeddings for retrieval tasks. These results suggest that other tasks benefiting from global embeddings may also benefit from local embeddings.

papers-810**papers-810**

14:00–14:20

Community Question Answering (cQA) services like Yahoo! Answers, Baidu Zhidao, Quora, StackOverflow etc. provide a platform for interaction with experts and help users to obtain precise and accurate answers to their questions. The time lag between the user posting a question and receiving its answer could be reduced by retrieving similar historic questions from the cQA archives. The main challenge in this task is the “lexico-syntactic” gap between the current and the previous questions. In this paper, we propose a novel approach called “Siamese Convolutional Neural Network for cQA (SCQA)” to find the semantic similarity between the current and the archived questions. SCQA consist of twin convolutional neural networks with shared parameters and a contrastive loss function joining them. SCQA learns the similarity metric for question-question pairs by leveraging the question-answer pairs available in cQA forum archives. The model projects semantically similar question pairs nearer to each other and dissimilar question pairs farther away from each other in the semantic space. Experiments on large scale real-life “Yahoo! Answers” dataset reveals that SCQA outperforms current state-of-the-art approaches based on translation models, topic models and deep neural network based models which use non-shared parameters.

papers-870**papers-870**

14:20–14:40

In this work, we focus on the problem of news citation recommendation. The task aims to recommend news citations for both authors and readers to create and search news references. Due to the sparsity issue of news citations and the engineering difficulty in obtaining information on authors, we focus on content similarity-based methods instead of collaborative filtering-based approaches. In this paper, we explore word embedding (i.e. implicit semantics) and grounded entities (i.e. explicit semantics) to address the variety and ambiguity issues of language. We formulate the problem as a re-ranking task and integrate different similarity measures under the learning to rank framework. We evaluate our approach on a real-world dataset. The experimental results show the efficacy of our method.

short-528**short-528**

14:40–15:00

The enormous scale of unlabeled text available today necessitates scalable schemes for representation learning in natural language processing. For instance, in this paper we are interested in classifying the intent of a user query. While our labeled data is quite limited, we have access to virtually an unlimited amount of unlabeled queries, which could be used to induce useful representations: for instance by principal component analysis(PCA). However, it is prohibitive to even store the data in memory due to its sheer size, let alone apply conventional batch algorithms. In this work, we apply the recently proposed matrix sketching algorithm to entirely obviate the problem with scalability (Liberty, 2013). This algorithm approximates the data within a specified memory bound while preserving the covariance structure necessary for PCA. Using matrix sketching, we significantly improve the user intent classification accuracy by leveraging large amounts of unlabeled queries.

2F: Phonology and morphology

Plaza Ballroom F

*Chair:***papers-742****papers-742**

13:40–14:00

Grapheme-to-phoneme (g2p) models are rarely available in low-resource languages, as the creation of training and evaluation data is expensive and time-consuming. We use Wiktionary to obtain more than 650k word-pronunciation pairs in more than 500 languages. We then develop phoneme and language distance metrics based on phonological and linguistic knowledge; applying those, we adapt g2p models for high-resource languages to create models for related low-resource languages. We provide results for models for 229 adapted languages.

tacl-905**tacl-905**

14:00–14:20

Morpho-syntactic lexicons provide information about the morphological and syntactic roles of words in a language. Such lexicons are not available for all languages and even when available, their coverage can be limited. We present a graph-based semi-supervised learning method that uses the morphological, syntactic and semantic relations between words to automatically construct wide coverage lexicons from small seed sets. Our method is language-independent, and we show that we can expand a 1000 word seed lexicon to more than 100 times its size with high quality for 11 languages. In addition, the automatically created lexicons provide features that improve performance in two downstream tasks: morphological tagging and dependency parsing.

papers-320**papers-320**

14:20–14:40

Most previous approaches to Chinese word segmentation formalize this problem as a character-based sequence labeling task so that only contextual information within fixed sized local windows and simple interactions between adjacent tags can be captured. In this paper, we propose a novel neural framework which thoroughly eliminates context windows and can utilize complete segmentation history. Our model employs a gated combination neural network over characters to produce distributed representations of word candidates, which are then given to a long short-term memory (LSTM) language scoring model. Experiments on the benchmark datasets show that without the help of feature engineering as most existing approaches, our models achieve competitive or better performances with previous state-of-the-art methods.

papers-651**papers-651**

14:40–15:00

Character-based and word-based methods are two main types of statistical models for Chinese word segmentation, the former exploiting sequence labeling models over characters and the latter typically exploiting a transition-based model, with the advantages that word-level features can be easily utilized. Neural models have been exploited for character-based Chinese word segmentation, giving high accuracies by making use of external character embeddings, yet requiring less feature engineering. In this paper, we study a neural model for word-based Chinese word segmentation, by replacing the manually-designed discrete features with neural features in a word-based segmentation framework. Experimental results demonstrate that word features lead to comparable performances to the best systems in the literature, and a further combination of discrete and neural features gives top accuracies.

Session 3 Overview – Monday, August 8, 2016

	Track A <i>Question answering I</i>	Track B <i>Sentence vectors</i>	Track C <i>Parsing III</i>	Track D <i>Dialog</i>	Track E <i>Generation</i>	Track F <i>Entities and coreference</i>	Track G <i>Topics</i>
	Plaza Ballroom	Plaza Ballroom	Plaza Ballroom	Plaza Ballroom	Plaza Ballroom	Plaza Ballroom	Plaza Ballroom F
03:30	papers-434 papers-434	papers-148 papers-148	tacl-919 tacl-919	papers-088 papers-088	papers-185 papers-185	papers-488 papers-488	papers-604 papers-604
03:50	papers-776 papers-776	tacl-917 tacl-917	tacl-899 tacl-899	papers-725 papers-725	papers-706 papers-706	tacl-900 tacl-900	papers-337 papers-337
04:10	papers-021 papers-021	papers-499 papers-499	tacl-922 tacl-922	papers-549 papers-549	papers-509 papers-509	tacl-902 tacl-902	tacl-910 tacl-910
04:30	papers-244 papers-244	papers-355 papers-355	tacl-918 tacl-918	papers-387 papers-387	papers-253 papers-253	papers-401 papers-401	papers-707 papers-707
04:50	papers-082 papers-082		tacl-920 tacl-920	papers-114 papers-114	papers-322 papers-322	papers-717 papers-717	papers-631 papers-631

Parallel Session 3

3A: Question answering I

Plaza Ballroom F

Chair:

papers-434

papers-434

15:30–15:50

Understanding unstructured text is a major goal within natural language processing. Comprehension tests pose questions based on short text passages to evaluate such understanding. In this work, we investigate machine comprehension on the challenging MCTest benchmark. Partly because of its limited size, prior work on MCTest has focused mainly on engineering better features. We tackle the dataset with a neural approach, harnessing simple neural networks arranged in a parallel hierarchy. The parallel hierarchy enables our model to compare the passage, question, and answer from a variety of trainable perspectives, as opposed to using a manually designed, rigid feature set. Perspectives range from the word level to sentence fragments to sequences of sentences; the networks operate only on word-embedding representations of text. When trained with a methodology designed to help cope with limited training data, our Parallel-Hierarchical model sets a new state of the art for MCTest, outperforming previous feature-engineered approaches slightly and previous neural approaches by a significant margin (over 15 percentage points)

papers-776

papers-776

15:50–16:10

Broad-domain question answering often requires shallow textual methods in order to gain coverage beyond facts that are available in structured knowledge bases, but simultaneously can benefit from the high precision of logical reasoning. We propose an approach for incorporating both of these signals in a unified framework based on natural logic. We train an evaluation function — akin to gameplaying — to evaluate the expected truth of candidate premises on the fly. As an added contribution, we extend the breadth of inferences afforded by natural logic to include relational entailment (e.g., buy → own) and meronymy (e.g., a person born in a city is born in the city's country). We evaluate our approach on answering multiple-choice science questions, outperforming prior work on the dataset.

papers-021

papers-021

16:10–16:30

Cognitive science researchers have emphasized the importance of ordering a complex task into a sequence of easy to hard problems. Such an ordering provides an easier path to learning and increases the speed of acquisition of the task compared to conventional learning. Recent works in machine learning have explored a curriculum learning approach called self-paced learning which orders data samples on the easiness scale so that easy samples can be introduced to the learning algorithm first and harder samples can be introduced successively. We explore curriculum learning in the context of non-convex models for question answering and show that these approaches lead to improvements. We argue that incorporating easy, yet, a diverse set of questions, which are different from the questions already seen by the learner, can further improve learning. Our experiments support this hypothesis.

papers-244

papers-244

16:30–16:50

Passage-level question answer matching is a challenging task since it requires effective representations that capture the complex semantic relations between questions and answers. In this work, we propose a series of deep learning models to address passage answer selection. To match passage answers to questions accommodating their complex semantic relations, unlike most previous work that utilizes a single deep learning structure, we develop hybrid models that process the text using both convolutional and recurrent neural networks, combining the merits on extracting linguistic information from both structures. Additionally, we also develop a simple but effective attention mechanism for the purpose of constructing better answer representations according to the input question, which is imperative for better modeling long answer sequences. The results on two public benchmark datasets, InsuranceQA and TREC-QA, show that our proposed models outperform a variety of strong baselines.

papers-082

papers-082

16:50–17:10

Question answering requires access to a knowledge base to check facts and reason about information. Knowledge in the form of natural language text is easy to acquire, but difficult for automated reasoning. Highly-structured knowledge bases can facilitate reasoning, but are difficult to acquire. In this paper we explore tables as a semi-structured formalism that provides a balanced compromise to this trade-off. We first use the structure of tables to guide the construction of a dataset of over 9000 multiple-choice questions with rich alignment annotations, easily and efficiently via crowd-sourcing. We then use this annotated data to train a semi-structured feature-driven model for question answering that uses tables as a knowledge base. In benchmark evaluations, we significantly outperform both a strong unstructured retrieval baseline and a highly-structured Markov Logic Network model.

3B: Sentence vectors

Plaza Ballroom F

*Chair:***papers-148****papers-148**

15:30–15:50

Traditional approaches to extractive summarization rely heavily on human-engineered features. In this work we propose a data-driven approach based on neural networks and continuous sentence features. We develop a general framework for single-document summarization composed of a hierarchical document encoder and an attention-based extractor. This architecture allows us to develop different classes of summarization models which can extract sentences or words. We train our models on large scale corpora containing hundreds of thousands of document-summary pairs. Experimental results on two summarization datasets demonstrate that our models obtain results comparable to the state of the art without any access to linguistic annotation.

tacl-917**tacl-917**

15:50–16:10

How to model a pair of sentences is a critical issue in many NLP tasks such as answer selection (AS), paraphrase identification (PI) and textual entailment (TE). Most prior work (i) deals with one individual task by fine-tuning a specific system; (ii) models each sentence's representation separately, rarely considering the impact of the other sentence; or (iii) relies fully on manually designed, task-specific linguistic features. This work presents a general Attention Based Convolutional Neural Network (ABCNN) for modeling a pair of sentences. We make three contributions. (i) The ABCNN can be applied to a wide variety of tasks that require modeling of sentence pairs. (ii) We propose three attention schemes that integrate mutual influence between sentences into CNNs; thus, the representation of each sentence takes into consideration its counterpart. These interdependent sentence pair representations are more powerful than isolated sentence representations. (iii) ABCNNs achieve state-of-the-art performance on AS, PI and TE tasks.

papers-499**papers-499**

16:10–16:30

Automatic negation scope detection is a task that has been tackled using different classifiers and heuristics. Most systems are however 1) highly-engineered, 2) English-specific, and 3) only tested on the same genre they were trained on. We start by addressing 1) and 2) using a neural network architecture. Results obtained on data from the *SEM2012 shared task on negation scope detection show that even a simple feed-forward neural network using word-embedding features alone, performs on par with earlier classifiers, with a bi-directional LSTM outperforming all of them. We then address 3) by means of a specially-designed synthetic test set; in doing so, we explore the problem of detecting the negation scope more in depth and show that performance suffers from genre effects and differs with the type of negation considered.

papers-355**papers-355**

16:30–16:50

Most sentence embedding models typically represent each sentence only using word surface, which makes these models indiscriminative for ubiquitous polysemy. In order to enhance semantic representation capability of sentence, we employ text conceptualization algorithm to assign associated concepts for each sentence, and then learn conceptual sentence embedding (CSE). Therefore, this concept-level representation is more expressive than some widely-used text representation models such as latent topic model, especially for short-text. Moreover, we further extend CSE models by utilizing a local attention-based model that select relevant words within the context to make more efficient prediction. The experimental results on text classification task and information retrieval task, show that the proposed models outperform the topic model-based sentence representation methods and other sentence embedding models.

3C: Parsing III

Plaza Ballroom F

*Chair:***tacl-919****tacl-919**

15:30–15:50

We propose a novel approach to cross-lingual part-of-speech tagging and dependency parsing for truly low-resource languages. Our annotation projection-based approach yields tagging and parsing models for over 100 languages. All that is needed are freely available parallel texts, and taggers and parsers for resource-rich languages. The empirical evaluation across 30 test languages shows that our method consistently provides top-level accuracies, close to established upper bounds, and outperforms several competitive baselines.

tacl-899**tacl-899**

15:50–16:10

Adaptor grammars are a flexible, powerful formalism for defining nonparametric, unsupervised models of grammar productions. This flexibility comes at the cost of expensive inference. We address the difficulty of inference through an online algorithm which uses a hybrid of Markov chain Monte Carlo and variational inference. We show that this inference strategy improves scalability without sacrificing performance on unsupervised word segmentation and topic modeling tasks.

tacl-922**tacl-922**

16:10–16:30

We train one multilingual model for dependency parsing and use it to parse sentences in several languages. The parsing model uses (i) multilingual word clusters and embeddings; (ii) token-level language information; and (iii) language-specific features (fine-grained POS tags). This input representation enables the parser not only to parse effectively in multiple languages, but also to generalize across languages based on linguistic universals and typological similarities, making it more effective to learn from limited annotations. Our parser's performance compares favorably to strong baselines in a range of data scenarios, including when the target language has a large treebank, a small treebank, or no treebank for training.

tacl-918**tacl-918**

16:30–16:50

We tackle unsupervised part-of-speech (POS) tagging by learning hidden Markov models (HMMs) that are particularly well-suited for the problem. These HMMs, which we call anchor HMMs, assume that each tag is associated with at least one word that can have no other tag, which is a relatively benign condition for POS tagging (e.g., “the” is a word that appears only under the determiner tag). We exploit this assumption and extend the non-negative matrix factorization framework of Arora et al. (2013) to design a consistent estimator for anchor HMMs. In experiments, our algorithm is competitive with strong baselines such as the clustering method of Brown et al. (1992) and the log-linear model of Berg-Kirkpatrick et al. (2010). Furthermore, it produces an interpretable model in which hidden states are automatically lexicalized by words.

tacl-920**tacl-920**

16:50–17:10

The growing work in multi-lingual parsing faces the challenge of fair comparative evaluation and performance analysis across languages and their treebanks. The difficulty lies in teasing apart the properties of treebanks, such as their size or average sentence length, from those of the annotation scheme, and from the linguistic properties of languages. We propose a method to evaluate the effects of word order of a language on dependency parsing performance, while controlling for confounding treebank properties. The method uses artificially-generated treebanks that are minimal permutations of actual treebanks with respect to two word order properties: word order variation and dependency lengths. Based on these artificial data on twelve languages, we show that longer dependencies and higher word order variability degrade parsing performance. Our method also extends to minimal pairs of individual sentences, leading to a finer-grained understanding of parsing errors.

3D: Dialog

Plaza Ballroom F

*Chair:***papers-088****papers-088**

15:30–15:50

Most current chatbot engines are designed to reply to user utterances based on existing utterance-response (or U-R) pairs. In this paper, we present DocChat, a novel information retrieval approach for chatbot engines that can leverage unstructured documents, instead of U-R pairs, to respond to utterances. A learning to rank model with features designed at different levels of granularity is proposed to measure the relevance between utterances and responses directly. We evaluate our proposed approach in both English and Chinese: (i) For English, we evaluate DocChat on WikiQA and QASent, two answer sentence selection tasks, and compare it with state-of-the-art methods. Reasonable improvements and good adaptability are observed. (ii) For Chinese, we compare DocChat with XiaoIce, a famous Chinese chitchat engine, and side-by-side evaluation shows that DocChat is a perfect complement for chatbot engines using U-R pairs as main source of responses.

papers-725**papers-725**

15:50–16:10

In conversation, speakers tend to “accommodate” or “align” to their partners, changing the style and substance of their communications to be more similar to their partners’ utterances. We focus here on “linguistic alignment,” changes in word choice based on others’ choices. Although linguistic alignment is observed across many different contexts and its degree correlates with important social factors such as power and likability, its sources are still uncertain. We build on a recent probabilistic model of alignment, using it to separate out alignment attributable to words versus word categories. We model alignment in two contexts: telephone conversations and microblog replies. Our results show evidence of alignment, but it is primarily lexical rather than categorical. Furthermore, we find that discourse acts modulate alignment substantially. This evidence supports the view that alignment is shaped by strategic communicative processes related to the ongoing discourse.

papers-549**papers-549**

16:10–16:30

The applicability of entropy rate constancy to dialogue is examined on two spoken dialogue corpora. The principle is found to hold; however, new entropy change patterns within the topic episodes of dialogue are described, which are different from written text. Speaker’s dynamic roles as topic initiators and topic responders are associated with decreasing and increasing entropy, respectively, which results in local convergence between these speakers in each topic episode. This implies that the sentence entropy in dialogue is conditioned on different contexts determined by the speaker’s roles. Explanations from the perspectives of grounding theory and interactive alignment are discussed, resulting in a novel, unified information-theoretic approach of dialogue.

papers-387**papers-387**

16:30–16:50

To establish sophisticated dialogue systems, text planning needs to cope with congruent as well as incongruent interlocutor interests as given in everyday dialogues. Little attention has been given to this topic in text planning in contrast to dialogues that are fully aligned with anticipated user interests. When considering dialogues with congruent and incongruent interlocutor interests, dialogue partners are facing the constant challenge of finding a balance between cooperation and competition. We introduce the concept of fairness that operationalize an equal and adequate, i.e. equitable satisfaction of all interlocutors’ interests. Focusing on Question-Answering (QA) settings, we describe an answer planning approach that support fair dialogues under congruent and incongruent interlocutor interests. Due to the fact that fairness is subjective per se, we present promising results from an empirical study ($N=107$) in which human subjects interacted with a QA system implementing the proposed approach.

papers-114**papers-114**

16:50–17:10

Modeling interactions between two sentences is crucial for a number of natural language processing tasks including Answer Selection, Dialogue Act Analysis, etc. While deep learning methods like Recurrent Neural Network or Convolutional Neural Network have been proved to be powerful for sentence modeling, prior studies paid less attention on interactions between sentences. In this work, we propose a Sentence Interaction

Network (SIN) for modeling the complex interactions between two sentences. By introducing “interaction states” for word and phrase pairs, SIN is powerful and flexible in capturing sentence interactions for different tasks. We obtain significant improvements on Answer Selection and Dialogue Act Analysis without any feature engineering.

3E: Generation

Plaza Ballroom F

Chair:

papers-185**papers-185**

15:30–15:50

In this study, we introduce a non-deterministic method for referring expression generation. We describe two models that account for individual variation in the choice of referential form in automatically generated text: a Naive Bayes model and a Recurrent Neural Network. Both are evaluated using the VaREG corpus. Then we select the best performing model to generate referential forms in texts from the GREC-2.0 corpus and conduct an evaluation experiment in which humans judge the coherence and comprehensibility of the generated texts, comparing them both with the original references and those produced by a random baseline model.

papers-706**papers-706**

15:50–16:10

How much is 131 million US dollars? To help readers put such numbers in context, we propose a new task of automatically generating short descriptions known as perspectives, e.g. “\$131 million is about the cost to employ everyone in Texas over a lunch period”. First, we collect a new dataset of numeric mentions from news articles paired with ratings of generated perspectives using crowdsourcing. We then propose a system to generate these descriptions consisting of two steps: formula construction and description generation. In construction, we compose formulae from numeric facts in a knowledge base and rank the resulting formulas based on familiarity, numeric proximity and semantic compatibility. In generation, we convert a formula into natural language using a sequence-to-sequence neural network model. Our system has a 15.2% F1 improvement over a non-compositional baseline at formula construction and a 12.5 BLEU point improvement over a baseline at description generation.

papers-509**papers-509**

16:10–16:30

Over the past decade, large-scale supervised learning corpora have enabled machine learning researchers to make substantial advances. However, to this date, there are no large-scale question-answer corpora available. In this paper we present the 30M Factoid Question-Answer Corpus, an enormous question-answer pair corpus produced by applying a novel neural network architecture on the knowledge base Freebase to transduce facts into natural language questions. The produced question-answer pairs are evaluated both by human evaluators and using automatic evaluation metrics, including well-established machine translation and sentence similarity metrics. Across all evaluation criteria the question-generation model outperforms the competing template-based baseline. Furthermore, when presented to human evaluators, the generated questions appear to be comparable in quality to real human-generated questions.

papers-253**papers-253**

16:30–16:50

Many language generation tasks require the production of text conditioned on both structured and unstructured inputs. We present a novel neural network architecture which generates an output sequence conditioned on an arbitrary number of input functions. Crucially, our approach allows both the choice of conditioning context and the granularity of generation, for example characters or tokens, to be marginalised, thus permitting scalable and effective training. Using this framework, we address the problem of generating programming code from a mixed natural language and structured specification. We create two new data sets for this paradigm derived from the collectible trading card games Magic the Gathering and Hearthstone. On these, and a third preexisting corpus, we demonstrate that marginalising multiple predictors allows our model to outperform strong benchmarks.

papers-322**papers-322**

16:50–17:10

Research on generating referring expressions has so far mostly focussed on “one-shot reference”, where the aim is to generate a single, discriminating expression. In interactive settings, however, it is not uncommon for reference to be established in “installments”, where referring information is offered piecewise until success has been confirmed. We show that this strategy can also be advantageous in technical systems that only have uncertain access to object attributes and categories. We train a recently introduced model of grounded word meaning on a data set of REs for objects in images and learn to predict semantically appropriate expressions.

In a human evaluation, we observe that users are sensitive to inadequate object names - which unfortunately are not unlikely to be generated from low-level visual input. We propose a solution inspired from human task-oriented interaction and implement strategies for avoiding and repairing semantically inaccurate words. We enhance a word-based REG with context-aware, referential installments and find that they substantially improve the referential success of the system.

3F: Entities and coreference

Plaza Ballroom F

*Chair:***papers-488****papers-488**

15:30–15:50

Entity resolution is the task of linking each mention of an entity in text to the corresponding record in a knowledge base (KB). Coherence models for entity resolution encourage all referring expressions in a document to resolve to entities that are related in the KB. We explore attention-like mechanisms for coherence, where the evidence for each candidate is based on a small set of strong relations, rather than relations to all other entities in the document. The rationale is that document-wide support may simply not exist for non-salient entities, or entities not densely connected in the KB. Our proposed system outperforms state-of-the-art systems on the CoNLL 2003, TAC KBP 2010, 2011 and 2012 tasks.

tacl-900**tacl-900**

15:50–16:10

We present Plato, a probabilistic model for entity resolution that includes a novel approach for handling noisy or uninformative features, and supplements labeled training data derived from Wikipedia with a very large unlabeled text corpus. Training and inference in the proposed model can easily be distributed across many servers, allowing it to scale to over 10^7 entities. We evaluate Plato on three standard datasets for entity resolution. Our approach achieves the best results to-date on TAC KBP 2011 and is highly competitive on both the CoNLL 2003 and TAC KBP 2012 datasets.

tacl-902**tacl-902**

16:10–16:30

Methods for Named Entity Recognition and Disambiguation (NERD) perform NER and NED in two separate stages. Therefore, NED may be penalized with respect to precision by NER false positives, and suffers in recall from NER false negatives. Conversely, NED does not fully exploit information computed by NER such as types of mentions. This paper presents J-NERD, a new approach to perform NER and NED jointly, by means of a probabilistic graphical model that captures mention spans, mention types, and the mapping of mentions to entities in a knowledge base. We present experiments with different kinds of texts from the CoNLL'03, ACE'05, and ClueWeb'09-FACC1 corpora. J-NERD consistently outperforms state-of-the-art competitors in end-to-end NERD precision, recall, and F1.

papers-401**papers-401**

16:30–16:50

Interpretability and discriminative power are the two most basic requirements for an evaluation metric. In this paper, we report the mention identification effect in the B3, CEAf, and BLANC coreference evaluation metrics that makes it impossible to interpret their results properly. The only metric which is insensitive to this flaw is MUC, which, however, is known to be the least discriminative metric. It is a known fact that none of the current metrics are reliable. The common practice for ranking coreference resolvers is to use the average of three different metrics. However, one cannot expect to obtain a reliable score by averaging three unreliable metrics. We propose LEA, a Link-based Entity-Aware evaluation metric that is designed to overcome the shortcomings of the current evaluation metrics. LEA is available as branch LEA-scorer in the reference implementation of the official CoNLL scorer.

papers-717**papers-717**

16:50–17:10

A long-standing challenge in coreference resolution has been the incorporation of entity-level information - features defined over clusters of mentions instead of mention pairs. We present a neural network based coreference system that produces high-dimensional vector representations for pairs of coreference clusters. Using these representations, our system learns when combining clusters is desirable. We train the system with a learning-to-search algorithm that teaches it which local decisions (cluster merges) will lead to a high-scoring final coreference partition. The system substantially outperforms the current state-of-the-art on the English and Chinese portions of the CoNLL 2012 Shared Task dataset despite using few hand-engineered features.

3G: Topics

Plaza Ballroom F

*Chair:***papers-604****papers-604**

15:30–15:50

Properties of corpora, such as the diversity of vocabulary and how tightly related texts cluster together, impact the best way to cluster short texts. We examine several such properties in a variety of corpora and track their effects on various combinations of similarity metrics and clustering algorithms. We show that semantic similarity metrics outperform traditional n -gram and dependency similarity metrics for k-means clustering of a linguistically creative dataset, but do not help with less creative texts. Yet the choice of similarity metric interacts with the choice of clustering method. We find that graph-based clustering methods perform well on tightly clustered data but poorly on loosely clustered data. Semantic similarity metrics generate loosely clustered output even when applied to a tightly clustered dataset. Thus, the best performing clustering systems could not use semantic metrics.

papers-337**papers-337**

15:50–16:10

Word embedding maps words into a low-dimensional continuous embedding space by exploiting the local word collocation patterns in a small context window. On the other hand, topic modeling maps documents onto a low-dimensional topic space, by utilizing the global word collocation patterns in the same document. These two types of patterns are complementary. In this paper, we propose a generative topic embedding model to combine the two types of patterns. In our model, topics are represented by embedding vectors, and are shared across documents. The probability of each word is influenced by both its local context and its topic. A variational inference method yields the topic embeddings as well as the topic mixing proportions for each document. Jointly they represent the document in a low-dimensional continuous space. In two document classification tasks, our method performs better than eight existing methods, with fewer features. In addition, we illustrate with an example that our method can generate coherent topics even based on only one document.

tacl-910**tacl-910**

16:10–16:30

Understanding cross-cultural differences has important implications for world affairs and many aspects of the life of society. Yet, the majority of text-mining methods to date focus on the analysis of monolingual texts. In contrast, we present a statistical model that simultaneously learns a set of common topics from multilingual, non-parallel data and automatically discovers the differences in perspectives on these topics across linguistic communities. We perform a behavioural evaluation of a subset of the differences identified by our model in English and Spanish to investigate their psychological validity.

papers-707**papers-707**

16:30–16:50

News reader comments found in many on-line news websites are typically massive in amount. We investigate the task of Cultural-common Topic Detection (CTD), which is aimed at discovering common discussion topics from news reader comments written in different languages. We propose a new probabilistic graphical model called MCTA which can cope with the language gap and capture the common semantics in different languages. We also develop a partially collapsed Gibbs sampler which effectively incorporates the term translation relationship into the detection of cultural-common topics for model parameter learning. Experimental results show improvements over the state-of-the-art model.

papers-631**papers-631**

16:50–17:10

Document collections often have links between documents—citations, hyperlinks, or revisions—and which links are added is often based on topical similarity. To model these intuitions, we introduce a new topic model for documents situated within a network structure, integrating latent blocks of documents with a max-margin learning criterion for link prediction using topic- and word-level features. Experiments on a scientific paper dataset and collection of webpages show that, by more robustly exploiting the rich link structure within a document network, our model improves link prediction, topic quality, and block distributions.

Poster and dinner session I (includes Demonstrations A and SRW)

Time: 6:00–9:00

Location: Plaza Ballroom A, B, & C

Posters: long papers

papers-554

papers-554

Sentiment Analysis (SA) is an active research area nowadays due to the tremendous interest in aggregating and evaluating opinions being disseminated by users on the Web. SA of English has been thoroughly researched; however research on SA of Arabic has just flourished. Twitter is considered a powerful tool for disseminating information and a rich resource for opinionated text containing views on many different topics. In this paper we attempt to bridge a gap in Arabic SA of Twitter which is the lack of sentiment lexicons that are tailored for the informal language of Twitter. We generate two lexicons extracted from a large dataset of tweets using two approaches and evaluate their use in a simple lexicon based method. The evaluation is performed on internal and external datasets. The performance of these automatically generated lexicons was very promising, albeit the simple method used for classification. The best F-score obtained was 89.58% on the internal dataset and 63.1-64.7% on the external datasets.

papers-347

papers-347

This paper proposes an unsupervised approach for segmenting a multi-author document into authorial components. The key novelty is that we utilize the sequential patterns hidden among document elements when determining their authorships. For this purpose, we adopt Hidden Markov Model (HMM) and construct a sequential probabilistic model to capture the dependencies of sequential sentences and their authorships. An unsupervised learning method is developed to initialize the HMM parameters. Experimental results on benchmark datasets have demonstrated the significant benefit of our idea and our approach has outperformed the state-of-the-arts on all tests. As an example of its applications, the proposed approach is applied for attributing authorship of a document and has also shown promising results.

papers-805

papers-805

Automated Text Scoring (ATS) provides a cost-effective and consistent alternative to human marking. However, in order to achieve good performance, the predictive features of the system need to be manually engineered by human experts. We introduce a model that forms word representations by learning the extent to which specific words contribute to the text's score. Using Long-Short Term Memory networks to represent the meaning of texts, we demonstrate that a fully automated framework is able to achieve excellent results over similar approaches. In an attempt to make our results more interpretable, and inspired by recent advances in visualizing neural networks, we introduce a novel method for identifying the regions of the text that the model has found more discriminative.

papers-615

papers-615

Digital personal assistants are becoming both more common and more useful. The major NLP challenge for personal assistants is machine understanding: translating natural language user commands into an executable representation. This paper focuses on understanding rules written as If-Then statements, though the techniques should be portable to other semantic parsing tasks. We view understanding as structure prediction and show improved models using both conventional techniques and neural network models. We also discuss various ways to improve generalization and reduce overfitting: synthetic training data from paraphrase, grammar combinations, feature selection and ensembles of multiple systems. An ensemble of these techniques achieves a new state of the art result with 8% accuracy improvement.

papers-268

papers-268

We introduce the Treebank of Learner English (TLE), the first publicly available syntactic treebank for English as a Second Language (ESL). The TLE provides manually annotated POS tags and Universal Dependency (UD) trees for 5,124 sentences from the Cambridge First Certificate in English (FCE) corpus. The UD annotations are tied to a pre-existing error annotation of the FCE, whereby full syntactic analyses are provided for both the original and error corrected versions of each sentence. Further on, we delineate ESL annotation guidelines that allow for consistent syntactic treatment of ungrammatical English. Finally, we benchmark POS tagging and dependency parsing performance on the TLE dataset and measure the effect of grammatical errors on parsing accuracy. We envision the treebank to support a wide range of linguistic and computational research on second language acquisition as well as automatic processing of ungrammatical language.

papers-644

papers-644

Neuro-imaging studies on reading different parts of speech (PoS) report somewhat mixed results, yet some of them indicate different activations with different PoS. This paper addresses the difficulty of using fMRI to discriminate between linguistic tokens in reading of running text because of low temporal resolution. We show that once we solve this problem, fMRI data contains a signal of PoS distinctions to the extent that it improves PoS induction with error reductions of more than 4%.

papers-336

papers-336

Relation classification is an important semantic processing task in the field of natural language processing (NLP). In this paper, we present a novel model BRCNN to classify the relation of two entities in a sentence. Some state-of-the-art systems concentrate on modeling the shortest dependency path (SDP) between two entities leveraging convolutional or recurrent neural networks. We further explore how to make full use of the dependency relations information in the SDP, by combining convolutional neural networks and two-channel recurrent neural networks with long short term memory (LSTM) units. We propose a bidirectional architecture to learn relation representations with directional information along the SDP forwards and backwards at the same time, which benefits classifying the direction of relations. Experimental results show that our method outperforms the state-of-the-art approaches on the SemEval-2010 Task 8 dataset.

papers-312

papers-312

A major challenge of semantic parsing is the vocabulary mismatch problem between natural language and target ontology. In this paper, we propose a sentence rewriting based semantic parsing method, which can effectively resolve the mismatch problem by rewriting a sentence into a new form which has the same structure with its target logical form. Specifically, we propose two sentence-rewriting methods for two common types of mismatch: a dictionary-based method for 1-N mismatch and a template-based method for N-1 mismatch. We evaluate our sentence rewriting based semantic parser on the benchmark semantic parsing dataset – WEBQUESTIONS. Experimental results show that our system outperforms the base system with a 3.4% gain in F1, and generates logical forms more accurately and parses sentences more robustly.

papers-820

papers-820

While unsupervised anaphoric zero pronoun (AZP) resolvers have recently been shown to rival their supervised counterparts in performance, it is relatively difficult to scale them up to reach the next level of performance due to the large amount of feature engineering efforts involved and their ineffectiveness in exploiting lexical features. To address these weaknesses, we propose a supervised approach to AZP resolution based on deep neural networks, taking advantage of their ability to learn useful task-specific representations and effectively exploit lexical features via word embeddings. Our approach achieves state-of-the-art performance when resolving the Chinese AZPs in the OntoNotes corpus.

papers-192

papers-192

Supervised machine learning models for automated essay scoring (AES) usually require substantial task-specific training data in order to make accurate predictions for a particular writing task. This limitation hinders their utility, and consequently their deployment in real-world settings. In this paper, we overcome this

shortcoming using a constrained multi-task pairwise-preference learning approach that enables the data from multiple tasks to be combined effectively. Furthermore, contrary to some recent research, we show that high performance AES systems can be built with little or no task-specific training data. We perform a detailed study of our approach on a publicly available dataset in scenarios when we have varying amounts of task-specific training data and in scenarios where the number of tasks increases.

[papers-447](#)**[papers-447](#)**

How can we enable computers to automatically answer questions like “Who created the character Harry Potter”? Carefully built knowledge bases provide rich sources of facts. However, it remains a challenge to answer factoid questions raised in natural language due to numerous expressions of one question. In particular, we focus on the most common questions — ones that can be answered with a single fact in the knowledge base. We propose CFO, a Conditional Focused neural-network-based approach to answering factoid questions with knowledge bases. Our approach first zooms in a question to find more probable candidate subject mentions, and infers the final answers with a unified conditional probabilistic framework. Powered by deep recurrent neural networks and neural embeddings, our proposed CFO achieves an accuracy of 75.7% on a dataset of 108k questions - the largest public one to date. It outperforms the current state of the art by an absolute margin of 11.8%.

[papers-518](#)**[papers-518](#)**

We revisit Levin’s theory about the correspondence of verb meaning and syntax and infer semantic classes from a large syntactic classification of more than 600 German verbs taking clausal and non-finite arguments. Grasping the meaning components of Levin-classes is known to be hard. We address this challenge by setting up a multi-perspective semantic characterization of the inferred classes. To this end, we link the inferred classes and their English translation to independently constructed semantic classes in three different lexicons - the German wordnet GermaNet, VerbNet and FrameNet - and perform a detailed analysis and evaluation of the resulting German-English classification (available at www.ukp.tu-darmstadt.de/modality-verbclasses/).

[papers-486](#)**[papers-486](#)**

Most of the existing Neural Machine Translation (NMT) models focus on the conversion of sequential data and do not directly use syntactic information. We propose a novel end-to-end syntactic NMT model, extending a sequence-to-sequence model with the source-side phrase structure. Our model has an attention mechanism that enables the decoder to generate a translated word while softly aligning it with phrases as well as words of the source sentence. Experimental results on the WAT’15 English-to-Japanese dataset demonstrate that our proposed model considerably outperforms sequence-to-sequence attentional NMT models and compares favorably with the state-of-the-art tree-to-string SMT system.

[papers-074](#)**[papers-074](#)**

Coordination is an important and common syntactic construction which is not handled well by state of the art parsers. Coordinations in the Penn Treebank are missing internal structure in many cases, do not include explicit marking of the conjuncts and contain various errors and inconsistencies. In this work, we initiated manual annotation process for solving these issues. We identify the different elements in a coordination phrase and label each element with its function. We add phrase boundaries when these are missing, unify inconsistencies, and fix errors. The outcome is an extension of the PTB that includes consistent and detailed structures for coordinations. We make the coordination annotation publicly available, in hope that they will facilitate further research into coordination disambiguation.

[papers-152](#)**[papers-152](#)**

User traits disclosed through written text, such as age and gender, can be used to personalize applications such as recommender systems or conversational agents. However, human perception of these traits is not perfectly aligned with reality. In this paper, we conduct a large-scale crowdsourcing experiment on guessing age and gender from tweets. We systematically analyze the quality and possible biases of these predictions. We identify the textual cues which lead to miss-assessments of traits or make workers more or less confident

in their choice. Our study demonstrates that differences between real and perceived traits are noteworthy and elucidates inaccurately used stereotypes in human perception.

papers-775
papers-775

Motivated by the findings in social science that people's opinions are diverse and variable while together they are shaped by evolving social norms, we perform personalized sentiment classification via shared model adaptation over time. In our proposed solution, a global sentiment model is constantly updated to capture the homogeneity in which users express opinions, while personalized models are simultaneously adapted from the global model to recognize the heterogeneity of opinions from individuals. Global model sharing alleviates data sparsity issue, and individualized model adaptation enables efficient online model learning. Extensive experimentations are performed on two large review collections from Amazon and Yelp, and encouraging performance gain is achieved against several state-of-the-art transfer learning and multi-task learning based sentiment classification solutions.

papers-534
papers-534

Our goal is to generate reading lists for students that help them optimally learn technical material. Existing retrieval algorithms return items directly relevant to a query but do not return results to help users read about the concepts supporting their query. This is because the dependency structure of concepts that must be understood before reading material pertaining to a given query is never considered. Here we formulate an information-theoretic view of concept dependency and present methods to construct a "concept graph" automatically from a text corpus. We perform the first human evaluation of concept dependency edges (to be published as open data), and the results verify the feasibility of automatic approaches for inferring concepts and their dependency relations. This result can support search capabilities that may be tuned to help users learn a subject rather than retrieve documents based on a single query.

papers-441
papers-441

We prove that log-linearly interpolated backoff language models can be efficiently and exactly collapsed into a single normalized backoff model, contradicting Hsu (2007). While prior work reported that log-linear interpolation yields lower perplexity than linear interpolation, normalizing at query time was impractical. We normalize the model offline in advance, which is efficient due to a recurrence relationship between the normalizing factors. To tune interpolation weights, we apply Newton's method to this convex problem and show that the derivatives can be computed efficiently in a batch process. These findings are combined in new open-source interpolation tool, which is distributed with KenLM. With 21 out-of-domain corpora, log-linear interpolation yields 72.58 perplexity on TED talks, compared to 75.91 for linear interpolation.

papers-300
papers-300

Recently a few systems for automatically solving math word problems report promising results. However, the datasets used for evaluation have limitations in both scale and diversity. In this paper, we build a large-scale dataset which is more than 7 times the size of previous ones, and contains many more problem types. Problems in the dataset are semi-automatically obtained from community question-answering (CQA) web pages. A ranking SVM model is trained to automatically extract problem answers from the answer text provided by CQA users, which significantly reduces human annotation cost. Experiments conducted on the new dataset lead to interesting and surprising results.

papers-618
papers-618

Recent years have seen a dramatic growth in the popularity of word embeddings mainly owing to their ability to capture semantic information from massive amounts of textual content. As a result, many tasks in Natural Language Processing have tried to take advantage of the potential of these distributional models. In this work, we study how word embeddings can be used in Word Sense Disambiguation, one of the oldest tasks in Natural Language Processing and Artificial Intelligence. We propose different methods through which word embeddings can be leveraged in a state-of-the-art supervised WSD system architecture, and perform a deep analysis of how different parameters affect performance. We show how a WSD system that makes use of word

embeddings alone, if designed properly, can provide significant performance improvement over a state-of-the-art WSD system that incorporates several standard WSD features.

papers-413 **papers-413**

Several large cloze-style context-question-answer datasets have been introduced recently: the CNN and Daily Mail news data and the Children’s Book Test. Thanks to the size of these datasets, the associated text comprehension task is well suited for deep-learning techniques that currently seem to outperform all alternative approaches. We present a new, simple model that uses attention to directly pick the answer from the context as opposed to computing the answer using a blended representation of words in the document as is usual in similar models. This makes the model particularly suitable for question-answering problems where the answer is a single word from the document. Ensemble of our models sets new state of the art on all evaluated datasets.

papers-807 **papers-807**

We investigate the use of deep bi-directional LSTMs for joint extraction of opinion entities and the IS-FROM and IS-ABOUT relations that connect them – the first such attempt using a deep learning approach. Perhaps surprisingly, we find that standard LSTMs are not competitive with a state-of-the-art CRF+ILP joint inference approach (Yang and Cardie, 2013) to opinion entities extraction, performing below even the standalone sequence-tagging CRF. Incorporating sentence-level and a novel relation-level optimization, however, allows the LSTM to identify opinion relations and to perform within 1-3% of the state of-the-art joint model for opinion entities and the IS-FROM relation; and to perform as well as the state-of-the-art for the IS-ABOUT relation – all without access to opinion lexicons, parsers and other preprocessing components required for the feature-rich CRF+ILP approach.

papers-765 **papers-765**

This paper proposes a left-corner parser which can identify nonlocal dependencies. Our parser integrates non-local dependency identification into a transition-based system. We use a structured perceptron which enables our parser to utilize global features captured by nonlocal dependencies. An experimental result demonstrates that our parser achieves a good balance between constituent parsing and nonlocal dependency identification.

papers-376 **papers-376**

We present the Siamese Continuous Bag of Words (Siamese CBOW) model, a neural network for efficient estimation of high-quality sentence embeddings. Averaging the embeddings of words in a sentence has proven to be a surprisingly successful and efficient way of obtaining sentence embeddings. However, word embeddings trained with the methods currently available are not optimized for the task of sentence representation, and, thus, likely to be suboptimal. Siamese CBOW handles this problem by training word embeddings directly for the purpose of being averaged. The underlying neural network learns word embeddings by predicting, from a sentence representation, its surrounding sentences. We show the robustness of the Siamese CBOW model by evaluating it on 20 datasets stemming from a wide variety of sources.

papers-650 **papers-650**

Can we train a system that, on any new input, either says “don’t know” or makes a prediction that is guaranteed to be correct? We answer the question in the affirmative provided our model family is well-specified. Specifically, we introduce the unanimity principle: only predict when all models consistent with the training data predict the same output. We operationalize this principle for semantic parsing, the task of mapping utterances to logical forms. We develop a simple, efficient method that reasons over the infinite set of all consistent models by only checking two of the models. We prove that our method obtains 100% precision even with a modest amount of training data from a possibly adversarial distribution. Empirically, we demonstrate the effectiveness of our approach on the standard GeoQuery dataset.

papers-346 **papers-346**

Dialogue topic tracking is a sequential labelling problem of recognizing the topic state at each time step in given dialogue sequences. This paper presents various artificial neural network models for dialogue topic tracking, including convolutional neural networks to account for semantics at each individual utterance, and recurrent neural networks to account for conversational contexts along multiple turns in the dialogue history. The experimental results demonstrate that our proposed models can significantly improve the tracking performances in human-human conversations.

papers-652

papers-652

Lexico-semantic knowledge of our native language provides an initial foundation for second language learning. In this paper, we investigate whether and to what extent the lexico-semantic models of the native language (L1) are transferred to the second language (L2). Specifically, we focus on the problem of lexical choice and investigate it in the context of three typologically diverse languages: Russian, Spanish and English. We show that a statistical semantic model learned from L1 data improves automatic error detection in L2 for the speakers of the respective L1. Finally, we investigate whether the semantic model learned from a particular L1 is portable to other, typologically related languages.

papers-849

papers-849

Fill-in-the-blank items are commonly featured in computer-assisted language learning (CALL) systems. An item displays a sentence with a blank, and often proposes a number of choices for filling it. These choices should include one correct answer and several plausible distractors. We describe a system that, given an English corpus, automatically generates distractors to produce items for preposition usage. We report a comprehensive evaluation on this system, involving both experts and learners. First, we analyze the difficulty levels of machine-generated carrier sentences and distractors, comparing several methods that exploit learner error and learner revision patterns. We show that the quality of machine-generated items approaches that of human-crafted ones. Further, we investigate the extent to which mismatched L1 between the user and the learner corpora affects the quality of distractors. Finally, we measure the system's impact on the user's language proficiency in both the short and the long term. Further, we investigate the extent to which mismatching L1 between the user and the learner corpora affects the quality of distractors. Finally, we report the system's impact on learners' language proficiency in both the short and the long term.

papers-637

papers-637

We present persona-based models for handling the issue of speaker consistency in neural response generation. A speaker model encodes personas in distributed embeddings that capture individual characteristics such as background information and speaking style. A dyadic speaker-addressee model captures properties of interactions between two interlocutors. Our models yield qualitative performance improvements in both perplexity and BLEU scores over baseline sequence-to-sequence models, with similar gains in speaker consistency as measured by human judges.

papers-033

papers-033

Deep Random Walk (DeepWalk) can learn a latent space representation for describing the topological structure of a network. However, for relational network classification, DeepWalk can be suboptimal as it lacks a mechanism to optimize the objective of the target task. In this paper, we present Discriminative Deep Random Walk (DDRW), a novel method for relational network classification. By solving a joint optimization problem, DDRW can learn the latent space representations that well capture the topological structure and meanwhile are discriminative for the network classification task. Our experimental results on several real social networks demonstrate that DDRW significantly outperforms DeepWalk on multi-label network classification tasks, while retaining the topological structure in the latent space. DDRW is stable and consistently outperforms the baseline methods by various percentages of labeled data. DDRW is also an online method that is scalable and can be naturally parallelized.

papers-628

papers-628

Automatically recognising medical concepts mentioned in social media messages (e.g. tweets) enables several applications for enhancing health quality of people in a community, e.g. real-time monitoring of infectious diseases in population. However, the discrepancy between the type of language used in social media and medical ontologies poses a major challenge. Existing studies deal with this challenge by employing techniques, such as lexical term matching and statistical machine translation. In this work, we handle the medical concept normalisation at the semantic level. We investigate the use of neural networks to learn the transition between layman's language used in social media messages and formal medical language used in the descriptions of medical concepts in a standard ontology. We evaluate our approaches using three different datasets, where social media texts are extracted from Twitter messages and blog posts. Our experimental results show that our proposed approaches significantly and consistently outperform existing effective baselines, which achieved state-of-the-art performance on several medical concept normalisation tasks, by up to 44%.

papers-330**papers-330**

We introduce an agreement-based approach to learning parallel lexicons and phrases from non-parallel corpora. The basic idea is to encourage two asymmetric latent-variable translation models (i.e., source-to-target and target-to-source) to agree on identifying latent phrase and word alignments. The agreement is defined at both word and phrase levels. We develop a Viterbi EM algorithm for jointly training the two unidirectional models efficiently. Experiments on the Chinese-English dataset show that agreement-based learning significantly improves both alignment and translation performance.

papers-658**papers-658**

Recently, there is rising interest in modelling the interactions of text pair with deep neural networks. In this paper, we propose a model of deep fusion LSTMs (DF-LSTMs) to model the strong interaction of text pair in a recursive matching way. Specifically, DF-LSTMs consist of two interdependent LSTMs, each of which models a sequence under the influence of another. We also use external memory to increase the capacity of LSTMs, thereby possibly capturing more complicated matching patterns. Experiments on two very large datasets demonstrate the efficacy of our proposed architecture. Furthermore, we present an elaborate qualitative analysis of our models, giving an intuitive understanding how our model worked.

papers-698**papers-698**

We construct a humans-in-the-loop supervised learning framework that integrates crowdsourcing feedback and local knowledge to detect job-related tweets from individual and business accounts. Using data-driven ethnography, we examine discourse about work by fusing language-based analysis with temporal, geospatial, and labor statistics information.

papers-847**papers-847**

Nearly all previous work on neural machine translation (NMT) has used quite restricted vocabularies, perhaps with a subsequent method to patch in unknown words. This paper presents a novel word-character solution to achieving open vocabulary NMT. We build hybrid systems that translate mostly at the word level and consult the character components for rare words. Our character-level recurrent neural networks compute source word representations and recover unknown target words when needed. The twofold advantage of such a hybrid approach is that it is much faster and easier to train than character-based ones; at the same time, it never produces unknown words as in the case of word-based models. On the WMT'15 English to Czech translation task, this hybrid approach offers an addition boost of +2.1–1.4 BLEU points over models that already handle unknown words. Our best system achieves a new state-of-the-art result with 20.7 BLEU score. We demonstrate that our character models can successfully learn to not only generate well-formed words for Czech, a highly-inflected language with a very complex vocabulary, but also build correct representations for English source words.

papers-046**papers-046**

State-of-the-art sequence labeling systems traditionally require large amounts of task-specific knowledge in the form of hand-crafted features and data pre-processing. In this paper, we introduce a novel neural network

architecture that benefits from both word- and character-level representations automatically, by using combination of bidirectional LSTM, CNN and CRF. Our system is truly end-to-end, requiring no feature engineering or data pre-processing, thus making it applicable to a wide range of sequence labeling tasks. We evaluate our system on two data sets for two sequence labeling tasks — Penn Treebank WSJ corpus for part-of-speech (POS) tagging and CoNLL 2003 corpus for named entity recognition (NER). We obtain state-of-the-art performance on both datasets — 97.55% accuracy for POS tagging and 91.21% F1 for NER.

papers-553

papers-553

Automatic spoken language assessment systems are becoming increasingly important to meet the demand for English second language learning. This is a challenging task due to the high error rates of, even state-of-the-art, non-native speech recognition. Consequently current systems primarily assess fluency and pronunciation. However, content assessment is essential for full automation. As a first stage it is important to judge whether the speaker responds on topic to test questions designed to elicit spontaneous speech. Standard approaches to off-topic response detection assess similarity between the response and question based on bag-of-words representations. An alternative framework based on Recurrent Neural Network Language Models (RNNLM) is proposed in this paper. The RNNLM is adapted to the topic of each test question. It learns to associate example responses to questions with points in a topic space constructed using these example responses. Classification is done by ranking the topic-conditional posterior probabilities of a response. The RNNLMs associate a broad range of responses with each topic, incorporate sequence information and scale better with additional training data, unlike standard methods. On experiments conducted on data from the Business Language Testing Service (BULATS) this approach outperforms standard approaches.

papers-808

papers-808

Most machine translation systems construct translations from a closed vocabulary of target word forms, posing problems for translating into languages that have productive compounding processes. We present a simple and effective approach that deals with this problem in two phases. First, we build a classifier that identifies spans of the input text that can be translated into a single compound word in the target language. Then, for each identified span, we generate a pool of possible compounds which are added to the translation model as “synthetic” phrase translations. Experiments reveal that (i) we can effectively predict what spans can be compounded; (ii) our compound generation model produces good compounds; and (iii) modest improvements are possible in end-to-end English—German and English—Finnish translation tasks. We additionally introduce KomposEval, a new multi-reference dataset of English phrases and their translations into German compounds.

papers-485

papers-485

In this paper, we propose a novel mechanism for enriching the feature vector, for the task of sarcasm detection, with cognitive features extracted from eye-movement patterns of human readers. Sarcasm detection has been a challenging research problem, and its importance for NLP applications such as review summarization, dialog systems and sentiment analysis is well recognized. Sarcasm can often be traced to incongruity that becomes apparent as the full sentence unfolds. This presence of incongruity- implicit or explicit- affects the way readers’ eyes move through the text. We observe the difference in the behaviour of the eye, while reading sarcastic and non sarcastic sentences. Motivated by this observation, we augment traditional linguistic and stylistic features for sarcasm detection with the cognitive features obtained from readers’ eye movement data. We perform statistical classification using the enhanced feature set so obtained. The augmented cognitive features improve sarcasm detection by 3.7% (in terms of F-score), over the performance of the best reported system.

papers-126

papers-126

We present a novel end-to-end neural model to extract entities and relations between them. Our recurrent neural network based model captures both word sequence and dependency tree substructure information by stacking bidirectional tree-structured LSTM-RNNs on bidirectional sequential LSTM-RNNs. This allows our model to jointly represent both entities and relations with shared parameters in a single model. We further encourage detection of entities during training and use of entity information in relation extraction via entity pretraining and scheduled sampling. Our model improves over the state-of-the-art feature-based model on end-to-end relation extraction, achieving 12.1% and 5.7% relative error reductions in F1-score on ACE2005

and ACE2004, respectively. We also show that our LSTM-RNN based model compares favorably to the state-of-the-art CNN based model (in F1-score) on nominal relation classification (SemEval-2010 Task 8). Finally, we present an extensive ablation analysis of several model components.

papers-036**papers-036**

We present a new proof that O_2 is a multiple context-free language. It contrasts with a recent proof by Salvati (2015) in its avoidance of concepts that seem specific to two-dimensional geometry, such as the complex exponential function. Our simple proof creates realistic prospects of widening the results to higher dimensions. This finding is of central importance to the relation between extreme free word order and classes of grammars used to describe the syntax of natural language.

papers-610**papers-610**

Context is crucial for identifying argumentative relations in text, but many argument mining methods make little use of contextual features. This paper presents context-aware argumentative relation mining that uses features extracted from writing topics as well as from windows of context sentences. Experiments on student essays demonstrate that the proposed features improve predictive performance in two argumentative relation classification tasks.

papers-609**papers-609**

We present several methods for stemming and lemmatization based on discriminative string transduction. We exploit the paradigmatic regularity of semi-structured inflection tables to identify stems in an unsupervised manner with over 85% accuracy. Experiments on English, Dutch and German show that our stemmers substantially outperform Snowball and Morfessor, and approach the accuracy of a supervised model. Furthermore, the generated stems are more consistent than those annotated by experts. Our direct lemmatization model is more accurate than Morfette and Lemming on most datasets. Finally, we test our methods on the data from the shared task on morphological reinflection.

papers-514**papers-514**

A key goal in natural language generation (NLG) is to enable fast generation even with large vocabularies, grammars and worlds. In this work, we build upon a recently proposed NLG system, STRUCT. We describe four enhancements to this system: (i) pruning the grammar based on the world and the communicative goal, (ii) intelligently caching and pruning the combinatorial space of semantic bindings, (iii) reusing the lookahead search tree at different search depths, and (iv) learning and using a search control heuristic. We evaluate the resulting system on three datasets of increasing size and complexity, the largest of which has a vocabulary of about 10K words, a grammar of about 32K lexicalized trees and a world with about 11K entities and 23K relations between them. Our results show that the system has a median generation time of 8.5s and finds the best sentence on average within 25s. These results are based on a sequential, interpreted implementation and are significantly better than the state of the art for planning-based NLG systems.

papers-473**papers-473**

Effective text classification requires experts to annotate data with labels; these training data are time-consuming and expensive to obtain. If you know what labels you want, active learning can reduce the number of labeled documents needed. However, establishing the label set remains difficult. Annotators often lack the global knowledge needed to induce a label set. We introduce ALTO: Active Learning with Topic Overviews, an interactive system to help humans annotate documents: topic models provide a global overview of what labels to create and active learning directs them to the right documents to label. Our user study with forty annotators shows that while active learning by itself is best in extremely resource limited conditions, topic models (even by themselves) lead to better label sets, and ALTO's combination is best overall.

papers-635**papers-635**

Computationally modeling the evolution of science by tracking how scientific topics rise and fall over time has important implications for research funding and public policy. However, little is known about the mechanisms

underlying topic growth and decline. We investigate the role of rhetorical framing: whether the rhetorical role or function that authors ascribe to topics (as methods, as goals, as results, etc.) relates to the historical trajectory of the topics. We train topic models and a rhetorical function classifier to map topic models onto their rhetorical roles in 2.4 million abstracts from the Web of Science from 1991-2010. We find that a topic's rhetorical function is highly predictive of its eventual growth or decline. For example, topics that are rhetorically described as results tend to be in decline, while topics that function as methods tend to be in early phases of growth.

papers-523

papers-523

In this paper, we present the first experiments using neural network models for the task of error detection in learner writing. We perform a systematic comparison of alternative compositional architectures and propose a framework for error detection based on bidirectional LSTMs. Experiments on the CoNLL-14 shared task dataset show the model is able to outperform other participants on detecting errors in learner writing. Finally, the model is integrated with a publicly deployed self-assessment system, leading to performance comparable to human annotators.

papers-230

papers-230

This paper introduces a novel model for semantic role labeling that makes use of neural sequence modeling techniques. Our approach is motivated by the observation that complex syntactic structures and related phenomena, such as nested subordinations and nominal predicates, are not handled well by existing models. Our model treats such instances as sub-sequences of lexicalized dependency paths and learns suitable embedding representations. We experimentally demonstrate that such embeddings can improve results over previous state-of-the-art semantic role labelers, and showcase qualitative improvements obtained by our method.

papers-183

papers-183

Intelligent assistants on mobile devices, such as Siri, have recently gained considerable attention as novel applications of dialogue technologies. A tremendous amount of real users of intelligent assistants provide us with an opportunity to explore a novel task of predicting whether users will continually use their intelligent assistants in the future. We developed prediction models of prospective user engagement by using large-scale user logs obtained from a commercial intelligent assistant. Experiments demonstrated that our models can predict prospective user engagement reasonably well, and outperforms a strong baseline that makes prediction based past utterance frequency.

papers-364

papers-364

A common use of language is to refer to visually present objects. Modelling it in computers requires modelling the link between language and perception. The “words as classifiers” model of grounded semantics views words as classifiers of perceptual contexts, and composes the meaning of a phrase through composition of the denotations of its component words. It was recently shown to perform well in a game-playing scenario with a small number of object types. We apply it to two large sets of real-world photographs that contain a much larger variety of types and for which referring expressions are available. Using a pre-trained convolutional neural network to extract image features, and augmenting these with in-picture positional information, we show that the model achieves performance competitive with the state of the art in a reference resolution task (given expression, find bounding box of its referent), while, as we argue, being conceptually simpler and more flexible.

papers-089

papers-089

Event extraction is a particularly challenging task of information extraction, which intend to identify and classify event trigger and arguments from raw text. In recent works, when deciding the event types (trigger classification), most of the works are either pattern-only or feature-only. However, although patterns cannot cover all representations of an event, it is still a very important feature. In addition, when identifying and classifying arguments, previous works consider each candidate argument separately while ignoring the relation between arguments. In this paper, we propose Regularization-Based Pattern Balancing Method (RPPB).

Inspired by the progress in representation learning, we use trigger embedding, sentence-level embedding and pattern features together as our features for trigger classification. So that the effect of patterns and other useful features can be balanced. In addition, RBPB use regularization method to take advantage of the relation between arguments. Experiments show that we achieved a rather good result than the current state-of-art work. We will release our code later.

papers-836**papers-836**

This paper presents a novel model for Japanese predicate argument structure (PAS) analysis based on a neural network framework. Japanese PAS analysis is challenging due to the tangled characteristics of the Japanese language, such as case disappearance and argument omission. To unravel this problem, we learn selectional preferences from a large raw corpus, and incorporate them into a SOTA PAS analysis model, which considers the consistency of all PASs in a given sentence. We demonstrate that the proposed PAS analysis model significantly outperforms the base SOTA system.

papers-844**papers-844**

We seek to address the lack of labeled data (and high cost of annotation) for textual entailment in some domains. To that end, we first create (for experimental purposes) an entailment dataset for the clinical domain, and a highly competitive supervised entailment system that is effective (out of the box) on two domains. We then explore self-training and active learning strategies to address the lack of labeled data. With self-training, we successfully exploit unlabeled data to improve over the supervised system by 15% F-score on the newswire domain, and by 13% F-score on clinical data. On the other hand, our active learning experiments demonstrate that we can match (and even beat) the supervised system using only 6.6% of the training data in the clinical domain, and only 5.8% of the training data in the newswire domain.

papers-317**papers-317**

The distinction between restrictive and non-restrictive modification in noun phrases is a well studied subject in linguistics. Automatically identifying non-restrictive modifiers can provide NLP applications with shorter, more salient arguments, which were found beneficial by several recent works. While previous work showed that restrictiveness can be annotated with high agreement, no large scale corpus was created, hindering the development of suitable classification algorithms. In this work we devise a novel crowdsourcing annotation methodology, and an accompanying large scale corpus. Then, we present a robust automated system which identifies non-restrictive modifiers, notably improving over prior methods.

papers-102**papers-102**

This study proposes the bilingual segmented topic model (BiSTM), which hierarchically models documents by treating each document as a set of segments, e.g., sections. While previous bilingual topic models, such as bilingual latent Dirichlet allocation (BiLDA) (Mimno et al., 2009; Ni et al., 2009), consider only cross-lingual alignments between entire documents, the proposed model considers cross-lingual alignments between segments in addition to document-level alignments and assigns the same topic distribution to aligned segments. This study also presents a method for simultaneously inferring latent topics and segmentation boundaries, incorporating unsupervised topic segmentation (Du et al., 2013) into BiSTM. Experimental results show that the proposed model significantly outperforms BiLDA in terms of perplexity and demonstrates improved performance in translation pair extraction (up to +0.083 extraction accuracy).

papers-784**papers-784**

This paper connects a vector-based composition model to a formal semantics, the Dependency-based Compositional Semantics (DCS). We show theoretical evidence that the vector compositions in our model conform to the logic of DCS. Experimentally, we show that vector-based composition brings a strong ability to calculate similar phrases as similar vectors, achieving near state-of-the-art on a wide range of phrase similarity tasks and relation classification; meanwhile, DCS can guide building vectors for structured queries that can be directly executed. We evaluate this utility on sentence completion task and report a new state-of-the-art.

papers-019**papers-019**

Attention based recurrent neural networks have shown advantages in representing natural language sentences (Hermann et al., 2015; Rocktaschel et al., 2015; Tan et al., 2015). Based on recurrent neural networks (RNN), external attention information was added to hidden representations to get an attentive sentence representation. Despite the improvement over non-attentive models, the attention mechanism under RNN is not well studied. In this work, we analyze the deficiency of traditional attention based RNN models quantitatively and qualitatively. Then we present three new RNN models that add attention information before RNN hidden representation, which shows advantage in representing sentence and achieves new state-of-art results in answer selection task.

papers-069**papers-069**

Relation classification is a crucial ingredient in numerous information extraction systems seeking to mine structured facts from text. We propose a novel convolutional neural network architecture for this task, relying on two levels of attention in order to better discern patterns in heterogeneous contexts. This architecture enables us to learn directly from small amounts of task-specific labeled data, forgoing the need for external knowledge such as explicit dependency structures. Experiments show that our model outperforms previous state-of-the-art methods, including those relying on much richer forms of prior knowledge.

papers-263**papers-263**

Knowledge bases (KBs) are often greatly incomplete, necessitating a demand for KB completion. The path ranking algorithm (PRA) is one of the most promising approaches to this task. Previous work on PRA usually follows a single-task learning paradigm, building a prediction model for each relation independently with its own training data. It ignores meaningful associations among certain relations, and might not get enough training data for less frequent relations. This paper proposes a novel multi-task learning framework for PRA, referred to as coupled PRA (CPRA). It first devises an agglomerative clustering strategy to automatically discover relations that are highly correlated to each other, and then employs a multi-task learning strategy to effectively couple the prediction of such relations. As such, CPRA takes into account relation association and enables implicit data sharing among them. We empirically evaluate CPRA on benchmark data created from Freebase. Experimental results show that CPRA can effectively identify coherent clusters in which relations are highly correlated. By further coupling such relations, CPRA significantly outperforms PRA, in terms of both predictive accuracy and model interpretability.

papers-501**papers-501**

In this work, we propose a novel method to incorporate corpus-level discourse information into language modelling. We call this larger-context language model. We introduce a late fusion approach to a recurrent language model based on long short-term memory units (LSTM), which helps the LSTM unit keep intra-sentence dependencies and inter-sentence dependencies separate from each other. Through the evaluation on three corpora (IMDB, BBC, and Penn TreeBank), we demonstrate that the proposed model improves perplexity significantly. In the experiments, we evaluate the proposed approach while varying the number of context sentences and observe that the proposed late fusion is superior to the usual way of incorporating additional inputs to the LSTM. By analyzing the trained larger-context language model, we discover that content words, including nouns, adjectives and verbs, benefit most from an increasing number of context sentences. This analysis suggests that larger-context language model improves the unconditional language model by capturing the theme of a document better and more easily.

papers-569**papers-569**

Website privacy policies are often ignored by Internet users, because these documents tend to be long and difficult to understand. However, the significance of privacy policies greatly exceeds the attention paid to them: these documents are binding legal agreements between website operators and their users, and their opaqueness is a challenge not only to Internet users but also to policy regulators. One proposed alternative to the status quo is to automate or semi-automate the extraction of salient details from privacy policy text, using a combination of crowdsourcing, natural language processing, and machine learning. However, there has been a relative

dearth of datasets appropriate for identifying data practices in privacy policies. To remedy this problem, we introduce a corpus of 115 privacy policies (267K words) with manual annotations for 23K fine-grained data practices. We describe the process of using skilled annotators and a purpose-built annotation tool to produce the data. We provide findings based on a census of the annotations and show results toward automating the annotation procedure. Finally, we describe challenges and opportunities for the research community to use this corpus to advance research in both privacy and language technologies.

papers-510**papers-510**

We propose an approach for semantic parsing that uses a recurrent neural network to map a natural language question into a logical form representation of a KB query. Building on recent work by (Wang et al., 2015), the interpretable logical forms, which are structured objects obeying certain constraints, are enumerated by an underlying grammar and are paired with their canonical realizations. In order to use sequence prediction, we need to sequentialize these logical forms. We compare three sequentializations: a direct linearization of the logical form, a linearization of the associated canonical realization, and a sequence consisting of derivation steps relative to the underlying grammar. We also show how grammatical constraints on the derivation sequence can easily be integrated inside the RNN-based sequential predictor. Our experiments show important improvements over previous results for the same dataset, and also demonstrate the advantage of incorporating the grammatical constraints.

papers-206**papers-206**

Word embeddings – distributed representations of words – in deep learning are beneficial for many tasks in NLP. However, different embedding sets vary greatly in quality and characteristics of the captured information. Instead of relying on a more advanced algorithm for embedding learning, this paper proposes an ensemble approach of combining different public embedding sets with the aim of learning meta-embeddings. Experiments on word similarity and analogy tasks and on part-of-speech tagging show better performance of meta-embeddings compared to individual embedding sets. One advantage of meta-embeddings is the increased vocabulary coverage. We release our meta-embeddings publicly at <http://cistern.cis.lmu.de/meta-emb>.

papers-887**papers-887**

In this paper, we investigate the possibility to automatically generate sports news from live text commentary scripts. As a preliminary study, we treat this task as a special kind of document summarization based on sentence extraction. We formulate the task in a supervised learning to rank framework, utilizing both traditional sentence features for generic document summarization and novelly designed task-specific features. To tackle the problem of local redundancy, we also propose a probabilistic sentence selection algorithm. Experiments on our collected data from football live commentary scripts and corresponding sports news demonstrate the feasibility of this task. Evaluation results show that our methods are indeed appropriate for this task, outperforming several baseline methods in different aspects.

papers-311**papers-311**

One of the major challenges for statistical machine translation (SMT) is to choose the appropriate translation rules based on the sentence context. This paper proposes a continuous space rule selection (CSRS) model for syntax-based SMT to perform this context-dependent rule selection. In contrast to existing maximum entropy based rule selection (MERS) models, which use discrete representations of words as features, the CSRS model is learned by a feed-forward neural network and uses real-valued vector representations of words, allowing for better generalization. In addition, we propose a method to train the rule selection models only on minimal rules, which are more frequent and have richer training data compared to non-minimal rules. We tested our model on different translation tasks and the CSRS model outperformed a baseline without rule selection and the previous MERS model by up to 2.2 and 1.1 points of BLEU score respectively.

papers-052**papers-052**

This paper presents neural probabilistic parsing models which explore up to third-order graph-based parsing with maximum likelihood training criteria. Two neural network extensions are exploited for performance

improvement. Firstly, a convolutional layer that absorbs the influences of all words in a sentence is used so that sentence-level information can be effectively captured. Secondly, a linear layer is added to integrate different order neural models and trained with perceptron method. The proposed parsers are evaluated on English and Chinese Penn Treebanks and obtain competitive accuracies.

papers-433

papers-433

We propose a novel reranking method to extend a deterministic neural dependency parser. Different to conventional k-best reranking, the proposed model integrates search and learning by utilizing a dynamic action revising process, using the reranking model to guide modification for the base outputs and to rerank the candidates. The dynamic reranking model achieves an absolute 1.78% accuracy improvement over the deterministic baseline parser on PTB, which is the highest improvement by neural rerankers in the literature.

papers-405

papers-405

Cross-lingual sentiment classification aims to adapt the sentiment resource in a resource-rich language to a resource-poor language. In this study, we propose a representation learning approach which simultaneously learns vector representations for the texts in both the source and the target languages. Different from previous research which only gets bilingual word embedding, our Bilingual Document Representation Learning model BiDRL directly learns document representations. Both semantic and sentiment correlations are utilized to map the bilingual texts into the same embedding space. The experiments are based on the multilingual multi-domain Amazon review dataset. We use English as the source language and use Japanese, German and French as the target languages. The experimental results show that BiDRL outperforms the state-of-the-art methods by a large margin for all the target languages.

papers-323

papers-323

Most of the sequence tagging tasks in natural language processing require to recognize segments with certain syntactic role or semantic meaning in a sentence. They are usually tackled with Conditional Random Fields (CRFs), which do indirect word-level modeling over word-level features and thus cannot make full use of segment-level information. Semi-Markov Conditional Random Fields (Semi-CRFs) model segments directly but extracting segment-level features for Semi-CRFs is still a very challenging problem. This paper presents Gated Recursive Semi-CRFs (grSemi-CRFs), which model segments directly and automatically learn segment-level features through a gated recursive convolutional neural network. Our experiments on text chunking and named entity recognition (NER) demonstrate that grSemi-CRFs generally outperform other neural models.;

Posters: short papers

short-297

short-297

Low-dimensional vector representations are widely used as stand-ins for the text of words, sentences, and entire documents. These embeddings are used to identify similar words or make predictions about documents. In this work, we consider embeddings for social media users and demonstrate that these can be used to identify users who behave similarly or to predict attributes of users. In order to capture information from all aspects of a user's online life, we take a multiview approach, applying a weighted variant of Generalized Canonical Correlation Analysis (GCCA) to a collection of over 100,000 Twitter users. We demonstrate the utility of these multiview embeddings on three downstream tasks: user engagement, friend selection, and demographic attribute prediction.

short-159

short-159

This paper deals with a double-implicit problem in opinion mining and sentiment analysis. We aim at identifying aspects and polarities of opinionated statements not consisting of opinion words and aspect terms. As a case study, opinion words and aspect terms are first extracted from Chinese hotel reviews, and then grouped into positive (negative) clusters and aspect term clusters. We observe that an implicit opinion and its neighbor

explicit opinion tend to have the same aspect and polarity. Under the observation, we construct an implicit opinions corpus annotated with aspect class labels and polarity automatically. Aspect and polarity classifiers trained by using this corpus is used to recognize aspect and polarity of implicit opinions.

short-218**short-218**

Finding domain invariant features is critical for successful domain adaptation and transfer learning. However, in the case of unsupervised adaptation, there is a significant risk of overfitting on source training data. Recently, a regularization for domain adaptation was proposed for deep models by Ganin and Lempitsky. We build on their work by suggesting a more appropriate regularization for denoising autoencoders. Our model remains unsupervised and can be computed in closed form. On standard text classification adaptation tasks, our approach yield state of the art results, with an important reduction of the learning cost.

short-521**short-521**

Recently, neural network approaches for parsing have largely automated the combination of individual features, but still rely on atomic features created from human linguistic intuition, and potentially omitting important global context. To further reduce feature engineering, we use bi-directional LSTM sentence representations and model a parser state with only three sentence positions, which allows the model to automatically identify important aspects of the entire sentence. This model achieves state-of-the-art results among greedy dependency parsers for English. We also introduce a novel transition system for constituency parsing which does not require binarization, and together with the above architecture, achieves state-of-the-art results among greedy parsers for both English and Chinese.

short-242**short-242**

We present a novel technique for training translation models for statistical machine translation by aligning source sentences to their oracle-BLEU translations. In contrast to previous approaches which are constrained to phrase training, our method also allows the re-estimation of re-ordering models along with the translation model. Experiments show an improvement of up to 0.8 BLEU for our approach over a competitive Arabic-English baseline trained directly on the word-aligned bitext using heuristic extraction. As an additional benefit, the phrase table size is reduced dramatically to only 3% of the original size.

short-430**short-430**

We present a natural language generator based on the sequence-to-sequence approach that can be trained to produce natural language strings as well as deep syntax dependency trees from input dialogue acts, and we use it to directly compare two-step generation with separate sentence planning and surface realization stages to a joint, one-step approach. We were able to train both setups successfully using very little training data. The joint setup offers better performance, surpassing state-of-the-art with regards to n-gram-based scores while providing more relevant outputs.

short-208**short-208**

We consider two graph models of semantic change. The first is a time-series model that relates embedding vectors from one time period to embedding vectors of previous time periods. In the second, we construct one graph for each word: nodes in this graph correspond to time points and edge weights to the similarity of the word's meaning across two time points. We apply our two models to corpora across three different languages. We find that semantic change is linear in two senses. Firstly, today's embedding vectors (= meaning) of words can be derived as linear combinations of embedding vectors of their neighbors in previous time periods. Secondly, self-similarity of words decays linearly in time. We consider both findings as new laws/hypotheses of semantic change.

short-287**short-287**

We describe a model which jointly performs word segmentation and induces vowel categories from formant values. Vowel induction performance improves slightly over a baseline model which does not segment; segmentation performance decreases slightly from a baseline using entirely symbolic input. Our high joint perfor-

mance in this idealized setting implies that problems in unsupervised speech recognition reflect the phonetic variability of real speech sounds in context.

short-130

short-130

Event detection remains a challenge due to the difficulty at encoding the word semantics in various contexts. Previous approaches heavily depend on language-specific knowledge and pre-existing natural language processing (NLP) tools. However, compared to English, not all languages have such resources and tools available. A more promising approach is to automatically learn effective features from data, without relying on language-specific resources. In this paper, we develop a hybrid neural network to capture both sequence and chunk information from specific contexts, and use them to train an event detector for multiple languages without any manually encoded features. Experiments show that our approach can achieve robust, efficient and accurate results for multiple languages (English, Chinese and Spanish).

short-031

short-031

Syntactic parsers perform poorly in prediction of Argument-Cluster Coordination (ACC). We change the PTB representation of ACC to be more suitable for learning by a statistical PCFG parser, affecting 125 trees in the training set. Training on the modified trees yields a slight improvement in EVALB scores on sections 22 and 23. The main evaluation is on a corpus of 4th grade science exams, in which ACC structures are prevalent. On this corpus, we obtain an impressive *times* 2.7 improvement in recovering ACC structures compared to a parser trained on the original PTB trees.

short-352

short-352

In the translation industry, human translations are assessed by comparison with the source texts. In the Machine Translation (MT) research community, however, it is a common practice to perform quality assessment using a reference translation instead of the source text. In this paper we show that this practice has a serious issue - annotators are strongly biased by the reference translation provided, and this can have a negative impact on the assessment of MT quality.

short-313

short-313

This paper presents a cross-lingual projection technique for training class-based language models. We borrow from previous success in projecting POS tags and NER mentions to that of a trained class-based language model. We use a CRF to train a model to predict when a sequence of words is a member of a given class and use this to label our language model training data. We show that we can successfully project the contextual cues for these classes across pairs of languages and retain a high quality class model in languages with no supervised class data. We present empirical results that show the quality of the projected models as well as their effect on the down-stream speech recognition objective. We are able to achieve over 70% of the WER reduction when using the projected class models as compared to models trained on human annotations.

short-473

short-473

Retrieving similar short texts is crucial to many applications, e.g., web search, microblogging, ads matching, and so forth. Most of the traditional methods concentrate on how to improve the precision of the similarity measurement, while current real applications need to efficiently explore the top similar short texts semantically related to the query one. In this paper, we address the efficiency issue by investigating the similarity strategies and incorporating them into the FAST framework (efficient FrAmework for semantic similar Short Texts retrieval). We conduct comprehensive performance evaluation on real-life data which shows that our proposal outperforms the state-of-the-art techniques.

short-283

short-283

This paper presents a parsing algorithm that can jointly detect empty elements. This method uses the spinal Tree Adjoining Grammar formalism of (Carreras, et. al. 08). In this paper, we define a transition-based algorithm together with several operations to combine the spinal elementary trees for constructing a full parse tree with empty elements. We conduct experiments on the English Penn and the Japanese Keyaki Treebank

datasets. The experimental results show that our proposed method outperforms the previous systems.

short-340**short-340**

We investigate the effectiveness of semantic generalizations/classifications for capturing the regularities of the behavior of verbs in terms of their metaphoricality. Starting from orthographic word unigrams, we experiment with various ways of defining semantic classes for verbs (grammatical, resource-based, distributional) and measure the effectiveness of these classes for classifying all verbs in a running text as metaphor or non-metaphor.

short-391**short-391**

Over the past decade, e-Commerce has rapidly grown enabling customers to purchase products with the click of a button. But to be able to do so, one has to understand the semantics of a user query and identify that in “digital lifestyle tv”, “digital lifestyle” is a brand and “tv” is a product. In this paper, we develop a series of structured prediction algorithms for semantic tagging of shopping queries with the “product”, “brand”, “model” and “product family” types. We model wide variety of features and show an alternative way to capture knowledge base information using embeddings. We conduct an extensive study over 37,000 manually annotated queries and report performance of 90.92 F1 independent of the query length.

short-363**short-363**

Recent work in learning vector-space embeddings for multi-relational data has focused on combining relational information derived from knowledge bases with distributional information derived from large text corpora. We propose a simple approach that leverages the descriptions of entities or phrases available in lexical resources, in conjunction with distributional semantics, in order to derive a better initialization for training relational models. Applying this initialization to the TransE model results in significant new state-of-the-art performances on the WordNet dataset, decreasing the mean rank from the previous best 212 to 51. It also results in faster convergence of the entity representations, and achieves small improvements on Freebase for raw mean rank. We find that there is a trade-off between improving the mean rank and the hits10 with this approach. This illustrates that much remains to be understood regarding performance improvements in relational models.

short-392**short-392**

In unsupervised semantic role labeling, identifying the role of an argument is usually informed by its dependency relation with the predicate. In this work, we propose a neural model to learn argument embeddings from the context by explicitly incorporating dependency relations as multiplicative factors, which bias argument embeddings according to their dependency roles. Our model outperforms existing state-of-the-art embeddings in unsupervised semantic role induction on the CoNLL 2008 dataset. Qualitative results demonstrate our model can effectively bias argument embeddings based on their dependency role.

short-085**short-085**

In order to capture rich language phenomena, neural machine translation models have to use a large vocabulary size, which requires high computing time and large memory usage. In this paper, we alleviate this issue by introducing a sentence-level or batch-level vocabulary, which is only a very small sub-set of the full output vocabulary. For each sentence or batch, we only predict the target words in its sentence-level or batch-level vocabulary. Thus, we reduce both the computing time and the memory usage. Our method simply takes into account the translation options of each word or phrase in the source sentence, and picks a very small target vocabulary for each sentence %or mini-batch based on a word-to-word translation model or a bilingual phrase library learned from a traditional machine translation model. Experimental results on the large-scale English-to-French task show that our method achieves better translation performance by 1 BLEU point over the large vocabulary neural machine translation system of Jean et al. (2015).

short-395**short-395**

In this paper, we propose the TBCNN-pair model to recognize entailment and contradiction between two sentences. In our model, a tree-based convolutional neural network (TBCNN) captures sentence-level semantics;

then heuristic matching layers like concatenation, element-wise product/difference combine the information in individual sentences. Experimental results show that our model outperforms existing sentence encoding-based approaches by a large margin.

short-529

short-529

In this paper we improve over the hierarchical Pitman-Yor processes language model in a cross-domain setting by adding skipgrams as features. We find that adding skipgram features reduces the perplexity. This reduction is substantial when models are trained on a generic corpus and tested on domain-specific corpora. We also find that within-domain testing and cross-domain testing require different backoff strategies. We observe a 30-40% reduction in perplexity in a cross-domain language modelling task, and up to 6% reduction in a within-domain experiment, for both English and Flemish-Dutch.

short-245

short-245

We release the Simple Paraphrase Database, a subset of the Paraphrase Database (PPDB) adapted for the task of text simplification. We train a supervised model to associate simplification scores with each phrase pair, producing rankings competitive with state-of-the-art lexical simplification models. Our new simplification database contains 4.5 million paraphrase rules, making it the largest available resource for lexical simplification.

short-322

short-322

Named entity recognition, and other information extraction tasks, frequently use linguistic features such as part of speech tags or chunkings. For languages where word boundaries are not readily identified in text, word segmentation is a key first step to generating features for an NER system. While using word boundary tags as features are helpful, the signals that aid in identifying these boundaries may provide richer information for an NER system. New state-of-the-art word segmentation systems use neural models to learn representations for predicting word boundaries. We show that these same representations, jointly trained with an NER system, yield significant improvements in NER for Chinese social media. In our experiments, jointly training NER and word segmentation with an LSTM-CRF model yields nearly 5% absolute improvement over previously published results.

short-299

short-299

We introduce a new multilingual resource containing judgments about nominal compound compositionality in English, French and Portuguese. It covers 3 x 180 noun-noun and adjective-noun compounds for which we provide numerical compositionality scores for the head word, for the modifier and for the compound as a whole, along with possible paraphrases. This resource was constructed by native speakers via crowdsourcing. It can serve as basis for evaluating tasks such as lexical substitution and compositionality prediction.

short-052

short-052

We present an open web platform for developing, compiling, and running rule-based speech to sign language translation applications. Speech recognition is performed using the Nuance Recognizer 10.2 toolkit, and signed output, including both manual and non-manual components, is rendered using the JASigning avatar system. The platform is designed to make the component technologies readily accessible to sign language experts who are not necessarily computer scientists. Translation grammars are written in a version of Synchronous Context-Free Grammar adapted to the peculiarities of sign language. All processing is carried out on a remote server, with content uploaded and accessed through a web interface. Initial experiences show that simple translation grammars can be implemented on a time-scale of a few hours to a few days and produce signed output readily comprehensible to Deaf informants. Overall, the platform drastically lowers the barrier to entry for researchers interested in building applications that generate high-quality signed language.

short-344

short-344

In word alignment certain source words are only needed for fluency reasons and do not have a translation on the target side. Most word alignment models assume a target NULL word from which they generate these untranslatable source words. Hypothesising a target NULL word is not without problems, however. For

example, because this NULL word has a position, it interferes with the distribution over alignment jumps. We present a word alignment model that accounts for untranslatable source words by generating them from preceding source words. It thereby removes the need for a target NULL word and only models alignments between word pairs that are actually observed in the data. Translation experiments on English paired with Czech, German, French and Japanese show that the model outperforms its traditional IBM counterparts in terms of BLEU score.

short-109**short-109**

This work explores the use of unsupervised morph segmentation along with statistical language models for the task of vocabulary expansion. Unsupervised vocabulary expansion has large potential for improving vocabulary coverage and performance in different natural language processing tasks, especially in less-resourced settings on morphologically rich languages. We propose a combination of unsupervised morph segmentation and statistical language models and evaluate on languages from the Babel corpus. The method is shown to perform well for all the evaluated languages when compared to the previous work on the task.

short-126**short-126**

Here we seek to automatically identify Hungarian patients suffering from mild cognitive impairment (MCI) based on linguistic features collected from their speech transcripts. Our system uses machine learning techniques and is based on several linguistic features like characteristics of spontaneous speech as well as features exploiting morphological and syntactic parsing. Our results suggest that it is primarily morphological and speech-based features that help distinguish MCI patients from healthy controls.

short-316**short-316**

Recent work has revealed the potential of using visual representations for bilingual lexicon learning (BLL). Such image-based BLL methods, however, still fall short of linguistic approaches. In this paper, we propose a simple yet effective multi-modal approach that learns bilingual semantic representations that fuse linguistic and visual input. These new bilingual multi-modal embeddings display significant performance gains in the BLL task for three language pairs on two benchmarking test sets, outperforming linguistic-only BLL models using three different types of state-of-the-art bilingual word embeddings, as well as visual-only BLL models.

short-487**short-487**

In this paper we study how to identify persuasive posts in online forum discussions, using data from Change My View sub-Reddit. Our analysis confirms that the users' voting score for a comment is highly correlated with its metadata information such as published time and author reputation. In this work, we propose and evaluate other features to rank comments based on their persuasive scores, including textual information in the comment and social interaction related features. Our experiments show that the surface textual features do not perform well compared to the argumentation based features, and the social interaction based features are effective especially when more users participate in the discussion.

short-507**short-507**

We demonstrate the value of collecting semantic parse labels for knowledge base question answering. In particular, (1) unlike previous studies on small-scale datasets, we show that learning from labeled semantic parses significantly improves overall performance, resulting in absolute 5 point gain compared to learning from answers, (2) we show that with an appropriate user interface, one can obtain semantic parses with high accuracy and at a cost comparable or lower than obtaining just answers, and (3) we have created and shared the largest semantic-parse labeled dataset to date in order to advance research in question answering.

short-074**short-074**

Relation classification is an important semantic processing task in the field of natural language processing (NLP). State-of-the-art systems still rely on lexical resources such as WordNet or NLP systems like dependency parser and named entity recognizers (NER) to get high-level features. Another challenge is that important information can appear at any position in the sentence. To tackle these problems, we propose Attention-Based Bidirectional Long Short-Term Memory Networks(Att-BLSTM) to capture the most important semantic

information in a sentence. The experimental results on the SemEval-2010 relation classification task show that our method outperforms most of the existing methods, with only word vectors.

Posters: TACL papers

tacl-923

tacl-923

We present a model of unsupervised phonological lexicon discovery—the problem of simultaneously learning phoneme-like and word-like units from acoustic input. Our model builds on earlier models of unsupervised phone-like unit discovery from acoustic data (Lee and Glass, 2012), and unsupervised symbolic lexicon discovery using the Adaptor Grammar framework (Johnson et al., 2006), integrating these earlier approaches using a probabilistic model of phonological variation. We show that the model is competitive with state-of-the-art spoken term discovery systems, and present analyses exploring the model’s behavior and the kinds of linguistic structures it learns.

Posters: SRW

srw-009

srw-009

We propose a new dataset for evaluating a Japanese lexical simplification method. Previous datasets have several deficiencies. All of them substitute only a single target word, and some of them extract sentences only from newswire corpus. In addition, most of these datasets do not allow ties and integrate simplification ranking from all the annotators without considering the quality. In contrast, our dataset has the following advantages: (1) It is the first controlled and balanced dataset for Japanese lexical simplification with high correlation with human judgement and (2) the consistency of the simplification ranking is improved by allowing candidates to have ties and by considering the reliability of annotators.

srw-010

srw-010

A hierarchical word alignment model that searches for k-best partial alignments on target constituent 1-best parse trees has been shown to outperform previous models. However, relying solely on 1-best parses trees might hinder the search for good alignments because 1-best trees are not necessarily the best for word alignment tasks in practice. This paper introduces a dependency forest based word alignment model, which utilizes target dependency forests in an attempt to minimize the impact on limitations attributable to 1-best parse trees. We present how k-best alignments are constructed over target-side dependency forests. Alignment experiments on the Japanese-English language pair show a relative error reduction of 4% of the alignment score compared to a model with 1-best parse trees.

srw-018

srw-018

This paper presents an application of the meta-bootstrapping algorithm from Riloff and Jones 1999 as a means to handle difficulties associated with non-standard language usage in the social media domain when tweets were analyzed for potentially containing adverse drug events. The paper details the importance of adverse event detection, presents a large new dataset, and steps through several different evaluations exploring the effects of incrementing the amount of unannotated social media data given to the bootstrapping procedure. Modifications of the original algorithm are discussed in addition to highlighting several important examples of how meta-bootstrapped lexicons can handle internet language use within the clinical domain in a way studies using medical ontologies cannot.

srw-019

srw-019

Quantitative analysis of human brain activity based on language representations, such as the semantic categories of words, have been actively studied in the field of brain and neuroscience. Our study aims to generate natural language descriptions for human brain activation phenomena caused by visual stimulus by employing

deep learning methods, which have gained interest as an effective approach to automatically describe natural language expressions for various type of multi-modal information, such as images. We employed an image-captioning system based on a deep learning framework as the basis for our method by learning the relationship between the brain activity data and the features of an intermediate expression of the deep neural network owing to lack of training brain data. We conducted three experiments and were able to generate natural language sentences which enabled us to quantitatively interpret brain activity.

srw-021**srw-021**

Works on Twitter community detection have yielded new ways to extract valuable insights from social media. Through this technique, Twitter users can be grouped into different types of communities such as those who have the same interests, those who interact a lot, or those who have similar sentiments about certain topics. Computationally, information is represented as a graph, and community detection is the problem of partitioning the graph such that each community is more densely connected to each other than to the rest of the network. It has been shown that incorporating sentiment analysis can improve community detection when looking for sentiment-based communities. However, such works only perform sentiment analysis in isolation without considering the tweet's various contextual information. Examples of these contextual information are social network structure, and conversational, author, and topic contexts. Disregarding these information poses a problem because at times, context is needed to clearly infer the sentiment of a tweet. Thus, this research aims to improve detection of sentiment-based communities on Twitter by performing contextual sentiment analysis.

srw-024**srw-024**

This paper evaluates the challenges involved in shallow parsing of Dravidian languages which are highly agglutinative and morphologically rich. Text processing tasks in these languages are not trivial because multiple words concatenate to form a single string with morpho-phonemic changes at the point of concatenation. This phenomenon known as Sandhi, in turn complicates the individual word identification. Shallow parsing is the task of identification of correlated group of words given a raw sentence. The current work is an attempt to study the effect of Sandhi in building shallow parsers for Dravidian languages by evaluating its effect on Malayalam, one of the main languages from Dravidian family. We provide an in-depth analysis of effect of Sandhi in developing a robust shallow parser pipeline with experimental results emphasizing on how sensitive the individual components of shallow parser are, towards the accuracy of a sandhi splitter. Our work can serve as a guiding light for building robust text processing systems in Dravidian language.

srw-027**srw-027**

Online newspaper articles can accumulate comments at volumes that prevent close reading. Summarisation of the comments allows interaction at a higher level and can lead to an understanding of the overall discussion. Comment summarisation requires topic clustering, comment ranking and extraction. Clustering must be robust as the subsequent extraction relies on a good set of clusters. Comment data, as with many social media datasets, contains very short documents and the number of words in the documents is a limiting factors on the performance of LDA clustering. We evaluate whether we can combine comments to form larger documents to improve the quality of clusters. We find that combining comments with comments that reply to them produce the highest quality clusters.

srw-029**srw-029**

Arab users sometimes tend to express their natural dialectal Arabic voice in text using a Latin script on social media. This form of transliterated Arabic is called Arabizi. Extracting sentiment from Arabizi is considered complicated for the following reasons: 1. Arabizi is dialectal, differing among regions. 2. Arabizi users do not follow a unified orthography nor grammar. 3. To the best of our knowledge, there are no public lexical resources to process this data. Several researchers who worked on sentiment analysis for Arabic filtered Arabizi from their datasets. As this data might contain valuable information, we first investigate how frequent is Arabizi on Twitter to show the amount of Arabizi data being filtered out. Second we study which methods could be used to create Arabizi identification tools. We believe that identifying Arabizi from multilingual data brings us a step forward towards analysing sentiment from Arabic social media.

srw-030

srw-030

Previous optimisations of parameters affecting the word-context association measure used in distributional vector space models have focused either on high-dimensional vectors with hundreds of thousands of dimensions, or dense vectors with dimensionality of few hundreds; but dimensionality of few thousands is often applied in compositional tasks as it is still computationally feasible and does not require the dimensionality reduction step. We present a systematic study of the interaction of the parameters of the association measure and vector dimensionality, and derive parameter selection heuristics that achieve performance across word similarity and relevance datasets competitive with the results previously reported in the literature achieved by highly dimensional or dense models.

srw-031

srw-031

Singleton (or non-coreferential) mentions are a problem for coreference resolution systems, and identifying singletons before mentions are linked improves resolution performance. Here, a singleton detection system based on word embeddings and neural networks is presented, which achieves state-of-the-art performance (79.6% accuracy) on the CoNLL-2012 shared task development set. Extrinsic evaluation with the Stanford and Berkeley coreference resolution systems shows significant improvement for the first, but not for the latter. The results show the potential of using neural networks and word embeddings for improving both singleton detection and coreference resolution.

srw-032

srw-032

We present a new, sizeable dataset of noun–noun compounds with their syntactic analysis (bracketing) and semantic relations. Derived from several established linguistic resources, such as the Penn Treebank, our dataset enables experimenting with new approaches towards a holistic analysis of noun–noun compounds, such as joint-learning of noun–noun compounds bracketing and interpretation, as well as integrating compound analysis with other tasks such as syntactic parsing.

srw-037

srw-037

The idea behind this proposal is to investigate the possibility of utilizing NLP tools, statistical topic modelling techniques and freely available online resources to propose a system able to provide dialogue contribution suggestions which are relevant to the context, yet out of the main activity of the dialogue (i.e. off-activity talk). The aim is to evaluate the effects of a tool that automatically suggests off-activity talks in form of some sentences relevant to the dialogue context. The evaluation is to be done over two test-sets of open domain and closed-domain in a conversational quiz-like setting. The outcome of this work will be a satisfactory point of entry to investigate the hypothesis that adding automatically generated off-activity talks feature to a conversational agent can lead to building up engagement of the dialogue partner(s).

srw-038

srw-038

We propose a method for improving the dependency parsing of complex sentences. This method assumes segmentation of input sentences into clauses and does not require to re-train a parser of one's choice. We represent a sentence clause structure using clause charts that provide a layer of embedding for each clause in the sentence. Then we formulate a parsing strategy as a two-stage process where (1) coordinated and subordinated clauses of the sentence are parsed separately with respect to the sentence clause chart and (2) their dependency trees become subtrees of the final tree of the sentence. The object language is Czech and the parser used is a maximum spanning tree parser trained on the Prague Dependency Treebank. We have achieved an average 0.97% improvement in the unlabeled attachment score. Although the method has been designed for the dependency parsing of Czech, it is useful for other parsing techniques and languages.

srw-039

srw-039

The computing cost of many NLP tasks increases faster than linearly with the length of a sentence. In this paper we propose sentence chunking as a new task which simplifies further processing of long sentences. We use rules referring to the link structure of the sentence's Dependency Minimal Recursion Semantics (DMRS) representation to find appropriate chunking points in the graph representation and then to map those onto the sentence surface. Based on our preliminary experiments, we report an improvement in the precision of chunk-

ing from 19.6% to 42% compared with string heuristics. Currently we are working on using the rule-based system to generate a dataset of sufficient quality to be used as training material for a minimally supervised machine learning system. Then sentences could be chunked without creating a full parse first and we could apply chunking to an extended set of NLP tasks, such as statistical machine translation or parsing itself.

srw-040**srw-040**

The availability of large document-summary corpora have opened up new possibilities for using statistical text generation techniques for abstractive summarization. Progress in Extractive text summarization has become stagnant for a while now and in this work we compare the two possible alternates to it. We present an argument in favor of abstractive summarization compared to an ensemble of extractive techniques. Further we explore the possibility of using statistical machine translation as a generative text summarization technique and present possible research questions in this direction. We also report our initial findings and future direction of research.

srw-047**srw-047**

Sarcasm can radically alter or invert a phrase's meaning. Sarcasm detection can therefore help improve natural language processing (NLP) tasks. However, the majority of prior research has treated sarcasm detection as classification, with three important limitations: 1. Balanced datasets, when sarcasm is actually rather rare. 2. Using Twitter users' self-declarations in the form of hashtags to label data, when sarcasm can take many forms. To address these issues, we create an unbalanced corpus of manually annotated Twitter conversations. We compare human and machine ability to recognize sarcasm on this data under varying amounts of context. Results indicate that both class imbalance and labelling method affect performance, and are factors that should be considered when designing automatic sarcasm detection systems. We conclude that for progress to be made in real-world sarcasm detection, we will require a new class labelling scheme that is able to access the 'common ground' held between conversational parties.

srw-050**srw-050**

This paper proposes a new unsupervised technique for clustering a collection of documents written by distinct individuals into authorial components. We highlight the importance of utilizing syntactic structure to cluster documents by author, and demonstrate experimental results that show the method we outline performs on par with state-of-the-art techniques. Additionally, we argue that this feature set outperforms previous methods in cases where authors consciously emulate each other's style or are otherwise rhetorically similar.

srw-052**srw-052**

In addition to the sentiments expressed by a speaker, opinions on the web also convey suggestions. Such text comprise of advice, recommendations and tips on a variety of points of interest. We propose that suggestions can be extracted from the available opinionated text and put to several use cases. The problem has been identified only recently as a viable task, and there is a lot of scope for research in the direction of problem definition, datasets, and methods. From an abstract view, standard algorithms for tasks like sentence classification and keyphrase extraction appear to be usable for suggestion mining. However, initial experiments reveal that there is a need for new methods, or variations in the existing ones for addressing the problem specific challenges. We present a research proposal which divides the problem into three main research questions. We walk through the research questions, presenting our findings, results and future directions.

srw-056**srw-056**

In this paper, we propose a cross-lingual convolutional neural network (CNN) model that is based on word and phrase embeddings learned from unlabeled data in two languages and dependency grammar. Compared to traditional machine translation (MT) based methods for cross lingual sentence modeling, our model is much simpler and does not need parallel corpora or language specific features. We only use a bilingual dictionary and dependency parser. This makes our model particularly appealing for resource poor languages. We evaluate our model using English and Chinese data on several sentence classification tasks. We show that our model achieves a comparable and even better performance than the traditional MT-based method.

srw-057**srw-057**

This paper introduces a new corpus, QA-It, for the classification of non-referential it. Our dataset is unique in a sense that it is annotated on question answer pairs collected from multiple genres, useful for developing advanced QA systems. Our annotation scheme makes clear distinctions between 4 types of it, providing guidelines for many erroneous cases. Several statistical models are built for the classification of it, showing encouraging results. To the best of our knowledge, this is the first time that such a corpus is created for question answering.

srw-058

srw-058

In this research, we build a Wikification corpus for advancing Japanese Entity Linking. This corpus consists of 340 Japanese newspaper articles with 25,675 entity mentions. All entity mentions are labeled by a fine-grained semantic classes (200 classes), and 19,121 mentions were successfully linked to Japanese Wikipedia articles. Even with the fine-grained semantic classes, we found it hard to define the target of entity linking annotations and to utilize the fine-grained semantic classes to improve the accuracy of entity linking.

srw-060

srw-060

Text categorization is an important task in current natural language processing pipelines. However, most works in this area are focused on Western languages ignoring other Semitic languages like Arabic. These languages are of immense political and social importance necessitating robust categorization techniques. In this paper, we present a novel three-stage technique to efficiently classify Arabic documents into different categories based on the words they contain. We leverage the significance of root-words in Arabic and incorporate a combination of Markov clustering and Deep Belief networks to classify Arabic words into separate groups (clusters). Our approach is tested on two public datasets giving a F-Measure of 91%.

Demonstrations A

demos-030

demos-030

Semantic role labeling (SRL) identifies the predicate-argument structure in text with semantic labels. It plays a key role in understanding natural language. In this paper, we present POLYGLOT, a multilingual semantic role labeling system capable of semantically parsing sentences in 9 different languages from 4 different language groups. The core of POLYGLOT are SRL models for individual languages trained with automatically generated Proposition Banks (Akbik et al., 2015). The key feature of the system is that it treats the semantic labels of the English Proposition Bank as “universal semantic labels”: Given a sentence in any of the supported languages, POLYGLOT applies the corresponding SRL and predicts English PropBank frame and role annotation. The results are then visualized to facilitate the understanding of multilingual SRL with this unified semantic representation.

demos-061

demos-061

The reading material used in a language learning classroom should ideally be rich in terms of the grammatical constructions and vocabulary to be taught and in line with the learner’s interests. We developed an online Information Retrieval system that helps teachers search for texts appropriate in form, content, and reading level. It identifies the 87 grammatical constructions spelled out in the official English language curriculum of schools in Baden-Württemberg, Germany. The tool incorporates a classical efficient algorithm for reranking the results by assigning weights to selected constructions and prioritizing the documents containing them. Supplemented by an interactive visualization module, it allows for a multi-faceted presentation and analysis of the retrieved documents.

demos-011

demos-011

We introduce, TermSuite, a JAVA and UIMA-based toolkit to build terminologies from corpora. TermSuite follows the classic two steps of terminology extraction tools, the identification of term candidates and their ranking, but implements new features. It is multilingually designed, scalable, and handles term variants. We focus on the main components: UIMA Token RegExp for defining term and variant patterns over word

annotations, and the grouping component for clustering terms and variants that works both at morphological and syntactic levels.

demos-025

demos-025

Despite the abundance of biomedical literature and health discussions in online communities, it is often tedious to retrieve informative contents for health-centric information needs. Users can query scholarly work in PubMed by keywords and MeSH terms, and resort to Google for everything else. This demo paper presents the DeepLife system, to overcome the limitations of existing search engines for life science and health topics. DeepLife integrates large knowledge bases and harnesses entity linking methods, to support search and exploration of scientific literature, newspaper feeds, and social media, in terms of keywords and phrases, biomedical entities, and taxonomic categories. It also provides functionality for entity-aware text analytics over health-centric contents.

demos-048

demos-048

Publicly available knowledge repositories, such as Wikipedia and Freebase, benefit significantly from volunteers, whose contributions ensure that the knowledge keeps expanding and is kept up-to-date and accurate. User interactions are often limited to hypertext, tabular, or graph visualization interfaces. For spatio-temporal information, however, other interaction paradigms may be better-suited. We present an integrated system that combines crowdsourcing, automatic or semi-automatic knowledge harvesting from text, and visual analytics. It enables users to analyze large quantities of structured data and unstructured textual data from a spatio-temporal perspective and gain deep insights that are not easily observed in individual facts.

demos-070

demos-070

The current release of ODIN (Online Database of Interlinear Text) contains over 150,000 linguistic examples, from nearly 1,500 languages, extracted from PDFs found on the web, representing a significant source of data for language research, particularly for low-resource languages. Errors during PDF-to-text conversion or poorly formatted examples can make the task of automatically analyzing the data more difficult, so we aim to clean and normalize the examples in order to maximize accuracy during analysis. In this paper we describe a system that allows users to automatically and manually correct errors in the source data in order to get the best possible analysis of the data. We also describe a RESTful service for managing collections of linguistic examples on the web. All software is distributed under an open-source license.

demos-012

demos-012

Monitoring mobility- and industry-relevant events is important in areas such as personal travel planning and supply chain management, but extracting events pertaining to specific companies, transit routes and locations from heterogeneous, high-volume text streams remains a significant challenge. We present Spree, a scalable system for real-time, automatic event extraction from social media, news and domain-specific RSS feeds. Our system is tailored to a range of mobility- and industry-related events, and processes German texts within a distributed linguistic analysis pipeline implemented in Apache Flink. The pipeline detects and disambiguates highly ambiguous domain-relevant entities, such as street names, and extracts various events with their geo-locations. Event streams are visualized on a dynamic, interactive map for monitoring and analysis.

demos-024

demos-024

We present TranscRater, an open-source tool for automatic speech recognition (ASR) quality estimation (QE). The tool allows users to perform ASR evaluation bypassing the need of reference transcripts and confidence information, which is common to current assessment protocols. TranscRater includes: i) methods to extract a variety of quality indicators from (signal, transcription) pairs and ii) machine learning algorithms which make possible to build ASR QE models exploiting the extracted features. Confirming the positive results of previous evaluations, new experiments with TranscRater indicate its effectiveness both in WER prediction and transcription ranking tasks.

demos-055

demos-055

We present TMop, the first open-source tool for automatic Translation Memory (TM) cleaning. The tool imple-

ments a fully unsupervised approach to the task, which allows spotting unreliable translation units (sentence pairs in different languages, which are supposed to be translations of each other) without requiring labeled training data. TMop includes a highly configurable and extensible set of filters capturing different aspects of translation quality. It has been evaluated on a test set composed of 1,000 translation units (TUs) randomly extracted from the English-Italian version of MyMemory, a large-scale public TM. Results indicate its effectiveness in automatic removing “bad” TUs, with comparable performance to a state-of-the-art supervised method (76.3 vs. 77.7 balanced accuracy).

demos-015

demos-015

Research at the intersection of language and other modalities, most notably vision, is becoming increasingly important in our field. We introduce a toolkit that can be used to obtain feature representations for visual as well as auditory information. MMFEAT is an easy-to-use Python toolkit, which has been developed with the purpose of making non-linguistic modalities more accessible to natural language processing researchers.

demos-038

demos-038

Freebase contains entities and relation information but is highly incomplete. Relevant information is ubiquitous in web text, but extraction deems challenging. We present JEDI, an automated system to jointly extract typed named entities and Freebase relations using dependency pattern from text. An innovative method for constraint solving on entity types of multiple relations is used to disambiguate pattern. The high precision in the evaluation supports our claim that we can detect entities and relations together, alleviating the need to train a custom classifier for an entity type. A Demonstrator as well as the source code is available at jedi.textmining.tu-berlin.de

demos-034

demos-034

We present a new release of OpenDial, an open-source toolkit for building and evaluating spoken dialogue systems. The toolkit relies on an information-state architecture where the dialogue state is represented as a Bayesian network and acts as a shared memory for all system modules. The domain models are specified via probabilistic rules encoded in XML. OpenDial has been deployed in several application domains such as human-robot interaction, intelligent tutoring systems and multi-modal in-car driver assistants.

demos-060

demos-060

The MUSEEC (MULTilingual SEntence Extraction and Compression) summarization tool implements several extractive summarization techniques — at the level of complete and compressed sentences — that can be applied, with some minor adaptations, to documents in multiple languages. The current version of MUSEEC provides the following summarization methods: (1) MUSE — a supervised summarizer, based on a genetic algorithm (GA), that ranks document sentences and extracts top-ranking sentences into a summary, (2) POLY — an unsupervised summarizer, based on linear programming (LP), that selects the best extract of document sentences, and (3) WECOM — an unsupervised extension of POLY that compiles a document summary from compressed sentences. In this paper, we provide an overview of MUSEEC methods and its architecture in general.

demos-020

demos-020

Current education standards in the U.S. require school students to read and understand complex texts from different subject areas (e.g., social studies). However, such texts usually contain figurative language, complex phrases and sentences, as well as unfamiliar discourse relations. This may present an obstacle to students whose native language is not English – a growing sub-population in the US. One way to help such students is to create classroom activities centered around linguistic elements found in subject area texts (DelliCarpini, 2008). We present a web-based tool that uses NLP algorithms to automatically generate customizable linguistic activities that are grounded in language learning research.

demos-082

demos-082

We demonstrate a simple and easy-to-use system to produce logical semantic representations of sentences. Our software operates by composing semantic formulas bottom-up given a CCG parse tree. It uses flexible semantic templates to specify semantic patterns. Templates for English and Japanese accompany our software, and they are easy to understand, use and extend to cover other linguistic phenomena or languages. We also provide scripts to use our semantic representations in a textual entailment task, and a visualization tool to display semantically augmented CCG trees in HTML.

demos-049**demos-049**

MeTA is developed to unite machine learning, information retrieval, and natural language processing in one easy-to-use toolkit. Its focus on indexing allows it to perform well on large datasets, supporting online classification and other out-of-core algorithms. MeTAs liberal open source license encourages contributions, and its extensive online documentation, forum, and tutorials make this process straightforward. We run experiments and show MeTA's performance is competitive with or better than existing software.

demos-023**demos-023**

In this paper, we present MDSWriter, a novel open-source annotation tool for creating multi-document summarization corpora. A major innovation of our tool is that we divide the complex summarization task into multiple steps which enables us to efficiently guide the annotators, to store all their intermediate results, and to record user-system interaction data. This allows for evaluating the individual components of a complex summarization system and learning from the human writing process. MDSWriter is highly flexible and can be adapted to various other tasks.

demos-081**demos-081**

We present Jigg, a Scala (or JVM-based) NLP annotation pipeline framework, which is easy to use and is extensible. Jigg supports a very simple interface similar to Stanford CoreNLP, the most successful NLP pipeline toolkit, but has more flexibility to adapt to new types of annotation. On this framework, system developers can easily integrate their downstream system into a NLP pipeline from a raw text by just preparing a wrapper of it.

demos-051**demos-051**

In our media-driven world the perception of companies and institutions in the media is of major importance. The creation of press reviews analyzing the media response to company-related events is a complex and time-consuming task. In this demo we present a system that combines advanced text-mining and machine learning approaches in an extensible press review system. The system collects documents from heterogeneous sources and enriches the documents applying different mining, filtering, classification, and aggregation algorithms. We present a system tailored to the needs of the press department of a major German University. We explain how the different components have been trained and evaluated. The system enables us demonstrating the live analyzes of news and social media streams as well as the strength of advanced text-mining algorithms for creating a comprehensive media analysis.

4

Main Conference: Tuesday, August 9

Overview

9:00 – 10:10	Invited talk II: Mark Steedman	<i>Plaza Ballroom A, B, & C</i>						
10:10 – 10:40	Coffee break	<i>Plaza Ballroom D & E</i>						
Session 4								
10:40 – 12:00	Relations and knowledge bases	Semantic parsing II	Word meaning II	Tasks and datasets	Parsing IV	Document analysis		
Lunch break								
Session 5								
1:40 – 3:00	Deep learning (short papers)	Semantics and generation (short papers)	Machine translation II (short papers)	Text classification (short papers)	Potpourri I (short papers)			
3:00 – 3:30	Coffee break	<i>Plaza Ballroom D & E</i>						
Session 6								
3:30 – 5:10	Machine learning	Word vectors II	Machine translation III	Discourse	Language and vision	Summarization		
5:30 – 7:00	Poster and dinner session II (includes demonstrations B)							
<i>Plaza Ballroom A, B, & C</i>								
8:00 – 11:00	Social event: Boat trip	<i>Grand Ballroom</i>						

Session 4 Overview – Tuesday, August 9, 2016

Track A <i>Relations and knowledge bases</i>	Track B <i>Semantic parsing II</i>	Track C <i>Word meaning II</i>	Track D <i>Tasks and datasets</i>	Track E <i>Parsing IV</i>	Track F <i>Document analysis</i>
Plaza Ballroom	Plaza Ballroom	Plaza Ballroom	Plaza Ballroom	Plaza Ballroom	Plaza Ballroom F
papers-639 papers-639	papers-666 papers-666		papers-747 papers-747	tacl-921 tacl-921	papers-740 papers-740
papers-826 papers-826		tacl-914 tacl-914	papers-425 papers-425	papers-798 papers-798	papers-787 papers-787
papers-831 papers-831	papers-265 papers-265	papers-200 papers-200	papers-590 papers-590	papers-577 papers-577	papers-252 papers-252
tacl-911 tacl-911	tacl-907 tacl-907	papers-309 papers-309	papers-563 papers-563		papers-256 papers-256

11:40 11:20 11:00 10:40

Parallel Session 4

4A: Relations and knowledge bases

Plaza Ballroom F

Chair:

papers-639

10:40–11:00

The automatic detection of causal relationships in text is important for natural language understanding. This task has proven to be difficult, however, due to the need for world knowledge and inference. We focus on a sub-task of this problem where an open class set of linguistic markers can provide clues towards understanding causality. Unlike the explicit markers, a closed class, these markers vary significantly in their linguistic forms. We leverage parallel Wikipedia corpora to identify new markers that are variations on known causal phrases, creating a training set via distant supervision. We also train a causal classifier using features from the open class markers and semantic features providing contextual information. The results show that our features provide an 11.05 point absolute increase over the baseline on the task of identifying causality in text.

papers-826

11:00–11:20

Modeling relation paths has been shown to offer significant gains in embedding models for knowledge base (KB) completion. However, enumerating paths between two entities is very expensive, and existing approaches typically resort to an approximation using a sampled subset. This problem is particularly acute when text is jointly modeled with KB relations and used to provide direct evidence for facts mentioned in it. In this paper, we propose the first exact dynamic programming algorithm which enables efficient incorporation of all multi-step relation paths of bounded length, while also modeling both relation types and intermediate nodes in the compositional path representations. We conduct a theoretical analysis on the efficiency gain from such an approach. Experiments on biomedical domain show that it addresses representational limitations in prior approaches and improves accuracy in KB completion.

papers-831

11:20–11:40

We enrich a curated resource of commonsense knowledge by formulating the problem as one of knowledge base completion (KBC). Most work in KBC focuses on knowledge bases like Freebase that relate entities drawn from a fixed set. However, the tuples in ConceptNet (Speer and Havasi, 2012) define relations between an unbounded set of phrases. We develop neural network models for scoring tuples on arbitrary phrases and evaluate them by their ability to distinguish true held-out tuples from false ones. We find strong performance from a bilinear model using a simple additive architecture to model phrases. We manually evaluate our trained model’s ability to assign quality scores to novel tuples, finding that it can propose tuples at the same quality level as medium-confidence tuples from ConceptNet.

tacl-911

11:40–12:00

We present a method for unsupervised open-domain relation discovery. In contrast to previous (mostly generative and agglomerative clustering) approaches, our model relies on rich contextual features and makes minimal independence assumptions. The model is composed of two parts: a feature-rich relation extractor, which predicts a semantic relation between two entities, and a factorization model, which reconstructs arguments (i.e., the entities) relying on the predicted relation. The two components are estimated jointly so as to minimize errors in recovering arguments. We study factorization models inspired by previous work in relation factorization and selectional preference modeling. Our models substantially outperform the generative and agglomerative-clustering counterparts and achieve state-of-the-art performance.

4B: Semantic parsing II

Plaza Ballroom F

Chair:

papers-666

papers-666

10:40–11:00

We consider the task of learning a context-dependent mapping from utterances to denotations. With only denotations at training time, we must search over a combinatorially large space of logical forms, which is even larger with context-dependent utterances. To cope with this challenge, we perform successive projections of the full model onto simpler models that operate over equivalence classes of logical forms. Though less expressive, we find that these simpler models are much faster and can be surprisingly effective. Moreover, they can be used to bootstrap the full model. Finally, we collected three new context-dependent semantic parsing datasets, and develop a new left-to-right parser.

papers-265

papers-265

11:20–11:40

Tree-structured neural networks exploit valuable syntactic parse information as they interpret the meanings of sentences. However, they suffer from two key technical problems that make them slow and unwieldy for large-scale NLP tasks: they usually operate on parsed sentences and they do not directly support batched computation. We address these issues by introducing the Stack-augmented Parser-Interpreter Neural Network (SPINN), which combines parsing and interpretation within a single tree-sequence hybrid model by integrating tree-structured sentence interpretation into the linear sequential structure of a shift-reduce parser. Our model supports batched computation for a speedup of up to 25 times over other tree-structured models, and its integrated parser can operate on unparsed data with little loss in accuracy. We evaluate it on the Stanford NLI entailment task and show that it significantly outperforms other sentence-encoding models.

tacl-907

tacl-907

11:40–12:00

We introduce a new approach to training a semantic parser that uses textual entailment judgements as supervision. These judgements are based on high-level inferences about whether the meaning of one sentence follows from another. When applied to an existing semantic parsing task, they prove to be a useful tool for revealing semantic distinctions and background knowledge not captured in the target representations. This information is used to improve the quality of the semantic representations being learned and to acquire generic knowledge for reasoning. Experiments are done on the benchmark Sportscaster corpus (Chen and Mooney, 2008), and a novel RTE-inspired inference dataset is introduced. On this new dataset our method strongly outperforms several strong baselines. Separately, we obtain state-of-the-art results on the original Sportscaster semantic parsing task.

4C: Word meaning II

Plaza Ballroom F

Chair:

tacl-914

tacl-914

11:00–11:20

Word meanings change over time and an automated procedure for extracting this information from text would be useful for historical exploratory studies, information retrieval or question answering. We present a dynamic Bayesian model of diachronic meaning change, which infers temporal word representations as a set of senses and their prevalence. Unlike previous work, we explicitly model language change as a smooth, gradual process. We experimentally show that this modeling decision is beneficial: our model performs competitively on meaning change detection tasks whilst inducing discernible word senses and their development over time. Application of our model to the SemEval-2015 temporal classification benchmark datasets further reveals that it performs on par with highly optimized task-specific systems.

papers-200

papers-200

11:20–11:40

Recently, many NLP tasks have benefited from distributed word representation. However, it remains unknown whether embedding models are really immune to the typological diversity of languages, despite the language-independent architecture. Here we investigate three models on a large set of language samples by mapping dense embedding to sparse linguistic property space. Experiment results reveal the language universal and specific encoded in various word representation. Additionally, strong evidence supports the utility of word form, especially for inflectional languages.

papers-309

papers-309

11:40–12:00

Understanding how words change their meanings over time is key to models of language and cultural evolution, but historical data on meaning is scarce, making theories hard to develop and test. Word embeddings show promise as a diachronic tool, but have not been carefully evaluated. We develop a robust methodology for quantifying semantic change by evaluating word embeddings (PPMI, SVD, word2vec) against known historical changes. We then use this methodology to reveal statistical laws of semantic evolution. Using six historical corpora spanning four languages and two centuries, we propose two quantitative laws of semantic change: (i) the law of conformity—the rate of semantic change scales with an inverse power-law of word frequency; (ii) the law of innovation—Independent of frequency, words that are more polysemous have higher rates of semantic change.

4D: Tasks and datasets

Plaza Ballroom F

*Chair:***papers-747****papers-747**

10:40–11:00

This paper complements semantic role representations with spatial knowledge beyond indicating plain locations. Namely, we extract where entities are (and are not) located, and for how long (seconds, hours, days, etc.). Crowdsourced annotations show that this additional knowledge is intuitive to humans and can be annotated by non-experts. Experimental results show that the task can be automated.

papers-425**papers-425**

11:00–11:20

There has recently been a lot of interest in unsupervised methods for learning sense distributions, particularly in applications where sense distinctions are needed. This paper analyses a state-of-the-art method for sense distribution learning, and optimises it for application to the entire vocabulary of a given language. The optimised method is then used to produce LexSemTM: a sense frequency and semantic dataset of unprecedented size, spanning approximately 88% of polysemous, English simplex lemmas, which is released as a public resource to the community. Finally, the quality of this data is investigated, and the LexSemTM sense distributions are shown to be superior to those based on the WordNet first sense for lemmas missing from SEMCOR, and at least on par with SEMCOR-based distributions otherwise.

papers-590**papers-590**

11:20–11:40

We introduce LAMBADA, a dataset to evaluate the capabilities of computational models for text understanding by means of a word prediction task. LAMBADA is a collection of narrative passages sharing the characteristic that human subjects are able to guess their last word if they are exposed to the whole passage, but not if they only see the last sentence preceding the target word. To succeed on LAMBADA, computational models cannot simply rely on local context, but must be able to keep track of information in the broader discourse. We show that LAMBADA exemplifies a wide range of linguistic phenomena, and that none of several state-of-the-art language models reaches accuracy above 1% on this novel benchmark. We thus propose LAMBADA as a challenging test set, meant to encourage the development of new models capable of genuine understanding of broad context in natural language text.

papers-563**papers-563**

11:40–12:00

We present WikiReading, a large-scale natural language understanding task and publicly-available dataset with 18 million instances. The task is to predict textual values from the structured knowledge base Wikidata by reading the text of the corresponding Wikipedia articles. The task contains a rich variety of challenging classification and extraction sub-tasks, making it well-suited to end-to-end models such as deep neural networks (DNNs). We compare various state-of-the-art DNN-based architectures for document classification, information extraction, and question answering. We find that models supporting a rich answer space, such as word or character sequences, perform best. Our best-performing model, a word-level sequence to sequence model with a mechanism to copy out-of-vocabulary words, obtains an accuracy of 71.8%.

4E: Parsing IV

Plaza Ballroom F

Chair:

tacl-921**tacl-921**

10:40–11:00

We present a simple and effective scheme for dependency parsing which is based on bidirectional-LSTMs (BiLSTMs). Each sentence token is associated with a BiLSTM vector representing the token in its sentential context, and feature vectors are constructed by concatenating a few BiLSTM vectors. The BiLSTM is trained jointly with the parser objective, resulting in very effective feature extractors for parsing. We demonstrate the effectiveness of the approach by applying it to a greedy transition based parser as well as to a globally optimized graph-based parser. The resulting parsers have very simple architectures, and match or surpass the state-of-the-art accuracies on English and Chinese.

papers-798**papers-798**

11:00–11:20

We describe a search algorithm for optimizing the number of latent states when estimating latent-variable PCFGs with spectral methods. Our results show that contrary to the common belief that the number of latent states for each nonterminal in an L-PCFG can be decided in isolation with spectral methods, parsing results significantly improve if the number of latent states for each nonterminal is globally optimized, while taking into account interactions between the different nonterminals. In addition, we contribute an empirical analysis of spectral algorithms on eight morphologically rich languages: Basque, French, German, Hebrew, Hungarian, Korean, Polish and Swedish. Our results show that our estimation consistently performs better or close to coarse-to-fine expectation-maximization techniques for these languages.

papers-577**papers-577**

11:20–11:40

Traditional syntax models typically leverage part-of-speech (POS) information by constructing features from hand-tuned templates. We demonstrate that a better approach is to utilize POS tags as a *regularizer* of learned representations. We propose a simple method for learning a stacked pipeline of models which we call “stack-propagation.” We apply this to dependency parsing and tagging, where we use the hidden layer of the tagger network as a representation of the input tokens for the parser. At test time, our parser does not require predicted POS tags. On 19 languages from the Universal Dependencies, our method is 1.3% (absolute) more accurate than a state-of-the-art graph-based approach and 2.7% more accurate than the most comparable greedy model.

4F: Document analysis

Plaza Ballroom F

Chair:

papers-740

papers-740

10:40–11:00

We examine communications in a social network to study user emotional contrast - the propensity of users to express different emotions than those expressed by their neighbors. Our analysis is based on a large Twitter dataset, consisting of the tweets of 123,513 users from the USA and Canada. Focusing on Ekman's basic emotions, we analyze differences between the emotional tone expressed by these users and their neighbors and correlate these differences with perceived user demographics. We demonstrate that many perceived user demographic traits correlate with the emotional contrast between users and their neighbors. Unlike other approaches on inferring user demographics in social media that rely solely on user communications, we explore the network structure and show that it is possible to accurately predict a range of perceived demographic traits based solely on the emotions emanating from users and their neighbors.

papers-787

papers-787

11:00–11:20

State legislatures often rely on existing text when drafting new bills. Resource and expertise constraints, which often drive this copying behavior, can be taken advantage of by lobbyists and special interest groups. These groups provide model bills, which encode policy agendas, with the intent that the models become actual law. Unfortunately, model legislation is often opaque to the public—both in source and content. In this paper we present LOBBYBACK, a system that reverse engineers model legislation from observed text. LOBBYBACK identifies clusters of bills which have text reuse and generates “prototypes” that represent a canonical version of the text shared between the documents. We demonstrate that LOBBYBACK accurately reconstructs model legislation and apply it to a dataset of over 550k bills.

papers-252

papers-252

11:20–11:40

We propose a new task in the field of computational argumentation in which we investigate qualitative properties of Web arguments, namely their convincingness. We cast the problem as relation classification, where a pair of arguments having the same stance to the same prompt is judged. We annotate a large datasets of 16k pairs of arguments over 32 topics and investigate whether the relation “A is more convincing than B” exhibits properties of total ordering; these findings are used as global constraints for cleaning the crowdsourced data. We propose two tasks: (1) predicting which argument from an argument pair is more convincing and (2) ranking all arguments to the topic based on their convincingness. We experiment with feature-rich SVM and bidirectional LSTM and obtain 0.76-0.78 accuracy and 0.35-0.40 Spearman’s correlation in a cross-topic evaluation. We release the newly created corpus UKPConvArg1 and the experimental software under open licenses.

papers-256

papers-256

11:40–12:00

An extensive literature in computational social science examines how features of messages, advertisements, and other corpora affect individuals' decisions, but these analyses must specify the relevant features of the text before the experiment. Automated text analysis methods are able to discover features of text, but these methods cannot be used to obtain the estimates of causal effects—the quantity of interest for applied researchers. We introduce a new experimental design and statistical model to simultaneously discover treatments in a corpora and estimate causal effects for these discovered treatments. We prove the conditions to identify the treatment effects of texts and introduce the supervised Indian Buffet process to discover those treatments. Our method enables us to discover treatments in a training set using a collection of texts and individuals' responses to those texts, and then estimate the effects of these interventions in a test set of new texts and survey respondents. We apply the model to an experiment about candidate biographies, recovering intuitive features of voters' decisions and revealing a penalty for lawyers and a bonus for military service.

Session 5 Overview – Tuesday, August 9, 2016

Track A <i>Deep learning (short papers)</i>	Track B <i>Semantics and generation (short papers)</i>	Track C <i>Machine trans- lation II (short papers)</i>	Track D <i>Text classifica- tion (short papers)</i>	Track E <i>Potpourri I (short papers)</i>
Plaza Ballroom F	Plaza Ballroom F	Plaza Ballroom F	Plaza Ballroom F	Plaza Ballroom F
short-327	short-367	short-195	short-323	short-312
short-327	short-367	short-195	short-323	short-312
short-165	short-356	short-132	short-471	short-417
short-165	short-356	short-132	short-471	short-417
short-176	short-256	short-329	short-470	short-279
short-176	short-256	short-329	short-470	short-279
short-314	short-318	short-254	short-219	
short-314	short-318	short-254	short-219	
short-103	short-438	short-286	short-447	
short-103	short-438	short-286	short-447	

02:44 02:28 02:12 01:56 01:40

Parallel Session 5

5A: Deep learning (short papers)

Plaza Ballroom F

*Chair:***short-327**

13:40–13:56

As a first step towards agents learning to communicate about their visual environment, we propose a system that, given visual representations of a referent (cat) and a context (sofa), identifies their discriminative attributes, i.e., properties that distinguish them (has_tail). Moreover, despite the lack of direct supervision at the attribute level, the model learns to assign plausible attributes to objects (sofa-has_cushion). Finally, we present a preliminary experiment confirming the referential success of the predicted discriminative attributes.

short-165

13:56–14:12

We describe an efficient neural network method to automatically learn sentiment lexicons without relying on any manual resources. The method takes inspiration from the NRC method, which gives the best results in SemEval13 by leveraging emoticons in large tweets, using the PMI between words and tweet sentiments to define the sentiment attributes of words. We show that better lexicons can be learned by using them to predict the tweet sentiment labels. By using a very simple neural network, our method is fast and can take advantage of the same data volume as the NRC method. Experiments show that our lexicons give significantly better accuracies on multiple languages compared to the current best methods.

short-176

14:12–14:28

Dimensional sentiment analysis aims to recognize continuous numerical values in multiple dimensions such as the valence-arousal (VA) space. Compared to the categorical approach that focuses on sentiment classification such as binary classification (i.e., positive and negative), the dimensional approach can provide more fine-grained sentiment analysis. This study proposes a regional CNN-LSTM model consisting of two parts: regional CNN and LSTM to predict the VA ratings of texts. Unlike a conventional CNN which considers a whole text as input, the proposed regional CNN uses an individual sentence as a region, dividing an input text

into several regions such that the useful affective information in each region can be extracted and weighted according to their contribution to the VA prediction. Such regional information is sequentially integrated across regions using LSTM for VA prediction. By combining the regional CNN and LSTM, both local (regional) information within sentences and long-distance dependency across sentences can be considered in the prediction process. Experimental results show that the proposed method outperforms lexicon-based, regression-based, and NN-based methods proposed in previous studies.

short-314**short-314**

14:28–14:44

In all previous work on deep multi-task learning we are aware of, all task supervisions are on the same (outermost) layer. We present a multi-task learning architecture with deep bi-directional RNNs, where different tasks supervision can happen at different layers. We present experiments in syntactic chunking and CCG supertagging, coupled with the additional task of POS-tagging. We show that it is consistently better to have POS supervision at the innermost rather than the outermost layer. We argue that this is because “low-level” tasks are better kept at the lower layers, enabling higher-level tasks make use of the shared representation of the lower-level tasks. Finally, we also show how this architecture can be used for domain adaptation.

short-103**short-103**

14:44–15:00

In this paper, we propose a method for referring to the real world to improve named entity recognition (NER) specialized for a domain. Our method adds a stacked auto-encoder to a text-based deep neural network for NER. We first train the stacked auto-encoder only from the real world information, then the entire deep neural network from sentences annotated with NEs and accompanied by real world information. In our experiments, we took Japanese chess as the example. The dataset consists of pairs of a game state and commentary sentences about it annotated with game-specific NE tags. We conducted NER experiments and showed that referring to the real world improves the NER accuracy.

5B: Semantics and generation (short papers)

Plaza Ballroom F

*Chair:***short-367****short-367**

13:40–13:56

Finding quality descriptions on the web, such as those found in Wikipedia articles, of newer companies can be difficult: search engines show many pages with varying relevance, while multi-document summarization algorithms find it difficult to distinguish between core facts and other information such as news stories. In this paper, we propose an entity-focused, hybrid generation approach to automatically produce descriptions of previously unseen companies, and show that it outperforms a strong summarization baseline.

short-356**short-356**

13:56–14:12

We present a new annotation method for collecting data on relation inference in context. We convert the inference task to one of simple factoid question answering, allowing us to easily scale up to 16,000 high-quality examples. Our method corrects a major bias in previous evaluations, making our dataset much more realistic.

short-256**short-256**

14:12–14:28

This paper addresses an automatic classification of preposition types in German, comparing hard and soft clustering approaches and various window- and syntax-based co-occurrence features. We show that (i) the semantically most salient preposition features (i.e., subcategorised nouns) are the most successful, and that (ii) soft clustering approaches are required for the task but reveal quite different attitudes towards predicting ambiguity.

short-318**short-318**

14:28–14:44

Decision-making is often dependent on uncertain data, e.g. data associated with confidence scores or probabilities. We present a comparison of different information presentations for uncertain data and, for the first time, measure their effects on human decision-making. We show that the use of Natural Language Generation (NLG) improves decision-making under uncertainty, compared to state-of-the-art graphical-based representation methods. In a task-based study with 442 adults, we found that presentations using NLG lead to 24% better decision-making on average than the graphical presentations, and to 44% better decision-making when NLG is combined with graphics. We also show that women achieve significantly better results when presented with NLG output (an 87% increase on average compared to graphical presentations).

short-438**short-438**

14:44–15:00

Text from social media provides a set of challenges that can cause traditional NLP approaches to fail. Informal language, spelling errors, abbreviations, and special characters are all commonplace in these posts, leading to a prohibitively large vocabulary size for word-level approaches. We propose a character composition model, tweet2vec, which finds vector-space representations of whole tweets by learning complex, non-local dependencies in character sequences. The proposed model outperforms a word-level baseline at predicting user-annotated hashtags associated with the posts, doing significantly better when the input contains many out-of-vocabulary words or unusual character sequences. Our tweet2vec encoder is publicly available.

5C: Machine translation II (short papers)

Plaza Ballroom F

*Chair:***short-195****short-195**

13:40–13:56

Constrained translation has improved statistical machine translation (SMT) by combining it with translation memory (TM) at sentence-level. In this paper, we propose using a constrained word lattice, which encodes input phrases and TM constraints together, to combine SMT and TM at phrase-level. Experiments on English–Chinese and English–French show that our approach is significantly better than previous combination methods, including sentence-level constrained translation and a recent phrase-level combination.

short-132**short-132**

13:56–14:12

We present a neural network based automatic post-editing (APE) system to improve raw machine translation (MT) output. Our neural model of APE (NNAPE) is based on a bidirectional recurrent neural network (RNN) model and consists of an encoder that encodes an MT output into a fixed-length vector from which a decoder provides a post-edited (PE) translation. APE translations produced by NNAPE showed statistically significant improvements of 3.96, 2.68 and 1.35 BLEU points over the original MT, phrase-based APE and hierarchical APE outputs, respectively. Furthermore, human evaluation also revealed that the NNAPE generated PE translations are much better than the original MT output.

short-329**short-329**

14:12–14:28

We address the problem of automatically cleaning a large-scale Translation Memory (TM) in a fully unsupervised fashion, i.e. without human-labelled data. We approach the task by: i) designing a set of features that capture the similarity between two text segments in different languages, ii) use them to induce reliable training labels for a subset of the translation units (TUs) contained in the TM, and iii) use the automatically labelled data to train an ensemble of binary classifiers. We apply our method to clean a test set composed of 1,000 TUs randomly extracted from the English–Italian version of MyMemory, the world’s largest public TM. Our results show competitive performance not only against a strong baseline that exploits machine translation, but also against a state-of-the-art method that relies on human-labelled data.

short-254**short-254**

14:28–14:44

Recently, neural network models have achieved consistent improvements in statistical machine translation. However, most networks only use one-hot encoded input vectors of words as their input. In this work, we investigated the exponentially decaying bag-of-words input features for feed-forward neural network translation models and proposed to train the decay rates along with other weight parameters. This novel bag-of-words model improved our phrase-based state-of-the-art system, which already includes a neural network translation model, by up to 0.5% Bleu and 0.6% Ter on three different translation tasks and even achieved a similar performance to the bidirectional LSTM translation model.

short-286**short-286**

14:44–15:00

We investigate the use of hierarchical phrase-based SMT lattices in end-to-end neural machine translation (NMT). Weight pushing transforms the Hiero scores for complete translation hypotheses, with the full translation grammar score and full n-gram language model score, into posteriors compatible with NMT predictive probabilities. With a slightly modified NMT beam-search decoder we find gains over both Hiero and NMT decoding alone, with practical advantages in extending NMT to very large input and output vocabularies.

5D: Text classification (short papers)

Plaza Ballroom F

Chair:

short-323**short-323**

13:40–13:56

We automatically predict properties of wines on the basis of smell and flavor descriptions from experts' wine reviews. We show wine experts are capable of describing their smell and flavor experiences in wine reviews in a sufficiently consistent manner, such that we can use their descriptions to predict properties of a wine based solely on language. The experimental results show promising F-scores when using lexical and semantic information to predict the color, grape variety, country of origin, and price of a wine. This demonstrates, contrary to popular opinion, that wine experts' reviews really are informative.

short-471**short-471**

13:56–14:12

Writing style allows NLP tools to adjust to the traits of an author. In this paper, we explore the relation between stylistic and syntactic features and authors' age and income. We confirm our hypothesis that for numerous feature types writing style is predictive of income even beyond age. We analyze the predictive power of writing style features in a regression task on two data sets of around 5,000 Twitter users each. Additionally, we use our validated features to study daily variations in writing style of users from distinct income groups. Temporal stylistic patterns not only provide novel psychological insight into user behavior, but are useful for future research and applications in social media.

short-470**short-470**

14:12–14:28

Optimism is linked to various personality factors as well as both psychological and physical health, but how does it relate to the way a person tweets? We analyze the online activity of a set of Twitter users in order to determine how well machine learning algorithms can detect a person's outlook on life by reading their tweets. A sample of tweets from each user is manually annotated in order to establish ground truth labels, and classifiers are trained to distinguish between optimistic and pessimistic users. Our results suggest that the words in people's tweets provide ample evidence to identify them as optimists, pessimists, or somewhere in between. Additionally, several applications of these trained models are explored.

short-219**short-219**

14:28–14:44

Access to data is critical to any machine learning component aimed at training an accurate predictive model. In reality, data is often a subject of technical and legal constraints. Data may contain sensitive topics and data owners are often reluctant to share them. Instead of access to data, they make available decision making procedures to enable predictions on new data. Under the black box classifier constraint, we build an effective domain adaptation technique which adapts classifier predictions in a transductive setting. We run experiments on text categorization datasets and show that significant gains can be achieved, especially in the unsupervised case where no labels are available in the target domain.

short-447**short-447**

14:44–15:00

Microblogging sites have emerged as major platforms for bloggers to create and consume posts as well as to follow other bloggers and get informed of their updates. Due to the large number of users, and the huge amount of posts they create, it becomes extremely difficult to identify relevant and interesting blog posts. In this paper, we propose a novel convex collective matrix completion (CCMC) method that effectively utilizes user-item matrix and incorporates additional user activity and topic-based signals to recommend relevant content. The key advantage of CCMC over existing methods is that it can obtain a globally optimal solution and can easily scale to large-scale matrices using Hazan's algorithm. To the best of our knowledge, this is the first work which applies and studies CCMC as a recommendation method in social media. We conduct a large scale study and show significant improvement over existing state-of-the-art approaches.

5E: Potpourri I (short papers)

Plaza Ballroom F

Chair:

short-312

short-312

13:40–13:56

We present a new, structured approach to text simplification using conditional random fields over top-down traversals of dependency graphs that jointly predicts possible compressions and paraphrases. Our model reaches readability scores comparable to word-based compression approaches across a range of metrics and human judgements while maintaining more of the important information.

short-417

short-417

13:56–14:12

We present a named entity recognition (NER) system for tagging fiction: LitNER. Relative to more traditional approaches, litNER has two important properties: (1) it makes use of no hand-tagged data or gazetteers, instead bootstrapping from term clusters; and (2) it leverages multiple instances of the same name in a text. Our experiments show it to substantially outperform off-the-shelf supervised NER systems.

short-279

short-279

14:12–14:28

Online reviews are a growing market, but it is struggling with fake reviews. They undermine both the value of reviews to the user, and their trust in the review sites. However, fake positive reviews can boost a business, and so a small industry producing fake reviews has developed. The two sides are facing an arms race that involves more and more natural language processing (NLP). So far, NLP has been used mostly for detection, and works well on human-generated reviews. But what happens if NLP techniques are used to generate fake reviews as well? We investigate the question in an adversarial setup, by assessing the detectability of different fake-review generation strategies. We use generative models to produce reviews based on meta-information, and evaluate their effectiveness against deception-detection models and human judges. We find that meta-information helps detection, but that NLP-generated reviews conditioned on such information are also much harder to detect than conventional ones.

Session 6 Overview – Tuesday, August 9, 2016

Track A <i>Machine learning</i>	Track B <i>Word vectors II</i>	Track C <i>Machine translation III</i>	Track D <i>Discourse</i>	Track E <i>Language and vision</i>	Track F <i>Summarization</i>	Track G <i>Learner language</i>
Plaza Ballroom	Plaza Ballroom	Plaza Ballroom	Plaza Ballroom	Plaza Ballroom	Plaza Ballroom	Plaza Ballroom F
papers-231 tacl-909 tacl-909	papers-305 papers-305	papers-483 papers-483	papers-796 papers-796			papers-002 papers-002
papers-231 tacl-904 tacl-904	papers-353 papers-353	tacl-912 tacl-912	papers-225 papers-225	papers-288 papers-288		papers-461 papers-461
papers-470 papers-470	papers-693 papers-693	papers-388 papers-388	papers-580 papers-580	papers-041 papers-041		tacl-915 tacl-915
papers-752 papers-752	papers-655 papers-655	papers-506 papers-506	papers-382 papers-382	papers-793 papers-793	papers-508 papers-508	papers-832 papers-832
papers-529 papers-529		papers-145 papers-145	tacl-906 tacl-906	papers-596 papers-596	tacl-908 tacl-908	papers-240 papers-240

Parallel Session 6

6A: Machine learning

Plaza Ballroom F

Chair:

papers-231

papers-231

15:30–15:50

Structured prediction from bandit feedback describes a learning scenario where instead of having access to a gold standard structure, a learner only receives partial feedback in form of the loss value of a predicted structure. We present new learning objectives and algorithms for this interactive scenario, focusing on convergence speed and ease of elicitability of feedback. We present supervised-to-bandit simulation experiments for several NLP tasks (machine translation, sequence labeling, text classification), showing that bandit learning from relative preferences eases feedback strength and yields improved empirical convergence.

tacl-904

tacl-904

15:50–16:10

papers-470

papers-470

16:10–16:30

This paper introduces a novel architecture for reinforcement learning with deep neural networks designed to handle state and action spaces characterized by natural language, as found in text-based games. Termed a deep reinforcement relevance network (DRRN), the architecture represents action and state spaces with separate embedding vectors, which are combined with an interaction function to approximate the Q-function in reinforcement learning. We evaluate the DRRN on two popular text games, showing superior performance over other deep Q-learning architectures. Experiments with paraphrased action descriptions show that the model is extracting meaning rather than simply memorizing strings of text.

papers-752

papers-752

16:30–16:50

We address an important problem in sequence-to-sequence (Seq2Seq) learning referred to as copying, in which certain segments in the input sequence are selectively replicated in the output sequence. A similar phenomenon is observable in human language communication. For example, humans tend to repeat entity names or even long phrases in conversation. The challenge with regard to copying in Seq2Seq is that new machinery is needed to decide when to perform the operation. In this paper, we incorporate copying into neural network-based Seq2Seq learning and propose a new model called CopyNet with encoder-decoder structure. CopyNet can nicely integrate the regular way of word generation in the decoder with the new copying mechanism which can choose sub-sequences in the input sequence and put them at proper places in the output sequence. Our empirical study on both synthetic data sets and real world data sets demonstrates the efficacy of CopyNet. For example, CopyNet can outperform regular RNN-based model with remarkable margins on text summarization tasks.

papers-529

papers-529

16:50–17:10

Advances in transfer learning have let go the limitations of traditional supervised machine learning algorithms for being dependent on annotated training data for training new models for every new domains. Several real world applications encounter scenarios where models need to transfer/adapt across domains when the label sets vary both in terms of count of labels as well as their connotations. This paper presents first-of-its-kind transfer learning algorithm for cross-domain classification with multiple source domains and disparate label sets. It starts with identifying transferable knowledge from across multiple domains that can be useful for learning the target domain task. This knowledge in the form of selective labeled instances from different domains is congregated to form an auxiliary training set which is used for the target domain task. Experimental results validate the efficacy of the proposed algorithm against strong baselines on a real world social media and the \$20\$ Newsgroups datasets.

6B: Word vectors II

Plaza Ballroom F

Chair:

tacl-909**tacl-909**

15:30–15:50

Semantic word embeddings represent the meaning of a word via a vector, and are created by diverse methods. Many use nonlinear operations on co-occurrence statistics, and have hand-tuned hyperparameters and reweightings. This paper gives a new generative model, a dynamic version of the log-linear topic model of Mnih and Hinton (2007). The methodological novelty is to use the prior to compute closed form expressions for word statistics. This provides a theoretical explanation for nonlinear models like PMI, word2vec, and GloVe, as well as some hyperparameter choices. It also helps explain why low-dimensional semantic embeddings contain linear algebraic structure that allows solution of word analogies, as shown by Mikolov et al. (2013a) and many subsequent papers. Experimental support is provided for the generative model assumptions, the most important of which is that latent word vectors are fairly uniformly dispersed in space.

papers-353**papers-353**

15:50–16:10

Languages with rich inflectional morphology exhibit lexical data sparsity, since the word used to express a given concept will vary with the syntactic context. For instance, each count noun in Czech has 12 forms (where English uses only singular and plural). Even in large corpora, we are unlikely to observe all inflections of a given lemma. This reduces the vocabulary coverage of methods that induce continuous representations for words from distributional corpus information. We solve this problem by exploiting existing morphological resources that can enumerate a word's component morphemes. We present a latent-variable Gaussian graphical model that allows us to extrapolate continuous representations for words not observed in the training corpus, as well as smoothing the representations provided for the observed words. The latent variables represent embeddings of morphemes, which combine to create embeddings of words. Over five languages and training sizes, our model improves the embeddings for words, when evaluated on an analogy task, skip-gram predictive accuracy and word similarity.

papers-693**papers-693**

16:10–16:30

Despite interest in using cross-lingual knowledge to learn word embeddings for various tasks, a systematic comparison of the possible approaches is lacking in the literature. We perform an extensive evaluation of four popular approaches of inducing cross-lingual embeddings, each requiring a different form of supervision, on four typographically different language pairs. Our evaluation setup spans four different tasks, including intrinsic evaluation on mono-lingual and cross-lingual similarity, and extrinsic evaluation on downstream semantic and syntactic applications. We show that models which require expensive cross-lingual knowledge almost always perform better, but cheaply supervised models often prove competitive on certain tasks.

papers-655**papers-655**

16:30–16:50

Recent work has shown that simple vector subtraction over word embeddings is surprisingly effective at capturing different lexical relations, despite lacking explicit supervision. Prior work has evaluated this intriguing result using a word analogy prediction formulation and hand-selected relations, but the generality of the finding over a broader range of lexical relation types and different learning settings has not been evaluated. In this paper, we carry out such an evaluation in two learning settings: (1) spectral clustering to induce word relations, and (2) supervised learning to classify vector differences into relation types. We find that word embeddings capture a surprising amount of information, and that, under suitable supervised training, vector subtraction generalises well to a broad range of relations, including over unseen lexical items.

6C: Machine translation III

Plaza Ballroom F

*Chair:***papers-305****papers-305**

15:30–15:50

We propose minimum risk training for end-to-end neural machine translation. Unlike conventional maximum likelihood estimation, minimum risk training is capable of optimizing model parameters directly with respect to arbitrary evaluation metrics, which are not necessarily differentiable. Experiments show that our approach achieves significant improvements over maximum likelihood estimation on a state-of-the-art neural machine translation system across various languages pairs. Transparent to architectures, our approach can be applied to more neural networks and potentially benefit more NLP tasks.

tac1-912**tac1-912**

15:50–16:10

Existing work on domain adaptation for statistical machine translation has consistently assumed access to a small sample from the test distribution (target domain) at training time. In practice, however, the target domain may not be known at training time or it may change to match user needs. In such situations, it is natural to push the system to make safer choices, giving higher preference to domain invariant translations, which work well across domains, over risky domain-specific alternatives. We encode this intuition by (1) inducing latent subdomains from the training data only;(2)introducing features which measure how specialized phrases are to individual induced sub-domains; (3) estimating feature weights on out-of-domain data(rather than on the target domain). We conduct experiments on three language pairs and a number of different domains. We observe consistent improvements over a baseline which does not explicitly reward domain invariance.

papers-388**papers-388**

16:10–16:30

The existing machine translation systems, whether phrase-based or neural, have relied almost exclusively on word-level modelling with explicit segmentation. In this paper, we ask a fundamental question: can neural machine translation generate a character sequence without any explicit segmentation? To answer this question, we evaluate an attention-based encoder–decoder with a subword-level encoder and a character-level decoder on four language pairs—En-Cs, En-De, En-Ru and En-Fi— using the parallel corpora from WMT’15. Our experiments show that the models with a character-level decoder outperform the ones with a subword-level decoder on all of the four language pairs. Furthermore, the ensembles of neural models with a character-level decoder outperform the state-of-the-art non-neural machine translation systems on En-Cs, En-De and En-Fi and perform comparably on En-Ru.

papers-506**papers-506**

16:30–16:50

Discriminative translation models utilizing source context have been shown to help statistical machine translation performance. We propose a novel extension of this work using target context information. Surprisingly, we show that this model can be efficiently integrated directly in the decoding process. Our approach scales to large training data sizes and results in consistent improvements in translation quality. We also provide an analysis comparing the strengths of the baseline source-context model with our extended source-context and target-context model and we show that our extension allows us to better capture morphological coherence. Our work is freely available as part of Moses.

papers-145**papers-145**

16:50–17:10

Neural machine translation (NMT) models typically operate with a fixed vocabulary, but translation is an open-vocabulary problem. Previous work addresses the translation of out-of-vocabulary words by backing off to a dictionary. In this paper, we introduce a simpler and more effective approach, making the NMT model capable of open-vocabulary translation by encoding rare and unknown words as sequences of subword units. This is based on the intuition that various word classes are translatable via smaller units than words, for instance names (via character copying or transliteration), compounds (via compositional translation), and cognates and loanwords (via phonological and morphological transformations). We discuss the suitability of different word segmentation techniques, including simple character n-gram models and a segmentation based on the byte pair encoding compression algorithm, and empirically show that subword models improve over a back-off

dictionary baseline for the WMT 15 translation tasks English-German and English-Russian by up to 1.1 and 1.3 BLEU, respectively.

6D: Discourse

Plaza Ballroom F

Chair:

papers-483

papers-483

15:30–15:50

Word pairs, which are one of the most easily accessible features between two text segments, have been proven to be very useful for detecting the discourse relations held between text segments. However, because of the data sparsity problem, the performance achieved by using word pair features is limited. In this paper, in order to overcome the data sparsity problem, we propose the use of word embeddings to replace the original words. Moreover, we adopt a gated relevance network to capture the semantic interaction between word pairs, and then aggregate those semantic interactions using a pooling layer to select the most informative interactions. Experimental results on Penn Discourse Tree Bank show that the proposed method without using manually designed features can achieve better performance on recognizing the discourse level relations in all of the relations.

papers-225

papers-225

15:50–16:10

Quotation detection is the task of locating spans of quoted speech in text. The state of the art treats this problem as a sequence labeling task and employs linear-chain conditional random fields. We question the efficacy of this choice: The Markov assumption in the model prohibits it from making joint decisions about the begin, end, and internal context of a quotation. We perform an extensive analysis with two new model architectures. We find that (a), simple boundary classification combined with a greedy prediction strategy is competitive with the state of the art; (b), a semi-Markov model significantly outperforms all others, by relaxing the Markov assumption.

papers-580

papers-580

16:10–16:30

This paper addresses the problem of speech act recognition in written asynchronous conversations (e.g., fora, emails). We propose a class of conditional structured models defined over arbitrary graph structures to capture the conversational dependencies between sentences. Our models use sentence representations encoded by a long short term memory (LSTM) recurrent neural model. Empirical evaluation shows the effectiveness of our approach over existing ones: (i) - LSTMs provide better task-specific representations, and (ii) - the global joint model improves over local models.

papers-382

papers-382

16:30–16:50

This paper describes the first robust approach to automatically labeling clauses with their situation entity type (Smith 2003), capturing aspectual phenomena at the clause level which are relevant for interpreting both semantics at the clause level and discourse structure. Previous work on this task used a small data set from a limited domain, and relied mainly on words as features, an approach which is impractical in larger settings. We provide a new corpus of texts from 13 genres (40,000 clauses) annotated with situation entity types. We show that our sequence labeling approach using distributional information in the form of Brown clusters, as well as syntactic-semantic features targeted to the task, is robust across genres, reaching accuracies of up to 76%.

tacl-906

tacl-906

16:50–17:10

This paper presents an empirical study of linguistic formality. We perform an analysis of humans' perceptions of formality in four different genres. These findings are used to develop a statistical model for predicting formality, which is evaluated under different feature settings and genres. We apply our model to an investigation of formality in online discussion forums, and present findings consistent with theories of formality and linguistic coordination.

6E: Language and vision

Plaza Ballroom F

Chair:

papers-796**papers-796**

15:30–15:50

Activities and events in our lives are structural, be it a vacation, a camping trip, or a wedding. While individual details vary, there are characteristic patterns that are specific to each of these scenarios. For example, a wedding typically consists of a sequence of events such as walking down the aisle, exchanging vows, and dancing. In this paper, we present a data-driven approach to learning event knowledge from a large collection of photo albums. We formulate the task as constrained optimization to induce the prototypical temporal structure of an event, integrating both visual and textual cues. Comprehensive evaluation demonstrates that it is possible to learn multimodal knowledge of event structure from noisy web content.

papers-288**papers-288**

15:50–16:10

A fundamental problem in artificial intelligence is automatically generating a natural language description of an image. This task involves both computer vision and natural language processing and is called “image caption generation.” Research on image caption generation has typically focused on taking in an image and generating a caption in English as existing image caption corpora are mostly in English. The lack of corpora in languages other than English is an issue, especially for morphologically rich languages such as Japanese. There is thus a need for corpora sufficiently large for image captioning in other languages. We have developed a Japanese version of the MS COCO caption dataset and a generative model based on a deep recurrent architecture that takes in an image and uses this Japanese version of the dataset to generate a caption in Japanese. As the Japanese portion of the corpus is small, our model was designed to transfer the knowledge representation obtained from the English portion into the Japanese portion. Experiments showed that the resulting bilingual comparable corpus has better performance than a monolingual corpus, indicating that image understanding using a resource-rich language benefits a resource-poor language.

papers-041**papers-041**

16:10–16:30

We study the problem of automatically building hypernym taxonomies from textual and visual data. Previous works in taxonomy induction generally ignore the increasingly prominent visual data, which encode important perceptual semantics. Instead, we propose a probabilistic model for taxonomy induction by jointly leveraging text and images. To avoid hand-crafted feature engineering, we design end-to-end features based on distributed representations of images and words. The model is discriminatively trained given a small set of existing ontologies and is capable of building full taxonomies from scratch for a collection of unseen conceptual label items with associated images. We evaluate our model and features on the WordNet hierarchies, where our system outperforms previous approaches by a large gap.

papers-793**papers-793**

16:30–16:50

There has been an explosion of work in the vision & language community during the past few years from image captioning to video transcription, and answering questions about images. These tasks have focused on literal descriptions of the image. To move beyond the literal, we choose to explore how questions about an image are often directed at commonsense inference and the abstract events evoked by objects in the image. In this paper, we introduce the novel task of Visual Question Generation, where the system is tasked with asking a natural and engaging question when shown an image. We provide three datasets which cover a variety of images from object-centric to event-centric, with considerably more abstract training data than provided to state-of-the-art captioning systems thus far. We train and test several generative and retrieval models to tackle the task of VQG. Evaluation results show that while such models ask reasonable questions for a variety of images, there is still a wide gap with human performance which motivates further work on connecting images with commonsense knowledge and pragmatics. Our proposed task offers a new challenge to the community which we hope furthers interest in exploring deeper connections between vision & language.

papers-596**papers-596**

16:50–17:10

Linguistics studies have shown that action verbs often denote some Change of State (CoS) as the result of an action. However, the causality of action verbs and its potential connection with the physical world has not been systematically explored. To address this limitation, this paper presents a study on physical causality of action verbs and their implied changes in the physical world. We first conducted a crowd-sourcing experiment and identified eighteen categories of physical causality for action verbs. For a subset of these categories, we then defined a set of detectors that detect the corresponding change from visual perception of the physical environment. We further incorporated physical causality modeling and state detection in grounded language understanding. Our empirical studies have demonstrated the effectiveness of causality modeling in grounding language to perception.

6F: Summarization

Plaza Ballroom F

*Chair:***papers-508****papers-508**

16:30–16:50

This paper presents a problem-reduction approach to extractive multi-document summarization: we propose a reduction to the problem of scoring individual sentences with their ROUGE scores based on supervised learning. For the summarization, we solve an optimization problem where the ROUGE score of the selected summary sentences is maximized. To this end, we derive an approximation of the ROUGE-N score of a set of sentences, and define a principled discrete optimization problem for sentence selection. Mathematical and empirical evidence suggests that the sentence selection step is solved almost exactly, thus reducing the problem to the sentence scoring task. We perform a detailed experimental evaluation on two DUC datasets to demonstrate the validity of our approach.

tac1-908**tac1-908**

16:50–17:10

Most recent sentence simplification systems use basic machine translation models to learn lexical and syntactic paraphrases from a manually simplified parallel corpus. These methods are limited by the quality and quantity of manually simplified corpora, which are expensive to build. In this paper, we conduct an in-depth adaptation of statistical machine translation to perform text simplification, taking advantage of large-scale paraphrases learned from bilingual texts and a small amount of manual simplifications with multiple references. Our work is the first to design automatic metrics that are effective for tuning and evaluating simplification systems, which will facilitate iterative development for this task.

6G: Learner language

Plaza Ballroom F

*Chair:***papers-002****papers-002**

15:30–15:50

There has been almost no work on phrase structure annotation and parsing specially designed for learner English despite the fact that they are useful for representing the structural characteristics of learners. To address this problem, in this paper, we first propose a phrase structure annotation scheme for learner English and annotate two different learner corpora using it. Second, we show their usefulness, reporting on (a) inter-annotator agreement rate, (b) characteristic CFG rules in the corpora, and (c) parsing performance on them. In addition, we explore methods to improve phrase structure parsing for learner English (achieving an F-measure of 0.878). Finally, we release the full annotation guidelines, the annotated data, and the improved parser model for learner English to the public.

papers-461**papers-461**

15:50–16:10

We present half-life regression (HLR), a novel model for spaced repetition practice with applications to second language acquisition. HLR combines psycholinguistic theory with modern machine learning techniques, indirectly estimating the “half-life” of a word or concept in a student’s long-term memory. We use data from Duolingo — a popular online language learning application — to fit HLR models, reducing error by 45%+ compared to several baselines at predicting student recall rates. HLR model weights also shed light on which linguistic concepts are systematically challenging for second language learners. Finally, HLR was able to improve Duolingo daily student engagement by 12% in an operational user study.

tacl-915**tacl-915**

16:10–16:30

The field of grammatical error correction (GEC) has grown substantially in recent years, with research directed at both evaluation metrics and improved system performance against those metrics. One unvisited assumption, however, is the reliance of GEC evaluation on error-coded corpora, which contain specific labeled corrections. We examine current practices and show that GEC’s reliance on such corpora unnaturally constrains annotation and automatic evaluation, resulting in (a) sentences that do not sound acceptable to native speakers and (b) system rankings that do not correlate with human judgments. In light of this, we propose an alternate approach that jettisons costly error coding in favor of unannotated, whole-sentence rewrites. We compare the performance of existing metrics over different gold-standard annotations, and show that automatic evaluation with our new annotation scheme has very strong correlation with expert rankings ($\lambda = 0.82$). As a result, we advocate for a fundamental and necessary shift in the goal of GEC, from correcting small, labeled error types, to producing text that has native fluency.

papers-832**papers-832**

16:30–16:50

Foreign language learners can acquire new vocabulary by using cognate and context clues when reading. We devise an experimental framework to measure such incidental comprehension using Amazon Mechanical Turk (MTurk). With the data collected, we train a graphical model to jointly predict which words users will be able to comprehend, using features such as edit distance (cognate clues) and pointwise mutual information (context clues). The results show that our model’s predictions correlate with learner understanding, enabling the construction of comprehensible text for personalized foreign language education.

papers-240**papers-240**

16:50–17:10

We present a computational analysis of three varieties of language: native, advanced non-native, and translation. Our goal is to investigate the similarities and differences between non-native language productions and translations, contrasting both with native language. Using a collection of computational methods we establish three main results: (1) the three types of texts are easily distinguishable; (2) non-native language and translations are closer to each other than each of them is to native language; and (3) some of these characteristics are source- and native-language-dependent, while others are not, reflecting, perhaps, unified principles that similarly affect translations and non-native language.

Poster and dinner session II (includes Demonstrations B)

Time: 5:30–7:00

Location: Plaza Ballroom A, B, & C

Posters: Long papers

papers-663

papers-663

We present a pairwise context-sensitive Autoencoder for computing text pair similarity. Our model encodes input text into context-sensitive representations and uses them to compute similarity between text pairs. Our model outperforms the state-of-the-art models in two semantic retrieval tasks and a contextual word similarity task. For retrieval, our unsupervised approach that merely ranks inputs with respect to the cosine similarity between their hidden representations shows comparable performance with the state-of-the-art supervised models and in some cases outperforms them.

papers-513

papers-513

Online news editors ask themselves the same question many times: what is missing in this news article to go online? This is not an easy question to be answered by computational linguistic methods. In this work, we address this important question and characterise the constituents of news article editorial quality. More specifically, we identify 14 aspects related to the content of news articles. Through a correlation analysis, we quantify their independence and relation to assessing an article's editorial quality. We also demonstrate that the identified aspects, when combined together, can be used effectively in quality control methods for online news.

papers-184

papers-184

Named Entity Disambiguation (NED) algorithms disambiguate mentions of named entities with respect to a knowledge-base, but sometimes the context might be poor or misleading. In this paper we introduce the acquisition of two kinds of background information to alleviate that problem: entity similarity and selectional preferences for syntactic positions. We show, using a generative model for NED, that the additional sources of context are complementary, and improve results in the CoNLL2003 and TAC KBP DEL 2014 datasets, yielding the third best and the best results, respectively. We provide examples and analysis which show the value of the acquired background information.

papers-656

papers-656

Finding paraphrases in text is an important task with implications for generation, summarization and question answering, among other applications. Of particular interest to those applications is the specific formulation of the task where the paraphrases are templated, which provides an easy way to lexicalize one message in multiple ways by simply plugging in the relevant entities. Previous work has focused on mining paraphrases from parallel and comparable corpora, or mining very short sub-sentence synonyms and paraphrases. In this paper we present an approach which combines distributional and KB-driven methods to allow robust mining of sentence-level paraphrasal templates, utilizing a rich type system for the slots, from a plain text corpus.

papers-489

papers-489

We cast sentence boundary detection and syntactic parsing as a joint problem, so an entire text document forms a training instance for transition-based dependency parsing. When trained with an early update or max-violation strategy for inexact search, we observe that only a tiny part of these very long training instances is ever exploited. We demonstrate this effect by extending the ArcStandard transition system with swap for the joint prediction task. When we use an alternative update strategy, our models are considerably better on both tasks and train in substantially less time compared to models trained with early update/max-violation. A

comparison between a standard pipeline and our joint model furthermore empirically shows the usefulness of syntactic information on the task of sentence boundary detection.

papers-770
papers-770

Precise evaluation metrics are important for assessing progress in high-level language generation tasks such as machine translation or image captioning. Historically, these metrics have been evaluated using correlation with human judgment, however human-derived scores are often alarmingly inconsistent and also limited in their ability to identify precise areas of weakness. In this paper, we perform a case study for metric evaluation based on known sentence corruptions which serve as comprehensive unit tests for comparing metrics. We analyze three broad categories of sentence transformations and evaluate how successfully several common metrics can detect them. We find that not only are human annotations heavily inconsistent in this study, but that the Metric Unit TesT analysis is better able to capture precise shortcomings of particular metrics (e.g. comparing passive and active sentences) than a simple correlation with human judgment would.

papers-474
papers-474

For many applications, the query speed of N-gram language models is a computational bottleneck. Although massively parallel hardware like GPUs offer a potential solution to this bottleneck, exploiting this hardware requires a careful rethinking of basic algorithms and data structures. We present the first language model designed for such hardware, using B-trees to maximize data parallelism and minimize memory footprint and latency. Compared with a single-threaded instance of KenLM (Heafield, 2011), a highly optimized CPU-based language model, our GPU implementation produces identical results with a smaller memory footprint and a sixfold increase in throughput on a batch query task. When we saturate both devices, the GPU delivers nearly twice the throughput per hardware dollar even when the CPU implementation uses faster data structures. Our implementation is freely available at <https://github.com/XapaJlaMnu/gLM>

papers-600
papers-600

Morphologically rich languages often lack the annotated linguistic resources required to develop accurate natural language processing tools. We propose models suitable for training morphological taggers with rich tagsets for low-resource languages without using direct supervision. Our approach extends existing approaches of projecting part-of-speech tags across languages, using bitext to infer constraints on the possible tags for a given word type or token. We propose a tagging model using Wsabie, a discriminative embedding-based model with rank-based learning. In our evaluation on 11 languages, on average this model performs on par with a baseline weakly-supervised HMM, while being more scalable. Multilingual experiments show that the method performs best when projecting between related language pairs. Despite the inherently lossy projection, we show that the morphological tags predicted by our models improve the downstream performance of a parser by +0.6 LAS on average.

papers-733
papers-733

While end-to-end neural machine translation (NMT) has made remarkable progress recently, NMT systems only rely on parallel corpora for parameter estimation. Since parallel corpora are usually limited in quantity, quality, and coverage, especially for low-resource languages, it is appealing to exploit monolingual corpora to improve NMT. We propose a semi-supervised approach for training NMT models on the concatenation of labeled (parallel corpora) and unlabeled (monolingual corpora) data. The central idea is to reconstruct the monolingual corpora using an autoencoder, in which the source-to-target and target-to-source translation models serve as the encoder and decoder, respectively. Our approach can not only exploit the monolingual corpora of the target language, but also of the source language. Experiments on the Chinese-English dataset show that our approach achieves significant improvements over state-of-the-art SMT and NMT systems..

papers-096
papers-096

Training neural network language models over large vocabularies is computationally costly compared to count-based models such as Kneser-Ney. We present a systematic comparison of neural strategies to represent and train large vocabularies, including softmax, hierarchical softmax, target sampling, noise contrastive estimation

and self normalization. We extend self normalization to be a proper estimator of likelihood and introduce an efficient variant of softmax. We evaluate each method on three popular benchmarks, examining performance on rare words, the speed/accuracy trade-off, complementarity to Kneser-Ney.

papers-603**papers-603**

Distributional semantic models (DSMs) are often evaluated on artificial similarity datasets containing single words or fully compositional phrases. We present a large-scale multilingual evaluation of DSMs for predicting the degree of semantic compositionality of nominal compounds on 4 datasets for English and French. We build a total of 816 DSMs and perform 2,856 evaluations using word2vec, GloVe, and PPMI-based models. In addition to the DSMs, we compare the impact of different parameters, such as level of corpus preprocessing, context window size and number of dimensions. The results obtained have a high correlation with human judgments, being comparable to or outperforming the state of the art for some datasets (Spearman $r=.82$ for the Reddy dataset).

papers-051**papers-051**

We describe and evaluate a simple method to extract parallel sentences from comparable corpora. The approach, termed STACC, is based on expanded lexical sets and the Jaccard similarity coefficient. We evaluate our system against state-of-the-art methods on a large range of datasets in different domains, for ten language pairs, showing that it either matches or outperforms current methods across the board and gives significantly better results on the noisiest datasets. STACC is a portable method, requiring no particular adaptation for new domains or language pairs, thus enabling the efficient mining of parallel sentences in comparable corpora.

papers-654**papers-654**

We propose a joint formulation for learning task-specific cross-lingual word embeddings, along with classifiers for that task. Unlike prior work, which first learns the embeddings from parallel data and then plugs them in a supervised learning problem, our approach is one-shot: a single optimization problem combines a co-regularizer for the multilingual embeddings with a task-specific loss. We present theoretical results showing the limitation of Euclidean co-regularizers to increase the embedding dimension, a limitation which does not exist for other co-regularizers (such as the l_1 -distance). Despite its simplicity, our method achieves state-of-the-art accuracies on the RCV1/RCV2 dataset when transferring from English to German, with training times below 1 minute. On the TED Corpus, we obtain the highest reported scores on 10 out of 11 languages.

papers-153**papers-153**

The effort of understanding the meaning of words is central to the NLP community. Coarse-grained semantic categories such as supersenses were proven useful for a range of downstream tasks such as question answering or machine translation. To date, however, no effort has been put into integrating the supersenses into distributional word representations. In this work, we present a novel joint embedding model of words and supersenses, providing insights into the relationship between words and supersenses in the same vector space. Using these embeddings in a deep neural network model, we demonstrate that the supersense enrichment leads to a significant improvement in a range of downstream classification tasks. We also manifest the utility of these embeddings for predicting supersenses, achieving competitive results.

papers-289**papers-289**

We present a discriminative model for single-document summarization that integrally combines compression and anaphoricity constraints. Our model selects textual units to include in the summary based on a rich set of sparse features whose weights are learned on a large corpus. We allow for the deletion of content within a sentence when that deletion is licensed by compression rules; in our framework, these are implemented as dependencies between subsentential units of text. Anaphoricity constraints then improve cross-sentence coherence by guaranteeing that, for each pronoun included in the summary, the pronoun's antecedent is included as well or the pronoun is rewritten as a full mention. When trained end-to-end, our final system outperforms prior work on both ROUGE as well as on human judgments of linguistic quality.

papers-478

papers-478

Parsing for a wide variety of grammar formalisms can be performed by intersecting finite tree automata. However, naive implementations of parsing by intersection are very inefficient. We present techniques that speed up tree-automata-based parsing, to the point that it becomes practically feasible on realistic data when applied to context-free, TAG, and graph parsing. For graph parsing, we obtain the best runtimes in the literature.

papers-428

papers-428

Distributional semantics creates vector-space representations that capture many forms of semantic similarity, but their relation to semantic entailment has been less clear. We propose a vector-space model which provides a formal foundation for a distributional semantics of entailment. Using a mean-field approximation, we develop approximate inference procedures and entailment operators over vectors of probabilities of features being known (verses unknown). We use this framework to reinterpret an existing distributional-semantic model as approximating an entailment-based model of the distributions of words in contexts, thereby predicting lexical entailment relations. In both unsupervised and semi-supervised experiments on hyponymy detection, we get substantial improvements over previous results.

papers-688

papers-688

We propose a new unsupervised learning model, hidden softmax sequence model (HSSM), based on Boltzmann machine for dialogue structure analysis. The model employs three types of units in the hidden layer to discovery dialogue latent structures: softmax units which represent latent states of utterances; binary units which represent latent topics specified by dialogues; and a binary unit that represents the global general topic shared across the whole dialogue corpus. In addition, the model contains extra connections between adjacent hidden softmax units to formulate the dependency between latent states. Two different kinds of real world dialogue corpora, Twitter-Post and AirTicketBooking, are utilized for extensive comparing experiments, and the results illustrate that the proposed model outperforms state-of-the-art popular approaches.

papers-772

papers-772

High quality source code is often paired with high level summaries of the computation it performs, for example in code documentation or in descriptions posted in online forums. Such summaries are extremely useful for applications such as code search but are expensive to manually author, hence only done for a small fraction of all code that is produced. In this paper, we present the first completely data-driven approach for generating high level summaries of source code. Our model, CODE-NN, uses Long Short Term Memory (LSTM) networks with attention to produce sentences that describe C# code snippets and SQL queries. CODE-NN is trained on a new corpus that is automatically collected from StackOverflow, which we release. Experiments demonstrate strong performance on two tasks: (1) code summarization, where we establish the first end-to-end learning results and outperform strong baselines, and (2) code retrieval, where our learned model improves the state of the art on a recently introduced C# benchmark by a large margin.

papers-446

papers-446

To create accessible content for deaf users, we investigate automatically synthesizing animations of American Sign Language (ASL), including grammatically important facial expressions and head movements. Based on recordings of humans performing various types of syntactic face and head movements (which include idiosyncratic variation), we evaluate the efficacy of Continuous Profile Models (CPMs) at identifying an essential “latent trace” of the performance, for use in producing ASL animations. A metric-based evaluation and a study with deaf users indicated that this approach was more effective than a prior method for producing animations.

papers-416

papers-416

There are numerous studies suggesting that published news stories have an important effect on the direction of the stock market, its volatility, the volume of trades, and the value of individual stocks mentioned in the news. There is even some published research suggesting that automated sentiment analysis of news documents,quarterly reports, blogs and/or twitter data can be productively used as part of a trading strategy. This paper presents just such a family of trading strategies, and then uses this application to re-examine some of

the tacit assumptions behind how sentiment analyzers are generally evaluated, in spite of the contexts of their application. This discrepancy comes at a cost.

papers-251

papers-251

The run time complexity of state-of-the-art inference algorithms in graph-based dependency parsing is super-linear in the number of input words (n). Recently, pruning algorithms for these models have shown to cut a large portion of the graph edges, with minimal damage to the resulting parse trees. Solving the inference problem in run time complexity determined solely by the number of edges (m) is hence of obvious importance. We propose such an inference algorithm for first-order models, which encodes the problem as a minimum spanning tree (MST) problem in an *undirected graph*. This allows us to utilize state-of-the-art undirected MST algorithms whose run time is $O(m)$ at expectation and with a very high probability. A directed parse tree is then inferred from the undirected MST and is subsequently improved with respect to the directed parsing model through local greedy updates, both steps running in $O(n)$ time. In experiments with 18 languages, a variant of the first-order MSTParser (McDonald et al., 2005b) that employs our algorithm performs very similarly to the original parser that runs an $O(n^2)$ directed MST inference.

papers-593

papers-593

Conventional topic models are ineffective for topic extraction from microblog messages since the lack of structure and context among the posts renders poor message-level word co-occurrence patterns. In this work, we organize microblog posts as conversation trees based on reposting and replying relations, which enrich context information to alleviate data sparseness. Our model generates words according to topic dependencies derived from the conversation structures. In specific, we differentiate messages as leader messages, which initiate key aspects of previously focused topics or shift the focus to different topics, and follower messages that do not introduce any new information but simply echo topics from the messages that they repost or reply. Our model captures the different extents that leader and follower messages may contain the key topical words, thus further enhances the quality of the induced topics. The results of thorough experiments demonstrate the effectiveness of our proposed model.

papers-133

papers-133

Distant supervised relation extraction has been widely used in finding novel relational facts from text. However, distant supervision inevitably accompanies with the wrong labelling problem, and these noisy data will substantially hurt the performance of relation extraction. To alleviate this issue, we propose an attention-based model for relation extraction. In this model, we employ convolutional neural networks to embed the semantics of sentences. Afterwards, we build sentence-level attention over multiple instances, which is expected to dynamically reduce the weights of those noisy instances during learning. Experimental results on real-world datasets show that, our model can effectively reduce the influence of wrong labelling instances and achieves significant and consistent improvements on relation extraction as compared with baselines.

papers-139

papers-139

Frames defined in FrameNet (FN) share highly similar structures with events in ACE event extraction program. An event in ACE is composed of an event trigger and a set of arguments. Analogously, a frame in FN is composed of a lexical unit and a set of frame elements, which play similar roles as triggers and arguments of ACE events respectively. Besides having similar structures, many frames in FN actually express certain types of events. The above observations motivate us to explore whether there exists a good mapping from frames to event-types and if it is possible to improve event detection by using FN. In this paper, we propose a global inference approach to detect events in FN. Further, based on the detected results, we analyze possible mappings from frames to event-types. Finally, we improve the performance of event detection and achieve a new state-of-the-art result by using the events automatically detected from FN.

papers-361

papers-361

Solving simple arithmetic word problems is one of the challenges in Natural Language Understanding. This paper presents a novel method to learn to use formulas to solve simple arithmetic word problems. Our system,

analyzes each of the sentences to identify the variables and their attributes; and automatically maps this information into a higher level representation. It then uses that representation to recognize the presence of a formula along with its associated variables. An equation is then generated from the formal description of the formula. In the training phase, it learns to score the <formula, variables> pair from the systematically generated higher level representation. It is able to solve 86.07% of the problems in a corpus of standard primary school test questions and beats the state-of-the-art by a margin of 8.07%.

papers-620

papers-620

In this paper we explore the correlation between the sound of words and their meaning, by testing if the polarity ("good guy" or "bad guy") of a character's role in a work of fiction can be predicted by the name of the character in the absence of any other context. Our approach is based on phonological and other features proposed in prior theoretical studies of fictional names. These features are used to construct a predictive model over a manually annotated corpus of characters from motion pictures. By experimenting with different mixtures of features, we identify phonological features as being the most discriminative by comparison to social and other types of features, and we delve into a discussion of specific phonological and phonotactic indicators of a character's role's polarity.

papers-208

papers-208

We examine adjective-noun (AN) composition in the task of recognizing textual entailment (RTE). We analyze behavior of ANs in large corpora and show that, despite conventional wisdom, adjectives do not always restrict the denotation of the nouns they modify. We use natural logic to characterize the variety of entailment relations that can result from AN composition. Predicting these relations depends on context and on common-sense knowledge, making AN composition especially challenging for current RTE systems. We demonstrate the inability of current state-of-the-art systems to handle AN composition in a simplified RTE task which involves the insertion of only a single word.

papers-842

papers-842

Essay stance classification, the task of determining how much an essay's author agrees with a given proposition, is an important yet under-investigated subtask in understanding an argumentative essay's overall content. We introduce a new corpus of argumentative student essays annotated with stance information and propose a computational model to automatically predicting essay stance. In an evaluation on 826 essays, our approach significantly outperforms four baselines, one of which relies on features previously developed specifically for stance classification on student essays, yielding relative error reductions of at least 11.3% and 5.3%, in micro and macro f-score, respectively.

papers-196

papers-196

Word segmentation is a fundamental task for Chinese language processing. However, with the successive improvements, the standard metric is becoming hard to distinguish state-of-the-art word segmentation systems. In this paper, we propose a new psychometric-inspired evaluation metric for Chinese word segmentation, which addresses to balance the very skewed word distribution at different levels of difficulty. The performance on a real evaluation shows that the proposed metric gives more reasonable and distinguishable scores and correlates well with human judgement. In addition, the proposed metric can be easily extended to evaluate other sequence labelling-based NLP tasks.

papers-202

papers-202

Today's extraction of temporal information for events heavily depends on annotated temporal links. These so called TLINKs capture the relation between pairs of event mentions and time expressions. One problem is that the number of possible TLINKs grows quadratic with the number of event mentions, therefore most annotation studies concentrate on links for mentions in the same or in adjacent sentences. However, as our annotation study shows, this restriction results for 58% of the event mentions in a less precise information when the event took place. This paper proposes a new annotation scheme to anchor events in time. Not only is the annotation effort much lower as it scales linear with the number of events, it also gives a more precise

anchoring when the events have happened as the complete document can be taken into account. Using this scheme, we annotated a subset of the TimeBank Corpus and compare our results to other annotation schemes. Additionally, we present some baseline experiments to automatically anchor events in time.

[papers-686](#)**[papers-686](#)**

We focus on two leading state-of-the-art approaches to grammatical error correction — machine learning classification and machine translation. Based on the comparative study of the two learning frameworks and through error analysis of the output of the state-of-the-art systems, we identify key strengths and weaknesses of each of these approaches and demonstrate their complementarity. In particular, the machine translation method learns from parallel data without requiring further linguistic input and is better at correcting complex mistakes. The classification approach possesses other desirable characteristics, such as the ability to easily generalize beyond what was seen in training, the ability to train without human-annotated data, and the flexibility to adjust knowledge sources for individual error types. Based on this analysis, we develop an algorithmic approach that combines the strengths of both methods. We present several systems based on resources used in previous work with a relative improvement of over 20% (and 7.4 F score points) over the previous state-of-the-art.

[papers-773](#)**[papers-773](#)**

Hand-crafted features based on linguistic and domain-knowledge play crucial role in determining the performance of disease name recognition systems. Such methods are further limited by the scope of these features or in other words, their ability to cover the contexts or word dependencies within a sentence. In this work, we focus on reducing such dependencies and propose a domain-invariant framework for the disease name recognition task. In particular, we propose various end-to-end recurrent neural network (RNN) models with conditional random fields (CRF) for the tasks of disease name recognition and their classification into four pre-defined categories. We also utilize convolution neural network (CNN) in cascade of RNN to get character-based embedded features and employ it with word-embedded features in our model. We compare our models with the state-of-the-art results for the two tasks on NCBI disease dataset. Our results for the disease mention recognition task indicate that state-of-the-art performance can be obtained without relying on feature engineering. Further our results significantly improved the state-of-the-art performance for the classification task of disease names.

[papers-223](#)**[papers-223](#)**

We present the first domain adaptation model for authorship attribution to leverage unlabeled data. The model includes extensions to structural correspondence learning needed to make it appropriate for the task. For example, we propose a median-based classification instead of the standard binary classification used in previous work. Our results show that punctuation-based character n-grams form excellent pivot features. We also show how singular value decomposition plays a critical role in achieving domain adaptation, and that replacing (instead of concatenating) non-pivot features with correspondence features yields better performance.

[papers-619](#)**[papers-619](#)**

The canonical word order of Japanese double object constructions has attracted considerable attention among linguists and has been a topic of many studies. However, most of these studies require either manual analyses or measurements of human characteristics such as brain activities or reading times for each example. Thus, while these analyses are reliable for the examples they focus on, they cannot be generalized to other examples. On the other hand, the trend of actual usage can be collected from a large corpus. Thus, in this paper, we assume that there is a relationship between the canonical word order and the proportion of each word order in a large corpus and present a corpus-based analysis of canonical word order of Japanese double object constructions.

[papers-829](#)**[papers-829](#)**

In this paper, with the help of knowledge base, we build and formulate a semantic space to connect the source and target languages, and apply it to the sequence-to-sequence framework to propose a Knowledge-Based Semantic Embedding (KBSE) method. In our KBSE method, the source sentence is firstly mapped into a knowledge based semantic space, and the target sentence is generated using a recurrent neural network with

the internal meaning preserved. Experiments are conducted on two translation tasks, the electric business data and movie data, and the results show that our proposed method can achieve outstanding performance, compared with both the traditional SMT method and the existing encoder-decoder model.

papers-559

papers-559

Entity linking (EL) is the task of disambiguating mentions in text by associating them with entries in a pre-defined database of mentions (persons, organizations, etc). Most previous EL research has focused mainly on one language, English, with less attention being paid to other languages, such as Spanish or Chinese. In this paper, we introduce LIEL, a Language Independent Entity Linking system, which provides an EL framework which, once trained on one language, works remarkably well on a number of different languages without change. LIEL makes a joint global prediction over the entire document, employing a discriminative re-ranking framework with many domain and language-independent feature functions. Experiments on numerous benchmark datasets, show that the proposed system, once trained on one language, English, outperforms several state-of-the-art systems in English (by 4 points) and the trained model also works very well on Spanish (14 points better than a competitor system), demonstrating the viability of the approach.

papers-368

papers-368

We discuss an approximate similarity search for word embeddings, which is an operation to approximately find embeddings close to a given vector. We compared several metric-based search algorithms with hash-, tree-, and graph-based indexing from different aspects. Our experimental results showed that a graph-based indexing exhibits robust performance and additionally provided useful information, e.g., vector normalization achieves an efficient search with cosine similarity.

papers-811

papers-811

Learning distributed representations for relation instances is a central technique in downstream NLP applications. In order to address semantic modeling of relational patterns, this paper constructs a new dataset that provides multiple similarity ratings for every pair of relational patterns on the existing dataset (Zeichner et al., 2012). In addition, we conduct a comparative study of different encoders including additive composition, RNN, LSTM, and GRU for composing distributed representations of relational patterns. We also present Gated Additive Composition, which is an enhancement of additive composition with the gating mechanism. Experiments show that the new dataset does not only enable detailed analyses of the different encoders, but also provides a gauge to predict successes of distributed representations of relational patterns in the relation classification task.

papers-812

papers-812

Anaphor resolution is an important task in NLP with many applications. Despite much research effort, it remains an open problem. The difficulty of the problem varies substantially across different sub-problems. One sub-problem, in particular, has been largely untouched by prior work despite occurring frequently throughout corpora: the anaphor that has multiple antecedents, which here we call multi-antecedent anaphors or m-anaphors. Current coreference resolvers restrict anaphors to at most a single antecedent. As we show in this paper, relaxing this constraint poses serious problems in coreference chain-building, where each chain is intended to refer to a single entity. This work provides a formalization of the new task with preliminary insights into multi-antecedent noun-phrase anaphors, and offers a method for resolving such cases that outperforms a number of baseline methods by a significant margin. Our system uses local agglomerative clustering on candidate antecedents and an existing coreference system to score clusters to determine which cluster of mentions is antecedent for a given anaphor. When we augment an existing coreference system with our proposed method, we observe a substantial increase in performance (0.6 absolute CoNLL F1) on an annotated corpus.

papers-105

papers-105

Labeling topics learned by topic models is a challenging problem. Previous studies have used words, phrases and images to label topics. In this paper, we propose to use text summaries for topic labeling. Several sentences are extracted from the most related documents to form the summary for each topic. In order to obtain

summaries with both high relevance, coverage and discrimination for all the topics, we propose an algorithm based on submodular optimization. Both automatic and manual analysis have been conducted on two real document collections, and we find 1) the summaries extracted by our proposed algorithm are superior over the summaries extracted by existing popular summarization methods; 2) the use of summaries as labels has obvious advantages over the use of words and phrases.

[papers-168](#)

[papers-168](#)

In this paper, we propose a neural network model for graph-based dependency parsing which utilizes Bidirectional LSTM (BLSTM) to capture richer contextual information instead of using high-order factorization, and enable our model to use much fewer features than previous work. In addition, we propose an effective way to learn sentence segment embedding on sentence-level based on an extra forward LSTM network. Although our model uses only first-order factorization, experiments on English Penn Treebank and Chinese Penn Treebank show that our model could be competitive with previous higher-order graph-based dependency parsing models and state-of-the-art models.

[papers-045](#)

[papers-045](#)

Recently, knowledge graph embedding, which projects symbolic entities and relations into continuous vector space, has become a new, hot topic in artificial intelligence. This paper proposes a novel generative model (**TransG**) to address the issue of **multiple relation semantics** that a relation may have multiple meanings revealed by the entity pairs associated with the corresponding triples. The new model can discover latent semantics for a relation and leverage a mixture of relation-specific component vectors to embed a fact triple. To the best of our knowledge, this is the first generative model for knowledge graph embedding, and at the first time, the issue of multiple relation semantics is formally discussed. Extensive experiments show that the proposed model achieves substantial improvements against the state-of-the-art baselines.

[papers-035](#)

[papers-035](#)

Existing knowledge-based question answering systems often rely on small annotated training data. While shallow methods like relation extraction are robust to data scarcity, they are less expressive than the deep meaning representation methods like semantic parsing, thereby failing at answering questions involving multiple constraints. Here we alleviate this problem by empowering a relation extraction method with additional evidence from Wikipedia. We first present a neural network based relation extractor to retrieve the candidate answers from Freebase, and then infer over Wikipedia to validate these answers. Experiments on the WebQuestions question answering dataset show that our method achieves an F1 of 53.3%, a substantial improvement over the state-of-the-art.

[papers-640](#)

[papers-640](#)

Semantic deficit is a symptom of language impairment in Alzheimer's disease (AD). We present a generalizable method for automatic generation of information content units (ICUs) for a picture used in a standard clinical task, achieving high recall, 96.8%, of human-supplied ICUs. We use the automatically generated topic model to extract semantic features, and train a random forest classifier to achieve an F-score of 0.74 in binary classification of controls versus people with AD using a set of only 12 features. This is comparable to results (0.72 F-score) with a set of 85 manual features. Adding semantic information to a set of standard lexicosyntactic and acoustic features improves F-score to 0.80. While control and dementia subjects discuss the same topics in the same contexts, controls are more informative per second of speech.

[papers-411](#)

[papers-411](#)

Part of the unique cultural heritage of China is the Chinese couplet. Given a sentence (namely an antecedent clause), people reply with another sentence (namely a subsequent clause) equal in length. Moreover, a special phenomenon is that corresponding characters from the same position in the two clauses match each other by following certain constraints on semantic and/or syntactic relatedness. Automatic couplet generation by computer is viewed as a difficult problem and has not been fully explored. In this paper, we formulate the task as a natural language generation problem using neural network structures. Given the issued antecedent clause,

the system generates the subsequent clause via sequential language modeling. To satisfy special characteristics of couplets, we incorporate the attention mechanism and polishing schema into the encoding-decoding process. The couplet is generated incrementally and iteratively. A comprehensive evaluation, using perplexity and BLEU measurements as well as human judgments, has demonstrated the effectiveness of our proposed approach.

Posters: Short papers

short-303

short-303

Neural Machine Translation (MT) has reached state-of-the-art results. However, one of the main challenges that neural MT still faces is dealing with very large vocabularies and morphologically rich languages. In this paper, we propose a neural MT system using character-based embeddings in combination with convolutional and highway layers to replace the standard lookup-based word representations. The resulting unlimited-vocabulary and affixaware source word embeddings are tested in a state-of-the-art neural MT based on an attention-based bidirectional recurrent neural network. The proposed MT scheme provides improved results even when the source language is not morphologically rich. Improvements up to 3 BLEU points are obtained in the German-English WMT task.

short-461

short-461

Bilingual models that capture the semantics of sentences are typically only evaluated on cross-lingual transfer tasks such as cross-lingual document categorization or machine translation. In this work, we evaluate the quality of the monolingual representations learned with a variant of the bilingual compositional model of Hermann and Blunsom (2014), when viewing translations in a second language as a semantic annotation as the original language text. We show that compositional objectives based on phrase translation pairs outperform compositional objectives based on bilingual sentences and on monolingual paraphrases.

short-425

short-425

Traditional event detection methods heavily rely on manually engineered rich features. Recent deep learning approaches alleviate this problem by automatic feature engineering. But such efforts, like tradition methods, have so far only focused on single-token event mentions, whereas in practice events can also be a phrase. We instead use forward-backward recurrent neural networks (FBRNNs) to detect events that can be either words or phrases. To the best our knowledge, this is one of the first efforts to handle multi-word events and also the first attempt to use RNNs for event detection. Experimental results demonstrate that FBRNN is competitive with the state-of-the-art methods on the ACE 2005 and the Rich ERE 2015 event detection tasks.

short-498

short-498

We describe the Iraq Body Count Corpus (IBC-C) dataset, the first substantial armed conflict-related dataset which can be used for conflict analysis. IBC-C provides a ground-truth dataset for conflict specific named entity recognition, slot filling, and event de-duplication. IBC-C is constructed using data collected by the Iraq Body Count project which has been recording casualties resulting from the ongoing war in Iraq since 2003. We describe the dataset's creation, how it can be used for the above three tasks and provide initial baseline results for the first task (named entity recognition) using Hidden Markov Models, Conditional Random Fields, and Recursive Neural Networks.

short-119

short-119

Uncovering thematic structures of SNS and blog posts is a crucial yet challenging task, because of the severe data sparsity induced by the short length of texts and diverse use of vocabulary. This hinders effective topic inference of traditional LDA because it infers topics based on document-level co-occurrence of words. To robustly infer topics in such contexts, we propose a latent concept topic model (LCTM). Unlike LDA, LCTM reveals topics via co-occurrence of latent concepts, which we introduce as latent variables to capture conceptual similarity of words. More specifically, LCTM models each topic as a distribution over the latent concepts,

where each latent concept is a localized Gaussian distribution over the word embedding space. Since the number of unique concepts in a corpus is often much smaller than the number of unique words, LCTM is less susceptible to the data sparsity. Experiments on the 20Newsgroups show the effectiveness of LCTM in dealing with short texts as well as the capability of the model in handling held-out documents with a high degree of OOV words.

short-388

short-388

This paper studies the effect of limited precision data representation and computation on word embeddings. We present a systematic evaluation of word embeddings with limited memory and discuss methods that directly train the limited precision representation with limited memory. Our results show that it is possible to use and train an 8\$-\$bit fixed-point value for word embedding without loss of performance in word/phrase similarity and dependency parsing tasks.

short-102

short-102

Classification of temporal textual data sequences is a common task in various domains such as social media and the Web. In this paper we propose to use Hawkes Processes for classifying sequences of temporal textual data, which exploit both temporal and textual information. Our experiments on rumour stance classification on four Twitter datasets show the importance of using the temporal information of tweets along with the textual content.

short-479

short-479

There are different definitions of what a troll is. Certainly, a troll can be somebody who teases people to make them angry, or somebody who offends people, or somebody who wants to dominate any single discussion, or somebody who tries to manipulate people's opinion (sometimes for money), etc. The last definition is the one that dominates the public discourse in Bulgaria and Eastern Europe, and this is our focus in this paper. In our work, we examine two types of opinion manipulation trolls: paid trolls that have been revealed from leaked "reputation management contracts" and "mentioned trolls" that have been called such by several different people. We show that these definitions are sensible: we build two classifiers that can distinguish a post by such a paid troll from one by a non-troll with 81-82% accuracy; the same classifier also achieves 81-82% accuracy on so called mentioned troll vs. non-troll posts.

short-114

short-114

Phrase table pruning is the act of removing phrase pairs from a phrase table to make it smaller, ideally removing the least useful phrases first. We propose a phrase table pruning method that formulates the task as a submodular function maximization problem, and solves it by using a greedy heuristic algorithm. The proposed method can scale with input size and long phrases, and experiments show that it achieves higher BLEU scores than state-of-the-art pruning methods.

short-384

short-384

Bidirectional long short-term memory (bi-LSTM) networks have recently proven successful for various NLP sequence modeling tasks, but little is known about their reliance to input representations, target languages, data set size, and label noise. We address these issues and evaluate bi-LSTMs with word, character, and unicode byte embeddings for POS tagging. We compare bi-LSTMs to traditional POS taggers across languages and data sizes. We also present a novel bi-LSTM model, which combines the POS tagging loss function with an auxiliary loss function that accounts for rare words. The model obtains state-of-the-art performance across 22 languages, and works especially well for morphologically complex languages. Our analysis suggests that bi-LSTMs are less sensitive to training data size and label corruptions (at small noise levels) than previously assumed.

short-358

short-358

In this paper, we propose LexVec, a new method for generating distributed word representations that uses low-rank, weighted factorization of the Positive Point-wise Mutual Information matrix via stochastic gradient descent, employing % of the PPMI matrix that generates a weighting scheme that assigns heavier penalties

for errors on frequent co-occurrences while still accounting for negative co-occurrence. Evaluation on word similarity and analogy tasks shows that LexVec matches and often outperforms state-of-the-art methods on many of these tasks.

short-193

short-193

We introduce an approach to train lexicalized parsers using bilingual corpora obtained by merging harmonized treebanks of different languages, producing parsers that can analyze sentences in either of the learned languages, or even sentences that mix both. We test the approach on the Universal Dependency Treebanks, training with MaltParser and MaltOptimizer. The results show that these bilingual parsers are more than competitive, as most combinations not only preserve accuracy, but some even achieve significant improvements over the corresponding monolingual parsers. Preliminary experiments also show the approach to be promising on texts with code-switching and when more languages are added.

short-040

short-040

Modern coreference resolution systems require linguistic and general knowledge typically sourced from costly, manually curated resources. Despite their intuitive appeal, results have been mixed. In this work, we instead implement fine-grained surface-level features motivated by cognitive theory. Our novel fine-grained feature specialisation approach significantly improves the performance of a strong baseline, achieving state-of-the-art results of 65.29 and 61.13% on CoNLL 2012 using gold and automatic preprocessing, with system extracted mentions.

short-404

short-404

This paper presents a novel approach to automated sentence completion based on pointwise mutual information (PMI). Feature sets are created by fusing the various types of input provided to other classes of language models, ultimately allowing multiple sources of both local and distant information to be considered. Furthermore, it is shown that additional precision gains may be achieved by incorporating feature sets of higher-order n-grams. Experimental results demonstrate that the PMI model outperforms all prior models and establishes a new state-of-the-art result on the Microsoft Research Sentence Completion Challenge.

short-317

short-317

Using corpus data of spoken dialogue, we examine the convergence of syntactic complexity levels between interlocutors in natural conversations, as it occurs within spans of topic episodes. The findings of general convergence in the Switchboard and BNC corpora are compatible with an information-theoretic model of dialogue and with Interactive Alignment Theory.

short-523

short-523

Nowadays, many scholarly messages are posted on Chinese microblogs and more and more researchers tend to find scholarly information on microblogs. In order to exploit microblogging to benefit scientific research, we propose a scholarly microblog recommendation system in this study. It automatically collects and mines scholarly information from Chinese microblogs, and makes personalized recommendations to researchers. We propose two different neural network models which learn the vector representations for both users and microblog texts. Then the recommendation is accomplished based on the similarity between a user's vector and a microblog text's vector. We also build a dataset for this task. The two embedding models are evaluated on the dataset and show good results compared to several baselines.

Demonstrations B

demos-087

demos-087

We present a computer-assisted language learning (CALL) system that generates fill-in-the-blank items for preposition usage. The system takes a set of carrier sentences as input, chooses a preposition in each sentence as the key, and then automatically generates distractors. It personalizes item selection for the user in two

ways. First, it logs items to which the user previously gave incorrect answers, and offers similar items in a future session as review. Second, it progresses from easier to harder sentences, to minimize any hindrance on preposition learning that might be posed by difficult vocabulary.

demos-086

demos-086

Boulder Learning has developed and evaluated a conversational multimedia virtual tutor for elementary school science. The operating principles for the tutor are grounded on research from education and cognitive science. This virtual science tutor is unique in that it engages children in spoken dialogs in which they learn to construct explanations of science phenomena presented in illustrations, animations and interactive simulations. Summative evaluations indicate that students are highly engaged in the tutoring sessions, and achieve learning outcomes equivalent to expert human tutors. Tutorials are developed through a process of recording and annotating data from sessions with students, and then updating tutor models. In initial evaluations, My Science Tutor (MyST) has been shown to produce significant learning gains with effect size $d=0.53$ compared to controls. It is enthusiastically supported by students and teachers. Teachers report that it is feasible to integrate into their curriculum.

demos-033

demos-033

screencast in youtube: https://youtu.be/QiV_ow7k2J4 system online at: <http://hum.csse.unimelb.edu.au:5000/>
Abstract We present pigeo, a Python geolocation prediction tool that predicts a location for a given text input or Twitter user. We discuss the design, implementation and application of pigeo, and empirically evaluate it. pigeo is able to geolocate informal text and is a very useful tool for users who require a free and easy-to-use, yet accurate geolocation service based on pre-trained models. Additionally, users can train their own models easily using pigeo's API.

demos-085

demos-085

We present a prototype of a novel technology for second language instruction. Our learn-by-reading approach lets a human learner acquire new words and constructions by encountering them in context. To facilitate reading comprehension, our technology presents mixed native language (L1) and second language (L2) sentences to a learner and allows them to interact with the sentences to make the sentences easier (more L1-like) or harder (more L2-like) to read. Eventually, our system should continuously track a learner's knowledge and learning style by modeling their interactions, including performance on a pop quiz feature. This will allow our system to generate personalized mixed-language texts for learners.

demos-071

demos-071

In this paper, we present Roleo, a web tool for visualizing the vector spaces generated by the evaluation of distributional memory (DM) models over thematic fit judgements. A thematic fit judgement is a rating of the selectional preference of a verb for an argument that fills a given thematic role. The DM approach to thematic fit judgements involves the construction of a sub-space in which a prototypical role-filler can be built for comparison to the noun being judged. We describe a publicly-accessible web tool that allows for querying and exploring these spaces as well as a technique for visualizing thematic fit sub-spaces efficiently for web use.

demos-044

demos-044

We introduce MediaGist, an online system for crosslingual analysis of aggregated news and commentaries based on summarization and sentiment analysis technologies. It is designed to assist journalists to detect and explore news topics, which are controversially reported or discussed in different countries. News articles from current week are clustered separately in currently 5 languages and the clusters are then linked across languages. Sentiment analysis provides a basis to compute controversy scores and summaries help to explore the differences. Recognized entities play an important role in most of the system's modules and provide another way to explore the data. We demonstrate the capabilities of MediaGist by listing highlights from the last week and present a rough evaluation of the system.

demos-042

demos-042

We introduce GoWvis, an interactive web application that represents any piece of text inputted by the user as a graph-of-words and leverages graph degeneracy and community detection to generate an extractive summary (keyphrases and sentences) of the inputted text in an unsupervised fashion. The entire analysis can be fully customized via the tuning of many text preprocessing, graph building, and graph mining parameters. Our system is thus well suited to educational purposes, exploration and early research experiments. The new summarization strategy we propose also shows promise.

demos-072

demos-072

We present a robust and efficient parallelizable multilingual UIMA-based platform for automatically annotating textual inputs with different layers of linguistic description, ranging from surface level phenomena all the way to deep discourse-level information. In particular, given an input text, the pipeline extracts: a) sentences and tokens, b) entity mentions (named entities, nominals, personal pronouns), c) syntactic information (constituency and dependency trees), d) opinionated expressions, e) relations between entity mentions, f) co-reference chains and g) wikified entities. The system is available in two versions: a standalone distribution enables design and optimization of user-specific sub-modules, whereas a server-client distribution allows for straightforward high-performance NLP processing, reducing the engineering cost for higher-level tasks.

demos-065

demos-065

We present new/s/leak, a novel tool developed for and with the help of journalists, which enables the automatic analysis and discovery of newsworthy stories from large textual datasets. We rely on different NLP preprocessing steps such named entity tagging, extraction of time expressions, entity networks, relations and metadata. The system features an intuitive web-based user interface based on network visualization combined with data exploring methods and various search and faceting mechanisms. We report the current state of the software and exemplify it with the WikiLeaks PlusD (Cablegate) data.

ACL Social Event: Boat trip

Tuesday, August 9, 2016, 7:00pm – 11:00pm (boarding between 7:00pm and 8:00pm

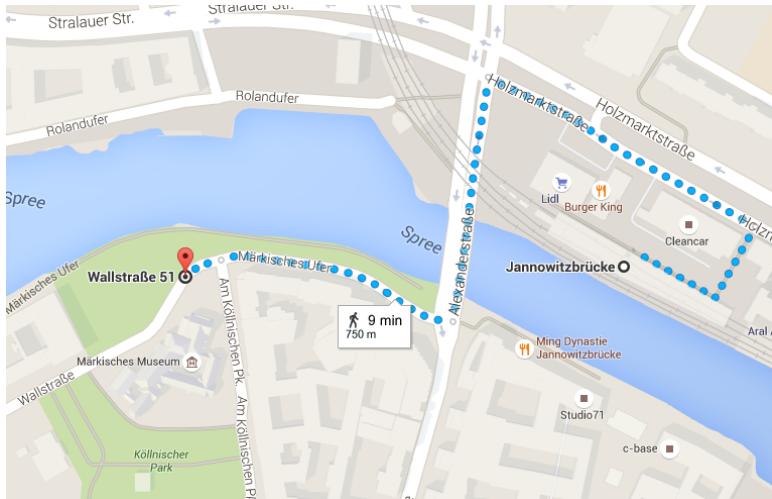
Märkisches Ufer, opposite of the Brazilian Embassy

Instead of just a dinner with a short social program, we would like to provide a unique cultural insight of Berlin. That is why we will take you on a boat tour of Berlin. The starting point is located within 5 minutes walking distance from U Märkisches Museum (line U2). The tour will take ca. 3 hours and finish at the starting point at Märkisches Ufer. People can enjoy the magnificent view which central Berlin offers during sunset as well as the show of lights which the city turns into in the evening.

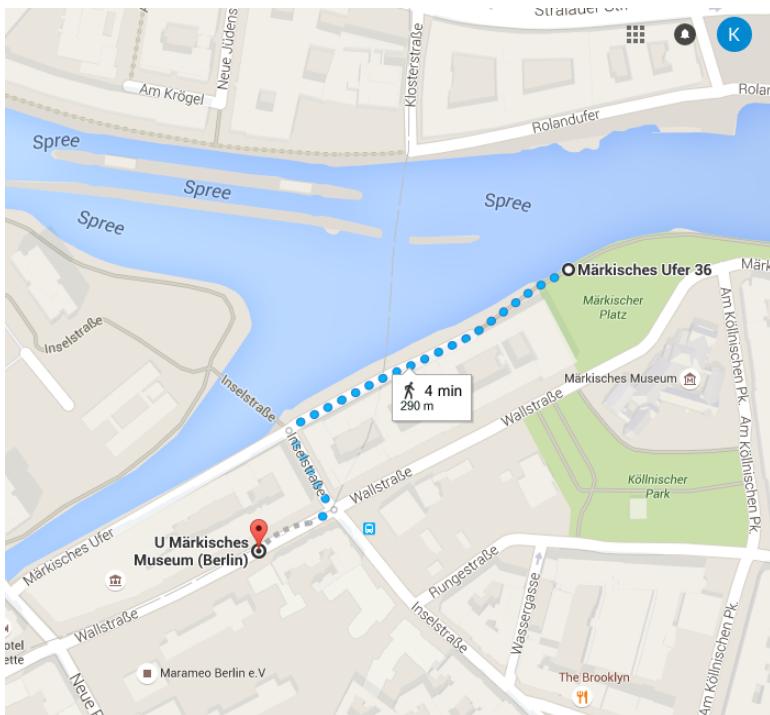
There will be English speaking guides who will give information about the main points of interests during the tour as well as tell interesting stories about Berlin, its people, and the city's history. The guides will be available for questions during the whole tour.

There will be a free selection of desserts as well as a free welcome drink. You can also purchase one additional free drink using a voucher that will be provided to you. You have to pay for any other drinks purchased in addition to these two free drinks.

Directions: Go to platform 5 on the top level of S+U Friedrichstraße (opposite of the poster session venue) and take any S-Bahn line departing from there for 3 stations to S+U Jannowitzbrücke. Finally, walk for ca. 8-10 minutes to the piers at Märkisches Ufer. The whole trip will take 15-2 minutes and there will be ACL personnel at both stations to direct and help you.



Alternatively, you make take subway line U6 from Friedrichstraße (direction Alt-Mariendorf) for 2 stations to U Stadmitte. There, change to U2 (direction Pankow) and get off at the third station which is U Märkisches Museum. The piers are located within 4 minutes walking distance. Note that there is no ACL personnel available to help you on this route. Nevertheless, this is also a convenient and easy way to get to the piers with less walking involved.



Time: Tuesday, August 9, boarding time is between 19:00 and 20:00. Whenever the capacity of a given ship is exhausted, it will depart immediately. Please try to be there at 19:00 or shortly after that for the boarding to go more quickly and smoothly!

Main Conference: Wednesday, August 10

Overview

9:00–9:40	President's talk	<i>dummy</i>			
9:40–10:10	Coffee break	<i>Plaza Ballroom D & E</i>			
10:10–11:50	Session 7 Outstanding papers I	Outstanding papers II			
11:50–1:30	Lunch break				
1:30–3:00	ACL business meeting (open to all)	<i>dummy</i>			
3:00–3:30	Coffee break	<i>Plaza Ballroom D & E</i>			
	Session 8				
3:30–4:50	Question answering II (short papers)	Word vectors III (short papers)	Topics and discourse (short papers)	Syntax and morphology (short papers)	Potpourri II (short papers)
5:00–6:30	Awards session	<i>dummy</i>			
6:30–7:00	Closing session	<i>dummy</i>			

Session 7 Overview – Wednesday, August 10, 2016

11:30 11:10 10:50 10:30 10:10

Track A	Track B
<i>Outstanding papers I</i> Plaza Ballroom F	<i>Outstanding papers II</i> Plaza Ballroom F
papers-863 papers-863	papers-232 papers-232
papers-696 papers-696	papers-730 papers-730
papers-121 papers-121	papers-258 papers-258
papers-094 papers-094	papers-142 papers-142
short-261 short-261	papers-587 papers-587

Parallel Session 7

7A: Outstanding papers I

Plaza Ballroom F

*Chair:***papers-863****papers-863**

10:10–10:30

Enabling a computer to understand a document so that it can answer comprehension questions is a central, yet unsolved goal of NLP. A key factor impeding its solution by machine learned systems is the limited availability of human-annotated data. Hermann et al. (2015) seek to solve this problem by creating over a million training examples by pairing CNN and Daily Mail news articles with their summarized bullet points, and show that a neural network can then be trained to give good performance on this task. In this paper, we conduct a thorough examination of this new reading comprehension task. Our primary aim is to understand what depth of language understanding is required to do well on this task. We approach this from one side by doing a careful hand-analysis of a small subset of the problems and from the other by showing that simple, carefully designed systems can obtain accuracies of 72.4% and 75.8% on these two datasets, exceeding current state-of-the-art results by over 5% and approaching what we believe is the ceiling for performance on this task.

papers-696**papers-696**

10:30–10:50

We introduce a new language learning setting relevant to building adaptive natural language interfaces. It is inspired by Wittgenstein's language games: a human wishes to accomplish some task (e.g., achieving a certain configuration of blocks), but can only communicate with a computer, who performs the actual actions (e.g., removing all red blocks). The computer initially knows nothing about language and therefore must learn it from scratch through interaction, while the human adapts to the computer's capabilities. We created a game in a blocks world and collected interactions from 100 people playing it. First, we analyze the humans' strategies, showing that using compositionality and avoiding synonyms correlates positively with task performance. Second, we compare computer strategies, showing how to quickly learn a semantic parsing model from scratch, and that modeling pragmatics further accelerates learning for successful players.

papers-121**papers-121**

10:50–11:10

Arbitrariness of the sign—the notion that the forms of words are unrelated to their meanings—is an underlying assumption of many linguistic theories. Two lines of research have recently challenged this assumption, but they produce differing characterizations of non-arbitrariness in language. Behavioral and corpus studies have confirmed the validity of localized form-meaning patterns manifested in limited subsets of the lexicon. Meanwhile, global (lexicon-wide) statistical analyses instead find diffuse form-meaning systematicity across the lexicon as a whole. We bridge the gap with an approach that can detect both local and global form-meaning systematicity in language. In the kernel regression formulation we introduce, form-meaning relationships can be used to predict words' distributional semantic vectors from their forms. Furthermore, we introduce a novel metric learning algorithm that can learn weighted edit distances that minimize kernel regression error. Our results suggest that the English lexicon exhibits far more global form-meaning systematicity than previously discovered, and that much of this systematicity is focused in localized form-meaning patterns.

papers-094**papers-094**

11:10–11:30

Detecting hypernymy relations is a key task in NLP, which is addressed in the literature using two complementary approaches. Distributional methods, whose supervised variants are the current best performers, and path-based methods, which received less research attention. We suggest an improved path-based algorithm, in which the dependency paths are encoded using a recurrent neural network, that achieves results comparable to distributional methods. We then extend the approach to integrate both path-based and distributional signals, significantly improving upon the state-of-the-art on this task.

short-261**short-261**

11:30–11:50

We propose a novel vector representation that integrates lexical contrast into distributional vectors and strengthens the most salient features for determining degrees of word similarity. The improved vectors significantly

outperform standard models and distinguish antonyms from synonyms with an average precision of 0.66—0.76 across word classes (adjectives, nouns, verbs). Moreover, we integrate the lexical contrast vectors into the objective function of a skip-gram model. The novel embedding outperforms state-of-the-art models on predicting word similarities in SimLex- 999, and on distinguishing antonyms from synonyms.

7B: Outstanding papers II

Plaza Ballroom F

Chair:

papers-232**papers-232**

10:10–10:30

We present an approach to improve statistical machine translation of image descriptions by multimodal pivots defined in visual space. The key idea is to perform image retrieval over a database of images that are captioned in the target language, and use the captions of the most similar images for crosslingual reranking of translation outputs. Our approach does not depend on the availability of large amounts of in-domain parallel data, but only relies on available large datasets of monolingually captioned images, and on state-of-the-art convolutional neural networks to compute image similarities. Our experimental evaluation shows improvements of 1 BLEU point over strong baselines.

papers-730**papers-730**

10:30–10:50

Combining deep neural networks with structured logic rules is desirable to harness flexibility and reduce uninterpretability of the neural models. We propose a general framework capable of enhancing various types of neural networks (e.g., CNNs and RNNs) with declarative first-order logic rules. Specifically, we develop an iterative distillation method that transfers the structured information of logic rules into the weights of neural networks. We deploy the framework on a CNN for sentiment analysis, and an RNN for named entity recognition. With a few highly intuitive rules, we obtain substantial improvements and achieve state-of-the-art or comparable results to previous best-performing systems.

papers-258**papers-258**

10:50–11:10

Linguistic drift is a process that produces slow irreversible changes in the grammar and function of a language's constructions. Importantly, changes in a part of a language can have trickle down effects, triggering changes elsewhere in that language. Although such causally triggered chains of changes have long been hypothesized by historical linguists, no explicit demonstration of the actual causality has been provided. In this study, we use co-occurrence statistics and machine learning to demonstrate that the functions of morphological cases experience a slow, irreversible drift along history, even in a language as conservative as is Icelandic. Crucially, we then move on to demonstrate –using the notion of Granger-causality– that there are explicit causal connections between the changes in the functions of the different cases, which are consistent with documented processes in the history of Icelandic. Our technique provides a means for the quantitative reconstruction of connected networks of subtle linguistic changes.

papers-142**papers-142**

11:10–11:30

The ability to compute an accurate reward functions is essential for optimising a dialogue policy via reinforcement learning. In real-world applications, using explicit user feedback as the reward signal is often unreliable and costly to collect. This problem can be mitigated if the user's intent is known in advance or data is available to pre-train a task success predictor off-line. In practice neither of these apply for most real world applications. Here we propose an on-line learning framework whereby the dialogue policy is jointly trained alongside the reward model via active learning with a Gaussian process model. This Gaussian process operates on a continuous space dialogue representation generated in an unsupervised fashion using a recurrent neural network encoder-decoder. The experimental results demonstrate that the proposed framework is able to significantly reduce data annotation costs and mitigate noisy user feedback in dialogue policy learning.

papers-587**papers-587**

11:30–11:50

We introduce a globally normalized transition-based neural network model that achieves state-of-the-art part-of-speech tagging, dependency parsing and sentence compression results. Our model is a simple feed-forward neural network that operates on a task-specific transition system, yet achieves comparable or better accuracies than recurrent models. We discuss the importance of global as opposed to local normalization: a key insight is that the label bias problem implies that globally normalized models can be strictly more expressive than locally normalized models.

Session 8 Overview – Wednesday, August 10, 2016

Track A <i>Question answering II (short papers)</i>	Track B <i>Word vectors III (short papers)</i>	Track C <i>Topics and discourse (short papers)</i>	Track D <i>Syntax and morphology (short papers)</i>	Track E <i>Potpourri II (short papers)</i>
Plaza Ballroom F	Plaza Ballroom F	Plaza Ballroom F	Plaza Ballroom F	Plaza Ballroom F
short-336 short-336	short-020 short-020	short-180 short-180	short-072 short-072	short-301 short-301
short-137 short-137	short-311 short-311	short-399 short-399	short-371 short-371	short-280 short-280
short-127 short-127	short-166 short-166	short-451 short-451	short-444 short-444	short-096 short-096
short-365 short-365	short-068 short-068	short-097 short-097	short-246 short-246	
short-012 short-012	short-243 short-243	short-422 short-422	short-413 short-413	

04:34 04:18 04:02 03:46 03:30

Parallel Session 8

8A: Question answering II (short papers)

Plaza Ballroom F

Chair:

short-336

short-336

15:30–15:46

We explore the applicability of machine translation evaluation (MTE) methods to a very different problem: answer ranking in community Question Answering. In particular, we adopt a pairwise neural network (NN) architecture, which incorporates MTE features, as well as rich syntactic and semantic embeddings, and which efficiently models complex non-linear interactions. The evaluation results show state-of-the-art performance, with sizeable contribution from both the MTE features and from the pairwise NN architecture.

short-137

short-137

15:46–16:02

We provide a solution for elementary science test using instructional materials. We posit that there is a hidden structure that explains the correctness of an answer given the question and instructional materials and present a unified max-margin framework that learns to find these hidden structures (given a corpus of question-answer pairs and instructional materials), and uses what it learns to answer novel elementary science questions. Our evaluation shows that our framework outperforms several strong baselines.

short-127

short-127

16:02–16:18

Prominent semantic annotations take an inclusive approach to argument span annotation, marking arguments as full constituency subtrees. Some works, however, showed that identifying a reduced argument span can be beneficial for various semantic tasks. While certain practical methods do extract reduced argument spans, such as in OpenIE, these solutions are often ad-hoc and system-dependent, with no commonly accepted standards. In this paper we propose a generic argument reduction criterion, along with an annotation procedure, and show that it can be consistently and intuitively annotated using the recent QA-SRL paradigm.

short-365

short-365

16:18–16:34

We present improvements to our incremental proposition-based summariser, which is inspired by Kintsch and van Dijk's (1978) text comprehension model. Argument overlap is a central concept in this summariser. Our new model replaces the old overlap method based on distributional similarity with one based on lexical chains. We evaluate on a new corpus of 124 summaries of educational texts, and show that our new system outperforms the old method and several state-of-the-art non-proposition-based summarisers. The experiment also verifies that the incremental nature of memory cycles is beneficial in itself, by comparing it to a non-incremental algorithm using the same underlying information.

short-012

short-012

16:34–16:50

Machine comprehension tests the system's ability to understand a piece of text through a reading comprehension task. For this task, we propose an approach using the Abstract Meaning Representation (AMR) formalism. We construct meaning representation graphs for the given text and for each question-answer pair by merging the AMRs of comprising sentences using cross-sentential phenomena such as coreference and rhetorical structures. Then, we reduce machine comprehension to a graph containment problem. We posit that there is a latent mapping of the question-answer meaning representation graph onto the text meaning representation graph that explains the answer. We present a unified max-margin framework that learns to find this mapping (given a corpus of texts and question-answer pairs), and uses what it learns to answer questions on novel texts. We show that this approach leads to state of the art results on the task.

8B: Word vectors III (short papers)

Plaza Ballroom F

Chair:

short-020

short-020

15:30–15:46

Cross-lingual word embeddings are used for cross-lingual information retrieval or domain adaptations. In this paper, we extend Eigenwords, spectral monolingual word embeddings based on canonical correlation analysis (CCA), to cross-lingual settings with sentence-alignment. For incorporating cross-lingual information, CCA is replaced with its generalization based on the spectral graph embeddings. The proposed method, which we refer to as Cross-Lingual Eigenwords (CL-Eigenwords), is fast and scalable for computing distributed representations of words via eigenvalue decomposition. Numerical experiments of English-Spanish word translation tasks show that CL-Eigenwords is competitive with state-of-the-art cross-lingual word embedding methods.

short-311

short-311

15:46–16:02

L2 learners often produce “ungrammatical” word combinations such as, e.g., *give a suggestion or *make a walk. This is because of the “collocationality” of one of their items (the base) that limits the acceptance of collocates to express a specific meaning ('perform' above). We propose an algorithm that delivers, for a given base and the intended meaning of a collocate, the actual collocate lexeme(s) (make / take above). The algorithm exploits the linear mapping between bases and collocates from examples and generates a collocation transformation matrix which is then applied to novel unseen cases. The evaluation shows a promising line of research in collocation discovery.

short-166

short-166

16:02–16:18

Incorporating lexical knowledge from semantic resources (e.g., WordNet) has been shown to improve the quality of distributed word representations. This knowledge often comes in the form of relational triplets ($x; r; y$) where words x and y are connected by a relation type r . Existing methods either ignore the relation types, essentially treating the word pairs as generic related words, or employ rather restrictive assumptions to model the relational knowledge. We propose a novel approach to model relational knowledge based on low-rank subspace regularization, and conduct experiments on standard tasks to evaluate its effectiveness.

short-068

short-068

16:18–16:34

We decompose a standard embedding space into interpretable orthogonal subspaces, including a one-dimensional polarity subspace as well as concreteness, frequency and part-of-speech (POS) subspaces. We introduce a new calculus for subspaces that supports operations like “ $-1 \text{ X hate} = \text{love}$ ” and “give me a neutral word for greasy” (i.e., oleaginous). This calculus extends analogy computations like “king - man + woman = queen”. For the tasks of Antonym Classification and POS Tagging our method outperforms the state of the art. We create new test sets for Morphological Analogies and for the task of Polarity Spectrum Creation.

short-243

short-243

16:34–16:50

Recent comparative studies have demonstrated the usefulness of dependency-based contexts (DEPS) for learning distributed word representations for similarity tasks. In English, DEPS tend to perform better than the more common, less informed bag-of-words contexts (BOW). In this paper, we present the first cross-linguistic comparison of different context types for three different languages. DEPS are extracted from “universal parses” without any language-specific optimization. Our results suggest that the universal DEPS (UDEPS) are useful for detecting functional similarity (e.g., verb similarity, solving syntactic analogies) among languages, but their advantage over BOW is not as prominent as previously reported on English. We also show that simple “post-parsing” filtering of useful UDEPS contexts leads to consistent improvements across languages.

8C: Topics and discourse (short papers)

Plaza Ballroom F

Chair:

short-180**short-180**

15:30–15:46

Computational Argumentation has two main goals - the detection and analysis of arguments on the one hand, and the synthesis of arguments on the other. Much attention has been given to the former, but considerably less to the latter. A key component in synthesizing arguments is the synthesis of claims. One way to do so is by employing argumentation mining to detect claims within an appropriate corpus. In general, this appears to be a hard problem. Thus, it is interesting to explore if - for the sake of synthesis - there may be other ways to generate claims. Here we explore such a method: we extract the predicate of simple, manually-detected, claims, and attempt to generate novel claims from them. Surprisingly, this simple method yields fairly good results.

short-399**short-399**

15:46–16:02

We propose a framework to model human comprehension of discourse connectives. Following the Bayesian pragmatic paradigm, we advocate that discourse connectives are interpreted based on a simulation of the production process by the speaker, who, in turn, considers the ease of interpretation for the listener when choosing connectives. Evaluation against the sense annotation of the Penn Discourse Treebank confirms the superiority of the model over literal comprehension. A further experiment demonstrates that the proposed model also improves automatic discourse parsing.

short-451**short-451**

16:02–16:18

Traditional topic models do not account for semantic regularities in language. Recent distributional representations of words exhibit semantic consistency over directional metrics such as cosine similarity. However, neither categorical nor Gaussian observational distributions used in existing topic models are appropriate to leverage such correlations. In this paper, we propose to use the von Mises-Fisher distribution to model the density of words over a unit sphere. Such a representation is well-suited for directional data. We use a Hierarchical Dirichlet Process for our base topic model and propose an efficient inference algorithm based on Stochastic Variational Inference. This model enables us to naturally exploit the semantic structures of word embeddings while flexibly discovering the number of topics. Experiments demonstrate that our method outperforms three competitive approaches in terms of topic coherence on two different text corpora while offering efficient inference.

short-097**short-097**

16:18–16:34

Big data presents new challenges for understanding large text corpora. Topic modeling algorithms help understand the underlying patterns, or “topics”, in data. Researchers author often read these topics in order to gain an understanding of the underlying corpus. It is important to evaluate the interpretability of these automatically generated topics. Methods have previously been designed to use crowdsourcing platforms to measure interpretability. In this paper, we demonstrate the necessity of a key concept, coherence, when assessing the topics and propose an effective method for its measurement. We show that the proposed measure of coherence captures a different aspect of the topics than existing measures. We further study the automation of these topic measures for scalability and reproducibility, showing that these measures can be automated.

short-422**short-422**

16:34–16:50

Scoring the quality of persuasive essays is an important goal of discourse analysis, addressed most recently with high-level persuasion-related features such as thesis clarity, or opinions and their targets. We investigate whether argumentation features derived from a coarse-grained argumentative structure of essays can help predict essay scores. We introduce a set of argumentation features related to argument components (e.g., the number of claims and premises), argument relations (e.g., the number of supported claims) and typology of argumentative structure (chains, trees). We show that these features are good predictors of human scores for TOEFL essays, both when the coarse-grained argumentative structure is manually annotated and automatically predicted.

8D: Syntax and morphology (short papers)

Plaza Ballroom F

*Chair:***short-072****short-072**

15:30–15:46

Morphological reinflection is the task of generating a target form given a source form, a source tag and a target tag. We propose a new way of modeling this task with neural encoder-decoder models. Our approach reduces the amount of required training data for this architecture and achieves state-of-the-art results, making encoder-decoder models applicable to morphological reinflection even for low-resource languages. We further present a new automatic correction method for the outputs based on edit trees.

short-371**short-371**

15:46–16:02

Most previous work on annotation projection has been limited to a subset of Indo-European languages, using only a single source language, and projecting annotation for one task at a time. In contrast, we present an Integer Linear Programming (ILP) algorithm that simultaneously projects annotation for multiple tasks from multiple source languages, relying on parallel corpora available for hundreds of languages. When training POS taggers and dependency parsers on jointly projected POS tags and syntactic dependencies using our algorithm, we obtain better performance than a standard approach on 20/23 languages using one parallel corpus; and 18/27 languages using another.

short-444**short-444**

16:02–16:18

Recently, many neural network models have been applied to Chinese word segmentation. However, such models focus more on collecting local information while long distance dependencies are not well learned. To integrate local features with long distance dependencies, we propose a dependency-based gated recursive neural network. Local features are first collected by bi-directional long short term memory network, then combined and refined to long distance dependencies via gated recursive neural network. Experimental results show that our model is a competitive model for Chinese word segmentation.

short-246**short-246**

16:18–16:34

Morphologically rich languages (MRL) are languages in which much of the structural information is contained at the word-level, leading to high level word-form variation. Historically, syntactic parsing has been mainly tackled using generative models. These models assume input features to be conditionally independent, making difficult to incorporate arbitrary features. In this paper, we investigate the greedy discriminative parser described in (Legrand and Collobert, 2015), which relies on word embeddings, in the context of MRL. We propose to learn morphological embeddings and propagate morphological information through the tree using a recursive composition procedure. Experiments show that such embeddings can dramatically improve the average performance on different languages. Moreover, it yields state-of-the art performance for a majority of languages.

short-413**short-413**

16:34–16:50

For many of the world’s languages, there are no or very few linguistically annotated resources. On the other hand, raw text, and often also dictionaries, can be harvested from the web for many of these languages, and part-of-speech taggers can be trained with these resources. At the same time, previous research shows that eye-tracking data, which can be obtained without explicit annotation, contains clues to part-of-speech information. In this work, we bring these two ideas together and show that given raw text, a dictionary, and eye-tracking data obtained from naive participants reading text, we can train a weakly supervised PoS tagger using a second-order HMM with maximum entropy emissions. The best model use type-level aggregates of eye-tracking data and significantly outperforms a baseline that does not have access to eye-tracking data.

8E: Potpourri II (short papers)

Plaza Ballroom F

Chair:

short-301

short-301

15:30–15:46

The aim of this paper is to investigate suitable evaluation strategies for the task of word-level quality estimation of machine translation. We suggest various metrics to replace $\$F_1\$$ -score for the “BAD” class, which is currently used as main metric. We compare the metrics’ performance on real system outputs and synthetically generated datasets and suggest a reliable alternative to the $\$F_1\$$ -BAD score — the multiplication of $\$F_1\$$ -scores for different classes. Other metrics have lower discriminative power and are biased by unfair labellings.

short-280

short-280

15:46–16:02

Medical sciences have long since established an ethics code for experiments, to minimize the risk of harm to subjects. Natural language processing (NLP) used to involve mostly anonymous corpora, with the goal of enriching linguistic analysis, and was therefore unlikely to raise ethical concerns. As NLP becomes increasingly wide-spread and uses more data from social media, however, the situation has changed: the outcome of NLP experiments and applications can now have a direct effect on individual users’ lives. Until now, the discourse on this topic in the field has not followed the technological development, while public discourse was often focused on exaggerated dangers. This position paper tries to take back the initiative and start a discussion. We identify a number of social implications of NLP and discuss their ethical significance, as well as ways to address them.

short-096

short-096

16:02–16:18

Increasing amounts of digital data in historical linguistics necessitate the development of automatic methods for the detection of cognate words across languages. Recently developed methods work well on language families with moderate time depths, but they are not capable of identifying cognate morphemes in words which are only partially related. Partial cognacy, however, is a frequently recurring phenomenon, especially in language families with productive derivational morphology. This paper presents a pilot approach for partial cognate detection in which networks are used to represent similarities between word parts and cognate morphemes are identified with help of state-of-the-art algorithms for network partitioning. The approach is tested on a newly created benchmark dataset with data from three sub-branches of Sino-Tibetan and yields very promising results, outperforming all algorithms which are not sensible to partial cognacy.

Message from the Program Committee Co-Chairs

*SEM, the Joint Conference on Lexical and Computational Semantics, has been organized yearly since 2012 under the auspices of ACL SIGLEX and SIGSEM. Its long term goal is to become a stable forum for the growing number of NLP researchers working on all aspects of semantics. To this end, each year it brings together researchers interested in the semantics of natural languages and its computational modeling. The conference embraces symbolic and probabilistic approaches, and everything in between. Theoretical contributions as well as practical applications are welcome.

The 2016 edition of *SEM takes place in Berlin on August 11 and 12 and is collocated with ACL. We accepted 27 papers (16 long and 11 short papers) for publication at the conference, out of 66 paper submissions (resulting in an overall acceptance rate of 40%). The selected papers cover a wide range of topics including formal and distributional semantics, lexical semantics, discourse semantics as well as more application-oriented themes.

The *SEM 2016 program consists of oral presentations for long papers, a poster session for short papers and three keynote talks by Yoav Artzi, Alexander Koller and Bonnie Webber.

Following the tradition initiated at *SEM 2015, *SEM 2016 will award two Adam Kilgarriff *SEM Best Paper Awards for Lexical Semantics.

We thank EACL and SIGLEX for sponsoring the three keynotes and Google and Lexical Computing for sponsoring the Adam Kilgarriff *SEM Best Paper Award. We would also like to thank Phong Le, *SEM 2016 Publication Chair, for his valuable work in editing these proceedings and the area chairs for their efforts in recruiting reviewers, stimulating discussion among them and for their dedication to carefully select the papers that make *SEM 2016 the high quality event we will all enjoy in Berlin. Last but not least, we thank the reviewers without whom *SEM could not be.

Claire Gardent, General Chair (CNRS and Université de Lorraine, Nancy, France) Raffaella Bernardi, Program Co-Chair (University of Trento, Italy) Ivan Titov, Program Co-Chair (University of Amsterdam, the Netherlands)

Claire Gardent, General Chair (CNRS and Université de Lorraine, Nancy, France)
Raffaella Bernardi, Program Co-Chair (University of Trento, Italy)

Ivan Titov, Program Co-Chair (University of Amsterdam, the Netherlands)
Sem 2016 conference co-chairs

*SEM 2016: The 5th Joint Conference on Lexical and Computational Semantics

Organizers: Claire Gardent, Raffaella Bernardi, Ivan Titov, Phong Le, Kevin Duh,
Diana McCarthy, Roberto Navigli, Jonathan Ginzburg, Yoav Artzi, Yonatan Bisk,
Annie Louis, Michael Roth and Elena Cabrio

Venue: Governor's Square 14

Thursday, August 11, 2016

9:00–9:10 Welcome

9:10–10:00 Invited Talk: Context and Non-compositional Phenomena in Language Understanding (Yoav Artzi)

10:00–10:30 starsem16-002
starsem16-002

10:30–11:00 Break

Lexical semantics

11:00–11:30 starsem16-004
starsem16-004
11:30–12:00 starsem16-008
starsem16-008
12:00–12:30 starsem16-063
starsem16-063

12:30–2:00 Lunch break

Semantic parsing and formal semantics

2:00–2:30 starsem16-051
starsem16-051
2:30–3:00 starsem16-065
starsem16-065
3:00–3:30 starsem16-064
starsem16-064

3:30–4:00 Break

Formal and linguistic semantics

4:00–4:30 starsem16-012
starsem16-012

4:30–5:30 Poster Session

- starsem16-036
- starsem16-036
- starsem16-041
- starsem16-041
- starsem16-044
- starsem16-044
- starsem16-045
- starsem16-045
- starsem16-003
- starsem16-003

- starsem16-035
starsem16-035
- starsem16-031
starsem16-031
- starsem16-033
starsem16-033
- starsem16-018
starsem16-018
- starsem16-026
starsem16-026
- starsem16-039
starsem16-039

Friday, August 12, 2016

9:10–10:00 **Invited talk: Top-down and bottom-up views on success in semantics**
(Alexander Koller)

10:00–10:30 starsem16-047
starsem16-047

10:30–11:00 **Break**

Semantics for applications and distributional semantics

11:00–11:30 starsem16-027
starsem16-027

11:30–12:00 starsem16-054
starsem16-054

12:00–12:30 starsem16-013
starsem16-013

12:30–2:10 **Lunch break**

2:10–3:00 **Invited talk: Exploring for Concurrent Discourse Relations** (Bonnie Webber)

3:00–3:30 starsem16-010
starsem16-010

3:30–4:00 **Break**

Distributional semantics

4:00–4:30 starsem16-060
starsem16-060

4:30–5:00 starsem16-049
starsem16-049

5:00–5:30 starsem16-034
starsem16-034

5:30–5:40 **Closing**

Message from the Program Committee Co-Chairs

The 2016 Conference on Computational Natural Language Learning is the twentieth in the series of annual meetings organized by SIGNLL, the ACL special interest group on natural language learning. CoNLL 2016 will be held on August 11-12, 2016, and is co-located with the 54th annual meeting of the Association for Computational Linguistics (ACL) in Berlin, Germany.

In order to accommodate papers with extended proofs and experimental material, CoNLL 2016 accepted only long papers, allowing 9 pages of content plus unlimited pages of references and supplementary material. We received 186 submissions in total, out of which 13 had to be rejected for formal reasons, and 21 were withdrawn by the authors. Of the remaining 149 papers, 30 papers were chosen to appear in the conference program, resulting in an overall acceptance rate of 20%. All accepted papers appear here in the proceedings.

As in previous years, CoNLL 2016 features a shared task, this year on Shallow Discourse Parsing. Papers accepted for the shared task are collected in a companion volume of CoNLL 2016.

To fit the paper presentations in a 2-day program, 21 long papers were selected for oral presentation, and 9 papers were presented as posters. The papers selected for oral presentation are distributed in six main sessions. Poster presenters were given the chance to present their poster in a short oral spotlight presentation.

For the first time, CoNLL 2016 announced a special topic on statistical natural language learning beyond linear models and convex optimization. The special topic was embraced by several authors and is reflected by the invited talks given by Jürgen Schmidhuber and Fernanda Ferreira.

We would like to thank all the authors who submitted their work to CoNLL 2016, as well as the program committee for helping us select the best papers out of many high-quality submissions. We are also grateful to our invited speakers, who graciously agreed to give talks at CoNLL.

Special thanks are due to the SIGNLL board members, Xavier Carreras and Julia Hockenmaier, for their valuable advice and assistance in putting together this year's program, and to Ben Verhoeven, for maintaining the CoNLL 2016 web page. We are grateful to the ACL organization for helping us with the program, proceedings and logistics. Finally, our gratitude goes to our sponsor, Google Inc., for supporting the best paper award at CoNLL 2016.

We hope you enjoy the conference!

Yoav Goldberg and Stefan Riezler
CoNLL 2016 conference co-chairs

CoNLL 2016: The SIGNLL Conference on Computational Natural Language Learning

Organizers: **Yoav Goldberg, Stefan Riezler**

Venue: dummy

Thursday, August 11, 2016

9:00–9:10 **Opening**

9:10–10:30 **Session 1: Word-Level Semantics**

- CoNLL2016-014
CoNLL2016-014
- CoNLL2016-025
CoNLL2016-025
- CoNLL2016-029
CoNLL2016-029
- CoNLL2016-092
CoNLL2016-092

10:30–11:00 **Break**

11:00–12:20 **Session 2: Sentence-Level Semantics**

- CoNLL2016-078
CoNLL2016-078
- CoNLL2016-145
CoNLL2016-145
- CoNLL2016-148
CoNLL2016-148
- CoNLL2016-170
CoNLL2016-170

12:20–12:50 **SIGNLL Business Meeting**

12:50–2:00 **Lunch Break**

2:00–3:40 **Session 3: Human Language Processing**

- CoNLL2016-055
CoNLL2016-055
- CoNLL2016-126
CoNLL2016-126
- CoNLL2016-163
CoNLL2016-163
- CoNLL2016-171
CoNLL2016-171
- CoNLL2016-185
CoNLL2016-185

3:40–4:00 **Break**

4:00–5:15 **Keynote 1: Human Processing of Disfluent Speech: Basic Findings, Theoretical Approaches, and Implications for Natural Language Processing (Fernanda Ferreira)**

5:15–6:15 **Session 4: Sarcasm / Sentiment**

- CoNLL2016-008
CoNLL2016-008
- CoNLL2016-124
CoNLL2016-124
- CoNLL2016-186
CoNLL2016-186

Friday, August 12, 2016

9:00–9:15 **SIGNLL Steering Committee Address**

9:15–10:30 **Keynote 2: RNNaissance (Jürgen Schmidhuber)**

10:30–11:00 **Break**

11:00–12:40 **Session 5: Syntax, Named Entities, and Machine Translation**

- CoNLL2016-012
CoNLL2016-012
- CoNLL2016-066
CoNLL2016-066
- CoNLL2016-105
CoNLL2016-105
- CoNLL2016-142
CoNLL2016-142
- CoNLL2016-166
CoNLL2016-166

12:40–2:00 **Lunch Break**

2:00–3:30 **Session 6: Shared Task on Shallow Discourse Parsing**

3:30–4:00 **Break**

4:00–4:20 **Session 7: Spotlight Presentations**

4:20–5:20 **Session 8: Posters**

- CoNLL2016-011
CoNLL2016-011
- CoNLL2016-013
CoNLL2016-013
- CoNLL2016-017
CoNLL2016-017
- CoNLL2016-021
CoNLL2016-021
- CoNLL2016-036
CoNLL2016-036
- CoNLL2016-076
CoNLL2016-076
- CoNLL2016-091
CoNLL2016-091
- CoNLL2016-128
CoNLL2016-128

- CoNLL2016-129
CoNLL2016-129

5:20–6:00 **Session 9: Best Paper Awards + Closing**

8

Workshops

Thursday–Friday

Governor's Square 17	First Conference on Machine Translation (WMT)	p.136
----------------------	-----------------------------------------------	-------

Thursday

Governor's Square 12	1st Workshop on Representation Learning for NLP (Repl4NLP)	p.142
Director's Row E	SIGANN's Linguistic Annotation Workshop (LAW X 2016)	p.144
Director's Row I	12th Workshop on Multiword Expressions (MWE 2016)	p.146
Tower Court A	7th Workshop on Cognitive Aspects of Computational Language Learning	p.148
Director's Row H	14th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology	p.150
Director's Row J	10th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH 2016)	p.152

Friday

Governors Square 11	5th Workshop on Vision and Language (VL'16)	p.154
Governors Square 17	15th Workshop on Biomedical Natural Language Processing	p.155
Director's Row E	SIGFSM Workshop on Statistical NLP and Weighted Automata	p.157
Director's Row I	1st Workshop on Evaluating Vector-Space Representations for NLP	p.158
Director's Row J	10th Web as Corpus Workshop	p.160
Governor's Square 11	6th NEWS Named Entities Workshop	p.161
Director's Row H	3rd Workshop on Argument Mining	p.162

Workshop 1: 1st Conference on Machine Translation

Organizers: **workshop1**
Venue: Governor's Square 17

Thursday, August 11, 2016

8:45–9:00 **Opening Remarks**

Session 1: Shared Tasks Overview Presentations I

9:00–9:20 **Shared Task: News Translation**

- WMT16-134
- WMT16-134

9:20–9:35 **Shared Task: IT-Domain Translation**

9:35–9:50 **Shared Task: Biomedical Translation**

9:50–10:10 **Shared Task: Metrics**

- WMT16-135
- WMT16-135

10:10–10:30 **Shared Task: Tuning**

- WMT16-136
- WMT16-136

10:30–11:00 **Coffee Break**

Session 2: Shared Tasks Poster Session I

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

- WMT16-110
- WMT16-110
- WMT16-118
- WMT16-118
- WMT16-123
- WMT16-123
- WMT16-117
- WMT16-117
- WMT16-141
- WMT16-141
- WMT16-130
- WMT16-130
- WMT16-127
- WMT16-127
- WMT16-108
- WMT16-108
- WMT16-082
- WMT16-082
- WMT16-063
- WMT16-063
- WMT16-097
- WMT16-097

- WMT16-058
WMT16-058
- WMT16-109
WMT16-109
- WMT16-128
WMT16-128
- WMT16-065
WMT16-065
- WMT16-139
WMT16-139
- WMT16-059
WMT16-059
- WMT16-064
WMT16-064
- WMT16-062
WMT16-062
- WMT16-056
WMT16-056
- WMT16-088
WMT16-088
- WMT16-122
WMT16-122
- WMT16-092
WMT16-092
- WMT16-120
WMT16-120
- WMT16-138
WMT16-138

11:00–12:30 **Shared Task: IT-domain Translation**

- WMT16-003
WMT16-003
- WMT16-078
WMT16-078
- WMT16-090
WMT16-090
- WMT16-084
WMT16-084
- WMT16-103
WMT16-103
- WMT16-087
WMT16-087

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

- WMT16-075
WMT16-075
- WMT16-100
WMT16-100
- WMT16-094
WMT16-094
- WMT16-055
WMT16-055

11:00–12:30 **Shared Task: Metrics**

- WMT16-107
WMT16-107
- WMT16-067
WMT16-067

- WMT16-004
WMT16-004
- WMT16-054
WMT16-054
- WMT16-079
WMT16-079

11:00–12:30 **Shared Task: Tuning**

- WMT16-074
WMT16-074

12:30–2:00 **Lunch**

Session 3: Invited Talk

2:00–3:30 **Spence Green (Lilt): Interactive Machine Translation: From Research to Practice**

3:30–4:00 **Coffee Break**

Session 4: Research Papers on Linguistic Modelling

- 4:00–4:20 WMT16-011
WMT16-011
- 4:20–4:40 WMT16-013
WMT16-013
- 4:40–5:00 WMT16-018
WMT16-018
- 5:00–5:20 WMT16-041
WMT16-041
- 5:20–5:40 WMT16-037
WMT16-037

Friday, August 12, 2016

Session 5: Shared Tasks Overview Presentations II

9:00–9:20 **Shared Task: Cross-Lingual Pronoun Prediction**

- WMT16-131
WMT16-131

9:20–9:40 **Shared Task: Multimodal Machine Translation and Cross-Lingual Image Description**

- WMT16-132
WMT16-132

9:40–9:55 **Shared Task: Bilingual Document Alignment**

- WMT16-133
WMT16-133

9:55–10:10 **Shared Task: Automatic Post-Editing**

10:10–10:30 **Shared Task: Quality Estimation**

10:30–11:00 **Coffee Break**

Session 6: Shared Tasks Poster Session II

11:00–12:30 **Shared Task: Cross-Lingual Pronoun Prediction**

- WMT16-111
- WMT16-111
- WMT16-081
- WMT16-081
- WMT16-068
- WMT16-068
- WMT16-070
- WMT16-070
- WMT16-071
- WMT16-071
- WMT16-089
- WMT16-089
- WMT16-076
- WMT16-076
- WMT16-066
- WMT16-066
- WMT16-102
- WMT16-102
- WMT16-083
- WMT16-083

11:00–12:30 **Shared Task: Multimodal Machine Translation and Cross-Lingual Image Description**

- WMT16-099
- WMT16-099
- WMT16-129
- WMT16-129
- WMT16-115
- WMT16-115
- WMT16-121
- WMT16-121
- WMT16-126
- WMT16-126
- WMT16-114
- WMT16-114

11:00–12:30 **Shared Task: Bilingual Document Alignment**

- WMT16-137
- WMT16-137
- WMT16-112
- WMT16-112
- WMT16-140
- WMT16-140
- WMT16-060
- WMT16-060
- WMT16-116
- WMT16-116
- WMT16-091
- WMT16-091
- WMT16-008
- WMT16-008
- WMT16-048
- WMT16-048

- WMT16-031
WMT16-031
- WMT16-042
WMT16-042
- WMT16-093
WMT16-093
- WMT16-113
WMT16-113

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

- WMT16-069
WMT16-069
- WMT16-105
WMT16-105
- WMT16-072
WMT16-072

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

- WMT16-106
WMT16-106
- WMT16-104
WMT16-104
- WMT16-124
WMT16-124
- WMT16-061
WMT16-061
- WMT16-073
WMT16-073
- WMT16-096
WMT16-096
- WMT16-086
WMT16-086
- WMT16-085
WMT16-085
- WMT16-095
WMT16-095
- WMT16-057
WMT16-057
- WMT16-077
WMT16-077
- WMT16-101
WMT16-101
- WMT16-098
WMT16-098
- WMT16-125
WMT16-125

12:30–2:00 **Lunch**

Session 7: Research Papers on Neural Translation

- 2:00–2:20 WMT16-043
WMT16-043
- 2:20–2:40 WMT16-025
WMT16-025
- 2:40–3:00 WMT16-027
WMT16-027
- 3:00–3:20 WMT16-036
WMT16-036

3:30–4:00 **Coffee Break**

Session 8: Research Papers on Translation Models

- | | |
|-----------|------------------|
| 4:00–4:20 | WMT16-005 |
| | WMT16-005 |
| 4:00–4:40 | WMT16-038 |
| | WMT16-038 |
| 4:00–5:00 | WMT16-047 |
| | WMT16-047 |
| 5:00–5:20 | WMT16-046 |
| | WMT16-046 |

Workshop 2: 1st Workshop on Representation Learning for NLP

Organizers: **workshop2**
Venue: Governor's Square 12

Thursday, August 11, 2016

9:30–9:40 **Welcome and Opening Remarks**

9:40–10:30 **Keynote: Katrin Erk (University of Texas at Austin)**

10:30–11:00 **Coffee Break**

11:00–11:50 **Keynote: Animashree Anandkumar (University of California, Irvine)**

11:50–12:10 **Best Papers**

12:10–1:30 **Lunch Break**

1:30–2:20 **Keynote: Hal Daumé III (University of Maryland)**

2:20–3:10 **Keynote: Raia Hadsell (DeepMind)**

3:10–3:30 **Poster Session**

- repl4nlp2016-033
repl4nlp2016-033
- repl4nlp2016-044
repl4nlp2016-044
- repl4nlp2016-037
repl4nlp2016-037
- repl4nlp2016-040
repl4nlp2016-040
- repl4nlp2016-011
repl4nlp2016-011
- repl4nlp2016-018
repl4nlp2016-018
- repl4nlp2016-048
repl4nlp2016-048
- repl4nlp2016-051
repl4nlp2016-051
- repl4nlp2016-008
repl4nlp2016-008
- repl4nlp2016-017
repl4nlp2016-017
- repl4nlp2016-042
repl4nlp2016-042
- repl4nlp2016-035
repl4nlp2016-035
- repl4nlp2016-023
repl4nlp2016-023
- repl4nlp2016-045
repl4nlp2016-045

- repl4nlp2016-029
repl4nlp2016-029
- repl4nlp2016-025
repl4nlp2016-025
- repl4nlp2016-019
repl4nlp2016-019
- repl4nlp2016-043
repl4nlp2016-043
- repl4nlp2016-010
repl4nlp2016-010
- repl4nlp2016-014
repl4nlp2016-014
- repl4nlp2016-005
repl4nlp2016-005
- repl4nlp2016-020
repl4nlp2016-020
- repl4nlp2016-041
repl4nlp2016-041
- repl4nlp2016-046
repl4nlp2016-046
- repl4nlp2016-052
repl4nlp2016-052
- repl4nlp2016-028
repl4nlp2016-028
- repl4nlp2016-027
repl4nlp2016-027
- repl4nlp2016-038
repl4nlp2016-038
- repl4nlp2016-013
repl4nlp2016-013
- repl4nlp2016-021
repl4nlp2016-021

3:30–4:00 **Poster Session Continues and Coffee break**

4:00–5:20 **Panel Discussion**

5:20–5:30 **Closing Remarks**

Workshop 3: 10th Linguistic Annotation Workshop

Organizers: **workshop3**
Venue: Director's Row E

Thursday, August 11, 2016

Session 1: Opening and Invited Talk

9:00–9:10 **Opening Remarks**

9:10–10:05 **Invited talk: Assessing the Consistency and Use of “Common Sense” and Dependency Annotations (Marie-Catherine de Marneffe)**

10:05–10:30 law-x-2016-044
law-x-2016-044

10:30–11:00 **Coffee Break**

Session 2: Dependency Annotation and Discourse

11:00–11:25 law-x-2016-037
law-x-2016-037
11:25–11:50 law-x-2016-040
law-x-2016-040
11:50–12:15 law-x-2016-026
law-x-2016-026
12:15–12:40 law-x-2016-046
law-x-2016-046

12:40–2:00 **Lunch Break**

Session 3: Evaluation of Agreement (Special Theme)

2:00–2:25 law-x-2016-031
law-x-2016-031
2:25–2:50 law-x-2016-016
law-x-2016-016

2:50–3:05 **Poster boasters**

3:05–4:00 **Session 4: Poster Presentations**

- law-x-2016-004
law-x-2016-004
- law-x-2016-005
law-x-2016-005
- law-x-2016-008
law-x-2016-008
- law-x-2016-009
law-x-2016-009
- law-x-2016-011
law-x-2016-011
- law-x-2016-012
law-x-2016-012
- law-x-2016-016
law-x-2016-016
- law-x-2016-022
law-x-2016-022

- law-x-2016-027
law-x-2016-027
- law-x-2016-031
law-x-2016-031
- law-x-2016-034
law-x-2016-034
- law-x-2016-035
law-x-2016-035
- law-x-2016-036
law-x-2016-036
- law-x-2016-043
law-x-2016-043
- law-x-2016-044
law-x-2016-044
- law-x-2016-045
law-x-2016-045
- law-x-2016-048
law-x-2016-048

3:30–4:00 **Poster presentations and coffee**

4:00–4:30 **Invited tutorial: How Item-Response Models Can Help us Take the Headache out of Annotation Projects (Dirk Hovy)**

4:30–5:15 **Discussion**

5:15–5:30 **Closing remarks: 10 years of LAW**

Workshop 4: 12th Workshop on Multiword Expressions

Organizers: **workshop4**
Venue: Director's Row I

Thursday, August 11, 2016

8:50–9:00 **Opening remarks**

Oral Session 1

9:00–9:30 **mwe-018**
mwe-018
9:30–10:00 **mwe-029**
mwe-029
10:00–10:30 **mwe-039**
mwe-039

10:30–11:00 **Coffee Break**

Oral Session 2

11:00–11:20 **mwe-047**
mwe-047
11:20–11:40 **mwe-006**
mwe-006
11:40–12:00 **mwe-037**
mwe-037

12:00–1:00 **Invited Talk**

1:00–2:00 **Lunch**

2:00–2:40 **Poster Booster Session (5 minutes per poster)**

- **mwe-011**
mwe-011
- **mwe-002**
mwe-002
- **mwe-031**
mwe-031
- **mwe-041**
mwe-041
- **mwe-013**
mwe-013
- **mwe-025**
mwe-025
- **mwe-014**
mwe-014
- **mwe-012**
mwe-012

2:40–3:30 **Poster Session**

3:30–4:00 **Coffee Break**

Oral Session 3

4:00–4:30 **mwe-030**
mwe-030

4:30–4:50 **mwe-019**
 mwe-019
4:50–5:10 **mwe-044**
 mwe-044
5:10–5:20 **Closing Remarks**

Workshop 5: 7th Workshop on Cognitive Aspects of Computational Language Learning

Organizers: **workshop5**
Venue: Tower Court A

Thursday, August 11, 2016

9:00–9:05 **Welcome and Opening Session**

Session 1 - Language in Clinical Conditions

9:05–9:35 **CogACLL-002**
CogACLL-002

9:35–10:30 **Invited Speaker I**

10:30–11:00 **Coffee Break**

11:00–11:30 **Poster Session**

- **CogACLL-007**
CogACLL-007
- **CogACLL-011**
CogACLL-011
- **CogACLL-019**
CogACLL-019
- **CogACLL-016**
CogACLL-016
- **CogACLL-018**
CogACLL-018

Session 2: Child Directed Language

11:30–12:00 **CogACLL-004**
CogACLL-004

12:00–12:30 **CogACLL-003**
CogACLL-003

12:30–2:00 **Lunch Break**

2:00–3:00 **Invited Talk II**

Session 3: Learning Artificial Languages

3:00–3:30 **CogACLL-010**
CogACLL-010

3:30–4:00 **Coffee Break**

Session 4: Language Acquisition

4:00–4:30 **CogACLL-013**
CogACLL-013

4:30–5:00 **CogACLL-006**
CogACLL-006

Panel and Closing Session

5:00–5:30 **Panel**

5:30–5:35 **Closing Session**

Workshop 6: 14th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology

Organizers: **workshop6**
Venue: Director's Row H

Thursday, August 11, 2016

Phonetics

9:00–9:30 **sigmorphon16-014**
sigmorphon16-014

9:30–10:30 **Invited Talk (Colin Wilson)**

10:30–11:00 **Coffee Break**

11:00–11:30 **sigmorphon16-035**
sigmorphon16-035

11:30–12:30 **Shared task poster session**

- **sigmorphon16-025**
sigmorphon16-025
- **sigmorphon16-026**
sigmorphon16-026
- **sigmorphon16-028**
sigmorphon16-028
- **sigmorphon16-029**
sigmorphon16-029
- **sigmorphon16-030**
sigmorphon16-030
- **sigmorphon16-031**
sigmorphon16-031
- **sigmorphon16-032**
sigmorphon16-032
- **sigmorphon16-033**
sigmorphon16-033
- **sigmorphon16-034**
sigmorphon16-034

12:00–2:00 **Lunch Break**

2:00–3:00 **Workshop poster session**

- **sigmorphon16-001**
sigmorphon16-001
- **sigmorphon16-002**
sigmorphon16-002
- **sigmorphon16-003**
sigmorphon16-003
- **sigmorphon16-004**
sigmorphon16-004

- **sigmorphon16-009**
sigmorphon16-009
- **sigmorphon16-018**
sigmorphon16-018

Phonology

- 3:00–3:30 **sigmorphon16-008**
sigmorphon16-008

3:30–4:00 **Coffee Break**

Morphology

- 4:00–4:30 **sigmorphon16-010**
sigmorphon16-010
- 4:30–5:00 **sigmorphon16-012**
sigmorphon16-012
- 5:00–5:30 **sigmorphon16-007**
sigmorphon16-007

Workshop 7: 10th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities

Organizers: **workshop7**
Venue: Director's Row J

Thursday, August 11, 2016

Session 1

9:00–9:30 latech-042
latech-042
9:30–10:00 latech-051
latech-051
10:00–10:30 latech-032
latech-032

Session 2

11:00–11:30 latech-015
latech-015
11:30–12:00 latech-022
latech-022
12:00–12:30 latech-030
latech-030

1:30–2:00 **SIGHUM Business Meeting**

Session 3

2:00–2:30 latech-034
latech-034
2:30–3:00 latech-035
latech-035

3:00–4:00 **Poster Session**

- latech-003
- latech-003
- latech-006
- latech-006
- latech-011
- latech-011
- latech-016
- latech-016
- latech-018
- latech-018
- latech-020
- latech-020
- latech-021
- latech-021
- latech-024
- latech-024
- latech-039
- latech-039
- latech-050
- latech-050

Session 4

4:00–4:30 latech-008
latech-008
4:30–5:00 latech-012
latech-012
5:00–5:30 latech-017
latech-017

Workshop 8: 5th Workshop on Vision and Language

Organizers: **workshop8**
Venue: Governors Square 17

Friday, August 12, 2016

9:00–9:30 **Opening Remarks**

9:30–10:00 **vl16-005**

vl16-005

10:00–10:30 **vl16-007**

vl16-007

10:30–11:00 **Coffee Break**

11:00–12:00 **Invited talk: Yejin Choi**

12:00–12:30 **vl16-015**

vl16-015

12:30–2:00 **Lunch Break**

2:00–2:30 **vl16-011**

vl16-011

2:30–3:30 **Quick-fire presentations for posters (5mins each)**

3:30–4:00 **Coffee Break**

4:00–5:30 **Poster Session**

• **vl16-001**

vl16-001

• **vl16-006**

vl16-006

• **vl16-008**

vl16-008

• **vl16-010**

vl16-010

• **vl16-017**

vl16-017

• **vl16-002**

vl16-002

• **vl16-009**

vl16-009

• **vl16-013**

vl16-013

• **vl16-016**

vl16-016

Workshop 9: 15th Workshop on Biomedical Natural Language Processing

Organizers: **workshop9**

Venue: Director's Row E

Friday, August 12, 2016

8:30–8:40 **Opening remarks**

Session 1: Entity extraction and representation

8:40–9:00 **BioNLP16-023**
BioNLP16-023
9:00–9:20 **BioNLP16-013**
BioNLP16-013
9:20–9:40 **BioNLP16-003**
BioNLP16-003
9:40–10:00 **BioNLP16-007**
BioNLP16-007

10:00–10:30 **Invited Talk: “The BioNLP-ST challenges on information extraction and knowledge acquisition in biology” – Robert Bossy and Jin-Dong Kim**

10:30–11:00 **Coffee Break**

Session 2: Event and Relation Extraction

11:00–11:20 **BioNLP16-025**
BioNLP16-025
11:40–12:00 **BioNLP16-011**
BioNLP16-011
12:00–12:20 **BioNLP16-022**
BioNLP16-022
12:20–12:40 **BioNLP16-024**
BioNLP16-024

12:40–2:00 **Lunch break**

Session 3: Disambiguation, Classification, and more

2:00–2:20 **BioNLP16-021**
BioNLP16-021
2:20–2:40 **BioNLP16-010**
BioNLP16-010
2:40–3:00 **BioNLP16-005**
BioNLP16-005
3:00–3:20 **BioNLP16-030**
BioNLP16-030
3:20–3:40 **BioNLP16-035**
BioNLP16-035

3:40–4:00 **Coffee Break**

4:00–4:30 **Invited Talk: “BioASQ: A challenge on large-scale biomedical semantic indexing and question answering” – Anastasia Krithara**

4:30–5:30 Poster Session

- BioNLP16-008
BioNLP16-008
- BioNLP16-001
BioNLP16-001
- BioNLP16-014
BioNLP16-014
- BioNLP16-016
BioNLP16-016
- BioNLP16-020
BioNLP16-020
- BioNLP16-040
BioNLP16-040
- BioNLP16-026
BioNLP16-026
- BioNLP16-041
BioNLP16-041
- BioNLP16-042
BioNLP16-042
- BioNLP16-038
BioNLP16-038
- BioNLP16-039
BioNLP16-039
- BioNLP16-027
BioNLP16-027
- BioNLP16-002
BioNLP16-002
- BioNLP16-004
BioNLP16-004
- BioNLP16-028
BioNLP16-028

Workshop 10: SIGFSM Workshop on Statistical NLP and Weighted Automata

Organizers: **workshop10**

Venue: Director's Row I

Friday, August 12, 2016

Welcome and keynote

9:00–9:30 **Opening (Organizers)**

9:30–10:30 **Probabilistic Models of Related Strings (Jason Eisner)**

10:30–11:00 **Morning break**

Weighted tree automata

11:00–11:30 **StatFSM-004**

StatFSM-004

11:30–12:00 **StatFSM-008**

StatFSM-008

12:00–12:30 **StatFSM-010**

StatFSM-010

12:30–2:00 **Lunch break**

Weighted finite-state transducers

2:00–2:30 **StatFSM-005**

StatFSM-005

2:30–3:00 **StatFSM-013**

StatFSM-013

3:00–3:30 **StatFSM-003**

StatFSM-003

3:30–4:00 **Afternoon break**

Various weighted automata

4:00–4:30 **StatFSM-002**

StatFSM-002

4:30–5:00 **StatFSM-011**

StatFSM-011

5:00–5:30 **StatFSM-012**

StatFSM-012

Workshop 11: 1st Workshop on Evaluating Vector-Space Representations for NLP

Organizers: **workshop11**

Venue: Director's Row J

Friday, August 12, 2016

9:00–9:15 **Opening Remarks**

9:15–10:00 **Analysis Track**

- repeval16-026
repeval16-026
- repeval16-038
repeval16-038
- repeval16-022
repeval16-022

10:00–10:20 **Proposal Track 1**

- repeval16-036
repeval16-036
- repeval16-037
repeval16-037

10:20–10:45 **Coffee Break**

10:45–12:30 **Poster Session**

10:45–11:00 **Lightning Talks**

- repeval16-011
repeval16-011
- repeval16-016
repeval16-016
- repeval16-005
repeval16-005
- repeval16-009
repeval16-009
- repeval16-017
repeval16-017
- repeval16-021
repeval16-021
- repeval16-023
repeval16-023
- repeval16-024
repeval16-024
- repeval16-025
repeval16-025
- repeval16-028
repeval16-028
- repeval16-031
repeval16-031

- repeval16-033
repeval16-033
- repeval16-035
repeval16-035

12:30–2:00 **Lunch Break**

2:00–3:30 **Proposal Track 2: Word Representations**

- repeval16-014
repeval16-014
- repeval16-013
repeval16-013
- repeval16-006
repeval16-006
- repeval16-039
repeval16-039

3:30–4:00 **Coffee Break**

4:00–5:30 **Proposal Track 3: Contextualized Representations**

- repeval16-019
repeval16-019
- repeval16-032
repeval16-032
- repeval16-027
repeval16-027
- repeval16-007
repeval16-007

5:30–6:15 **Open Discussion**

6:15–6:30 **Best Proposal Awards**

Workshop 12: 10th Web as Corpus Workshop

Organizers: **workshop12**
Venue: Governor's Square 11

9:30–9:40 **Welcome and Introduction**

WAC-X Morning Session

9:40–10:00 **WAC-X-012**

WAC-X-012

10:00–10:30 **WAC-X-010**

WAC-X-010

10:30–11:00 **Coffee Break**

WAC-X Noon Session

11:00–11:30 **WAC-X-017**

WAC-X-017

11:30–12:00 **WAC-X-020**

WAC-X-020

12:00–12:30 **WAC-X-014**

WAC-X-014

12:30–1:30 **Lunch Break**

EmpiriST Session

1:30–1:50 **WAC-X-022**

WAC-X-022

1:50–2:10 **WAC-X-004**

WAC-X-004

2:10–2:30 **WAC-X-002**

WAC-X-002

WAC-X and EmpiriST Teaser Talks

2:30–2:35 **WAC-X-003**

WAC-X-003

2:35–2:40 **WAC-X-011**

WAC-X-011

2:40–2:45 **WAC-X-013**

WAC-X-013

2:45–2:50 **WAC-X-021**

WAC-X-021

2:55–3:00 **WAC-X-015**

WAC-X-015

3:00–3:05 **WAC-X-016**

WAC-X-016

3:05–3:10 **WAC-X-005**

WAC-X-005

3:10–4:30 **WAC-X and EmpiriST Poster Session**

Discussion Session

4:30–5:30 **WAC-X and EmpiriST Closing Discussion**

5:30–6:30 **Panel Discussion: Corpora, open science, and copyright reforms**

Workshop 13: 6th NEWS Named Entities Workshop

Organizers: **workshop13**

Venue: Director's Row H

Research Papers I

9:10–9:40 NEWS2016-008

NEWS2016-008

9:40–10:10 NEWS2016-013

NEWS2016-013

10:10–10:30 NEWS2016-014

NEWS2016-014

10:30–11:00 **Coffee Break**

Research Papers II

11:00–11:30 NEWS2016-006

NEWS2016-006

11:30–12:00 NEWS2016-010

NEWS2016-010

12:00–12:30 NEWS2016-011

NEWS2016-011

12:30–2:30 **Lunch Break**

2:30–3:30 **Keynote Speech**

2:30–3:30 NEWS2016-019

NEWS2016-019

3:30–4:00 **Coffee Break**

Shared Task on Name Entity Transliteration

4:00–4:10 NEWS2016-020

NEWS2016-020

4:00–4:10 NEWS2016-016

NEWS2016-016

4:10–4:30 NEWS2016-004

NEWS2016-004

4:30–4:50 NEWS2016-005

NEWS2016-005

4:50–5:10 NEWS2016-007

NEWS2016-007

5:10–5:20 NEWS2016-003

NEWS2016-003

5:20–5:30 **Closing Remarks**

Workshop 14: 3rd Workshop on Argument Mining

Organizers: **workshop14**
Venue: Director's Row H

Friday, August 12, 2016

9:00–9:10 **Welcome**

Session I

9:10–9:30 argmining2016-005
argmining2016-005
9:30–9:50 argmining2016-027
argmining2016-027
9:50–10:10 argmining2016-023
argmining2016-023
10:10–10:30 argmining2016-011
argmining2016-011

10:30–11:00 **Coffee break**

Session II

11:00–11:20 argmining2016-020
argmining2016-020
11:20–11:40 argmining2016-025
argmining2016-025
11:40–12:00 argmining2016-022
argmining2016-022
12:00–12:15 argmining2016-007
argmining2016-007
12:15–12:30 argmining2016-002
argmining2016-002

12:30–2:00 **Lunch**

Session III

2:00–2:20 argmining2016-019
argmining2016-019
2:20–2:40 argmining2016-018
argmining2016-018
2:40–3:00 argmining2016-029
argmining2016-029
3:00–3:15 argmining2016-001
argmining2016-001
3:15–3:30 argmining2016-012
argmining2016-012

3:30–4:00 **Coffee break**

Session IV - Debating Technologies and the Unshared Task Panel

4:00–4:20 argmining2016-031
argmining2016-031

- 4:20–4:40 **argmining2016-010**
 argmining2016-010
4:40–5:00 **argmining2016-006**
 argmining2016-006
- 5:00–5:25 **Panel Discussion: Unshared Task**
- **argmining2016-004**
 argmining2016-004
 - **argmining2016-009**
 argmining2016-009
 - **argmining2016-014**
 argmining2016-014
- 5:25–5:30 **Closing Remarks**
- 5:30–5:45 **Close**

Local Guide

This year's Association for Computational Linguistics annual meeting (ACL 2016) takes place in Berlin, Germany. Here are some things that we think might be useful or enjoyable for conference participants.

Transportation

The city is served by two airports, Berlin Tegel and Berlin Schönefeld, linking it to major cities all over the world. The city is also one of the major nodes in Europe's rail and road networks. Thus it is an excellent starting point for exploring the rest of Germany or its neighboring countries. Finally, Berlin has an excellent and extensive public transportation system which operates 24 hours, 7 days a week.

Airports

Berlin is served by two airports, providing multiple flights a day to a huge number of European, North African and Middle Eastern destinations. There are also daily long-haul flights to a large number of destinations outside Europe. Most airlines fly from the Berlin-Tegel airport, situated 8km northwest of the city centre. Berlin-Schönefeld, on the other hand, is mainly used by charter and low-cost airlines such as EasyJet, Ryanair, Norwegian Air Shuttle, and Condor. This gives participants from Europe, North Africa, and Turkey the opportunity to fly directly to Berlin at very low prices. However, some legacy airlines such as Aeroflot, El Al, and TAP Portugal also use the airport. The airport is situated 18 km southeast of the city center.

Transport from airports to city

From Berlin-Tegel: There is a direct bus line TXL which connects the airport with the main train station (Hauptbahnhof). From Hauptbahnhof, there are three urban rail lines S5, S7 and S75 to the station "Friedrichstrasse" which is about 5-10 minutes walk from the conference venue. The trains depart from platform 15 at the top level of the train station. The whole trip from the airport takes between 30 and 40 minutes and costs 2,70€. There are also bus lines which connect the airport to other parts of Berlin for the same price. Depending on the time of day, the buses run every 6 to 15 minutes. Taxis are available outside the airport and a trip to the city center

would typically cost around 25 €. The trip takes between 20 and 30 minutes, depending on the traffic conditions.

From Berlin-Schönefeld: Between 4.30h. and 23h an airport express train travels in 30 minute intervals between Schönefeld and the city center. It provides a direct connection to the conference venue (station “Friedrichstrasse”) which can be reached within 25 minutes. The ticket to the city costs 3,20€. There are also two urban rail trains S9 and S45 which connect the airport to the southern and eastern parts of Berlin. Those have longer operating hours and the city center can be easily reached with a single change to the Berlin metro system. In this case the trip takes a bit longer but the cost of the ticket remains the same. Taxis are also available at a price of about 40€.

Train Transport

Berlin is a major hub of the Deutsche Bahn, the German national railway company. There are direct high speed train connections to all major German cities as well as to all neighboring countries. For example, the trip to Hamburg takes about 1.5 hours to Frankfurt about 3.5 hours and to Cologne about 4.5 hours. The new central station was opened in 2006 and it is a 5 minute walk from it to the Reichstag building and the Brandenburg Tor. This makes the station an excellent starting point for exploring Berlin. It is also connected to both airports by direct bus and train connections. Most trains, including high speed ones, also stop at other train stations within Berlin, so one can get on and off at another, more convenient location. Finally, the major German train stations feature extensive markings and information in English. Furthermore, written and voice information in English is given in all high speed trains. Therefore, conference participants can get around easily without knowing any German.

City transport

Berlin's local public transport network consists of city train/urban rail (S-Bahn), metro (U-Bahn), buses and trams. There are three zones: AB, BC and ABC.

The conference participants will need an AB or ABC ticket – AB ticket covers the whole Berlin area until the city boundary. However, Schönefeld Airport and Potsdam are outside of the Berlin city boundary, i.e. in the zone C. Therefore, participants intending to go to zone C are advised to buy the ABC ticket.

There are various types of tickets offered:

- short distance tickets – up to 3 urban rail or metro stations (change allowed) and up to 6 bus or tram stations (no change)
- single tickets
- daily tickets
- 7-day tickets
- and some more.

We recommend a 7-day ticket:

- for zones AB: 30 €
- for zones ABC: 37.20 € (includes Schönefeld Airpor and Potsdam)

If you fly from Schönefeld Airport or you would like to visit Potsdam, it will be more convenient for you to by the ABC ticket. With the 7-day ticket, on Saturdays, Sundays as well as on working days between 20h and 3h, you can bring with you one adult person and up to three 6 to 14 years old kids.

Buying tickets: Tickets can be purchased at multilingual ticket machines on the platforms of S-and U-Bahn stations. In buses, fares are paid to the bus driver, in trams at machines inside the trains. In larger stations the S-Bahn and BVG provide ticket counters.

Validation of tickets: Before the journey starts tickets must be validated by stamping them at the yellow or red boxes found on the platforms of all S-Bahn and U-Bahn stations. If you travel by bus or tram, there are also validating machines inside the vehicles. In case of inspection, a ticket that is not stamped is invalid. **Important:** if you have a daily or a 7-day ticket, you need to validate it only once, namely before your first journey.

Fare evasion: Anyone caught without a valid ticket must pay a higher fare of 60€. Even people who forgot to stamp their ticket must pay the fine. Note: Ticket inspectors are dressed in plain clothes and will not make any exceptions for tourists. Those who get caught have to show an ID, otherwise the police will be called.

Stops closest to the HU Berlin building are:

- S-Bahn: “S+U Friedrichstrasse” (S-Bahn lines S5, S7, S75, S1, S2, S25)
- U-Bahn: “U Friedrichstrasse, U Französsische Strasse” (U6)
- Bus: “Staatsoper” (lines 100, 200)
- Tram:”Universitätstrasse” (lines 12, M1)

Bicycles

Berlin also has the reputation of having excellent infrastructure for bicycles. The city features an extensive network of bicycle alleys and dedicated bicycle lanes as well as bicycle traffic lights. Bikers are also allowed to use the dedicated bus lanes at major streets. There are bicycle parking facilities throughout the city, including one at the conference venue.

Renting a bicycle in the summer is a must-do in Berlin – there are several bicycle renting places close to the conference venue. For planning a route from place A to place B (via C), visit a web site “Berlin By Bike” <http://www.bbbike.de>

It is also possible to take a bicycle in marked cars of S-Bahn, U-Bahn and tram. However, an additional bicycle ticket has to be purchased.

Food

There is a huge number of locations for having lunch or dinner within walking distance of the conference venue, ranging from simple bakeries and sandwich bars to gourmet restaurants. Since the area contains major tourist attractions, virtually all food locations provide menus in multiple languages.

International fast food chains such as McDonalds, Burger King, and Subway are located within walking distance. Additionally, the streets round the venue are full with local sandwich and salad bars, traditional German bakeries, doner and kebab shops, and Asian fast food places. It is also worth trying the traditional Berlin ‘Curry Wurst’ (curry sausage) sold by street sellers.

A number of restaurants from all price categories are also found in the surrounding area. Those offer traditional German as well as international cuisine. Many restaurants offer the so called ‘business lunch’ which typically includes a main dish and a drink for a fixed price. The prices of such a lunch usually start from 6€.

Last but not least, Berlin is famous for its large ‘green’ community. A lot of restaurants, bakeries, and sandwich bars offer food prepared only with biological ingredients. It is also easy to find vegetarian and vegan dishes and those dishes are clearly indicated in the menus of all restaurants.

Below, we list some of the venues located nearby the conference site together with the approximate prices of business lunch offers or those of the main dishes.

Name	Type of food	Distance	Prices
Cum laude	German and international	on site	from 6€
12 Apostel	pizza and pasta	behind the venue	business lunch, 7-10€
Deponie No. 3	German and Berlin	behind the venue	5-15€
Via Nova	Italian	behind the venue	business lunch, 7-11€
Leon	café and restaurant	3 min walking	5-15€
Restaurant Nolle	international and Berlin	3 min walking	5-15€
Die Maultasche	Swabian delicacies	3 min walking	5-9€
Mövenpick Restaurant	small international menu	2 min walking	6-13€
Daimlers	international	2 min walking	7-20€
Vapiano	Italian, self-service	4 min walking	moderate
Suppenbörse	soups	3 min walking	3-6 euros
Casa Italia	pizza, salads, pasta	4 min walking	6-16€
Block House	steak house	4 min walking	9-12€
Subway	fast food, sandwiches	4 min walking	low, moderate
Meyerbeers Coffee	café	3 min walking	moderate
Odeon	café, snacks, mainly cakes	behind the venue	moderate
Starbucks	café, snacks and cakes	5 min walking	moderate
Kofler Café	café, snacks and meals	2 min walking	5-20€
McDonalds	fast food	4 min walking	low, moderate
Burger King	fast food	4 min walking	low, moderate
Asia gourmet	Asian take-away	4 min walking	low
Klässig's Fish & Chips	fish fast food	4 min walking	low
Damisch	fast food	4 min walking	low
McCornells Obsttressen	fruit, snacks, fast food	4 min walking	low
Fischers Fritz***	German and international	10 min walking	25-48€
Dressler Restaurant	national and international	6 min walking	16-30€
Bocca di Bacco	Italian	5 min walking	17-35€

Weather

Summer in Berlin can have many faces, from rainy and windy 15° to sunny and hot 40°C. Check the weather forecast 2-3 days before the trip, for example here: <http://www.wetteronline.de/wetter/berlin> (Berlin weather forecast)

Safety

Berlin is generally a safe city. Nevertheless, be aware of pick-pocketing, especially at large S-Bahn and U-Bahn stations. In addition, do not sign any papers, petitions or similar offered by random people at the street or in the public transport.

ACL 2016 gratefully acknowledges the following sponsors for their support:

Platinum

Bloomberg

Gold

Google™

**Goldman
Sachs**

A9™

 **IBM Watson**

Silver


NUANCE


SDL Research

Bronze

YAHOO! LABS

Supporter


Digital roots.


**UNIVERSITY of
WASHINGTON**
Computational Linguistics

Student Research Workshop


NSF

Google™


BABAB


fuse | machines

Wireless


MZ

MACHINEZONE