emnlp2017

Copenhagen, Denmark
September 7-11, 2017

Conference on Empirical Methods in Natural Language Processing

Conference Proceedings

www.emnlp2017.net

The EMNLP organizers gratefully acknowledge support from the following sponsors:

Platinum

Gold

Silver

Bronze

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-945626-83-8

Preface by the General Chair

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

We have an exciting program lined up for you, with three Invited talks, fifteen workshops, seven tutorials, nine TACL presentations, 322 reviewed papers presented as both oral talks and posters, and twenty-one demos. I am especially grateful to our Program Chairs, Rebecca Hwa and Sebastian Riedel, who did a fantastic job managing a backbreaking 1,500 paper submissions (1466 reviewed papers). This involved 51 Area chairs and 980 reviewers. We tried some new things this year (never conducive to a smooth process) including a more careful handling of the COIs that result from Area Chair submissions, and the addition of a meta-review step to encourage more thoughtful reviewing. We are soliciting feedback on the meta-review process, from both reviewers and authors. Despite the additional time involvement, many of the Area Chairs embraced this new approach, and would like to repeat it. However, there are clearly a few dissenters, since Rebecca and Sebastian ended up writing around 200 meta-reviews themselves at the last minute! We are also trying to raise the visibility and status of the poster sessions by integrating them as parallel sessions alongside oral talks, with poster session chairs. This is in response to the survey results from EMNLP 2015 that indicated a decided preference for smaller, more frequent poster sessions during the day rather than evening mega-sessions. Finally, Rebecca and Sebastian are bringing you three outstanding invited speakers, Dan Jurafsky, Sharon Goldwater, and Nando de Freitas. No program chairs ever worked harder to bring you a superb set of presentations in an attendee friendly setting.

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

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

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

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

Best Regards,
Martha Palmer
EMNLP 2017 General Chair

Preface by the Program Committee Co-Chairs

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

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

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

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

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

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

- Martha Palmer, whose encouragement and advice as the general chair has been invaluable every step of the way;

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

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

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

Organizing Committee

General Chair
Martha Palmer, University of Colorado

Local Arrangements Chair
Priscilla Rasmussen, ACL Business Manager

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

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

Local Sponsorship Chair
Daniel Hardt, Copenhagen Business School

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

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

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

Publications Sr Chair
Siddharth Patwardhan, Apple

Publications Jr Chair
Preethi Raghavan, TJ Watson Lab, IBM

Publicity Chair
Isabelle Augenstein, University of Copenhagen

Web Chair
Anders Johannsen, Apple

Conference Handbook Chair
Joachim Bingel, University of Copenhagen

Conference Handbook Advisor
Matt Post, Johns Hopkins University

Handbook Proofreader
Pontus Stenetorp, University College London

Conference App Chair
Chloé Braud, University of Copenhagen

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

SIGDAT Liason
Chris Callison-Birch, University of Pennsylvania

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

Area Chairs
Information Extraction, Information Retrieval, and Question Answering
Mihai Surdeanu, University of Arizona
Jing Jiang, Singapore Management University
Hinrich Schütze, LMU Munich
Sameer Singh, UC Irvine
Scott Wen-tau Yih, MSR
Tomek Strzalkowski, SUNY Albany

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

Linguistic Theories and Psycholinguistics
William Schuler, The Ohio State University

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

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

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

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

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

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

Spoken Language Processing
Mari Ostendorf, University of Washington

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

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

Primary Reviewers

Muhammad Abdul-Mageed; Amjad Abu-Jbara; Heike Adel; Željko Agić; Eneko Agirre; Salah Ait-Mokhtar; Ahmet Aker; Cem Akkaya; Afra Alishahi; Alexandre Allauzen; Tim Althoff; Carlos Alzate; Bharat Ram Ambati; Antonios Anastasopoulos; Daniel Andor; Jacob Andreas; Nicholas Andrews;

Anietie Andy; Gabor Angeli; Marianna Apidianaki; Yuki Arase; Arturo Argueta; Ron Artstein; Yoav Artzi; Ehsaneddin Asgari; Nicholas Asher; Ramón Astudillo; Isabelle Augenstein; Michael Auli; Eleftherios Avramidis; amittai axelrod; Wilker Aziz; Yoram Bachrach; Hessam Bagherinezhad; Collin Baker; Niranjan Balasubramanian; Jason Baldridge; Timothy Baldwin; Tyler Baldwin; Kalika Bali; Miguel Ballesteros; Siddhartha Banerjee; Chitta Baral; Marco Baroni; Alberto Barrón-Cedeño; Pierpaolo Basile; Fernando Batista; Riza Theresa Batista-Navarro; Timo Baumann; Daniel Beck; Beata Beigman Klebanov; Núria Bel; Yonatan Belinkov; Dane Bell; Eric Bell; Kedar Bellare; Islam Beltagy; Anja Belz; Emily M. Bender; Darina Benikova; Luciana Benotti; Jonathan Berant; Taylor Berg-Kirkpatrick; Nicola Bertoldi; Laurent Besacier; Steven Bethard; Chandra Bhagavatula; Suma Bhat; Archna Bhatia; Joachim Bingel; Alexandra Birch; Arianna Bisazza; Yonatan Bisk; Johannes Bjerva; Anders Björkelund; Alan W Black; Eduardo Blanco; John Blitzer; Ivan Bogatyy; Bernd Bohnet; Gemma Boleda; Kalina Bontcheva; Stefan Bordag; Johan Bos; Houda Bouamor; Guillaume Bouchard; Samuel Bowman; Kristy Boyer; S.R.K. Branavan; Chloé Braud; Fabienne Braune; Felipe Bravo-Marquez; Chris Brew; Chris Brockett; Julian Brooke; Caroline Brun; Dominique Brunato; William Bryce; Christian Buck; Paul Buitelaar; Florin Bulgarov; Wray Buntine; Jill Burstein; Bill Byrne; Donna Byron; Elena Cabrio; Aoife Cahill; Nicoletta Calzolari; Jose Camacho-Collados; Erik Cambria; Nicola Cancedda; Marie Candito; Hailong Cao; Cornelia Caragea; Dallas Card; Xavier Carreras; Francisco Casacuberta; Tommaso Caselli; Taylor Cassidy; Vittorio Castelli; Asli Celikyilmaz; Daniel Cer; Özlem Çetinoğlu; Mauro Cettolo; Nathanael Chambers; Yee Seng Chan; Angel Chang; Baobao Chang; Kai-Wei Chang; Snigdha Chaturvedi; Wanxiang Che; Ciprian Chelba; Boxing Chen; Chen Chen; Danqi Chen; Hsin-Hsi Chen; John Chen; Lei Chen; Tao Chen; Wenliang Chen; Yidong Chen; Yun-Nung Chen; Zhiyuan Chen; Jianpeng Cheng; Colin Cherry; Jackie Chi Kit Cheung; Hai Leong Chieu; laura chiticariu; Eunsol Choi; Yejin Choi; Christos Christodoulopoulos; Grzegorz Chrupała; Tagyoung Chung; Philipp Cimiano; Kevin Clark; Stephen Clark; Ann Clifton; Maximin Coavoux; Anne Cocos; Nigel Collier; Michael Collins; Miriam Connor; John Conroy; Matthieu Constant; Danish Contractor; Ryan Cotterell; Benoit Crabbé; Danilo Croce; Montse Cuadros; Iria da Cunha; Andrew Dai; Bhavana Dalvi; Lena Dankin; Amitava Das; Pradeep Dasigi; Munmun De Choudhury; Adrià de Gispert; Daniël de Kok; Eric De La Clergerie; Gerard de Melo; Vera Demberg; Thomas Demeester; Dina Demner-Fushman; Steve DeNeefe; John DeNero; Lingjia Deng; Pascal Denis; Michael Denkowski; Tejaswini Deoskar; Valeria dePaiva; Leon Derczynski; Nina Dethlefs; Ann devitt; Jacob Devlin; Giuseppe Di Fabbrizio; Georgiana Dinu; Ellen Dodge; Jesse Dodge; A. Seza Doğruöz; Li Dong; Qing Dou; Doug Downey; Gabriel Doyle; Eduard Dragut; Mark Dras; Markus Dreyer; Lan Du; Loic Dugast; Kevin Duh; Greg Durrett; Chris Dyer; Marc Dymetman; Richard Eckart de Castilho; Judith Eckle-Kohler; Maud Ehrmann; Andreas Eisele; Jason Eisner; Asif Ekbal; Michael Elhadad; Messina Enza; Gülşen Eryiğit; Hugo Jair Escalante; Keelan Evanini; James Fan; Federico Fancellu; Hao Fang; Stefano Faralli; Richárd Farkas; Noura Farra; Manaal Faruqui; Benoit Favre; Afsaneh Fazly; Marcello Federico; Christian Federmann; Geli Fei; Anna Feldman; Yang Feng; Yansong Feng; Raquel Fernandez; Daniel Fernández-González; Olivier Ferret; Katja Filippova; Andrew Finch; Nicholas FitzGerald; Jeffrey Flanigan; Lucie Flekova; Michael Flor; Radu Florian; Antske Fokkens; José A. R. Fonollosa; Mikel L. Forcada; Eric Fosler-Lussier; George Foster; Jennifer Foster; Stefan L. Frank; Stella Frank; Alexander Fraser; Lea Frermann; Daniel Fried; Mario Fritz; Xiao Fu; Atsushi Fujii; Matthias Gallé; Michel Galley; Michael Gamon; Kuzman Ganchev; Juri Ganitkevitch; Wei Gao; Claire Gardent; Matt Gardner; Ekaterina Garmash; Dan Garrette; Milica Gasic; Tao Ge; Michaela Geierhos; Spandana Gella; Georgi Georgiev; Kallirroi Georgila; Ulrich Germann; George Giannakopoulos; Kevin Gimpel; Filip Ginter; Filip Ginter; Goran Glavaš; Alfio Gliozzo; Koldo Gojenola; Yoav Goldberg; Dan Goldwasser; Juan Carlos Gomez; Carlos Gómez-Rodríguez; Graciela

Gonzalez; Matthew R. Gormley; Cyril Goutte; Amit Goyal; Pawan Goyal; Joao Graca; David Grangier; Spence Green; Eleni Gregoromichelaki; Cyril Grouin; Adam Grycner; Curry Guinn; Hongyu GUO; Jiafeng Guo; Jiang Guo; Weiwei Guo; Nitish Gupta; Pankaj Gupta; Sonal Gupta; Francisco Guzmán; Nizar Habash; Barry Haddow; Gholamreza Haffari; Masato Hagiwara; Udo Hahn; Gus Hahn-Powell; Dilek Hakkani-Tur; David Hall; Keith Hall; Na-Rae Han; Oul Han; Shuguang Han; Xianpei Han; Sanda Harabagiu; Daniel Hardt; Mark Hasegawa-Johnson; Kazuma Hashimoto; Eva Hasler; Helen Hastie; Katsuhiko Hayashi; He He; Hua He; Luheng He; Wenqi He; Xiaodong He; Yifan He; Yulan He; Kenneth Heafield; Michael Heilman; James Henderson; John Henderson; Aron Henriksson; Aurélie Herbelot; Teresa Herrmann; Daniel Hershcovich; Jack Hessel; Ryuichiro Higashinaka; Derrick Higgins; Felix Hill; Erhard Hinrichs; Gerold Hintz; Tsutomu Hirao; Julia Hirschberg; Graeme Hirst; Hieu Hoang; Nathan Hodas; Kristy Hollingshead; Ales Horak; Chiori Hori; Julian Hough; Yifan Hu; Fei Huang; Heyan Huang; Liang Huang; Lifu Huang; Minlie Huang; Ruihong Huang; Shujian Huang; Xuanjing Huang; Zhongqiang Huang; Luwen Huangfu; Samar Husain; Young-Sook Hwang; Gonzalo Iglesias; Ryu Iida; Diana Inkpen; Naoya Inoue; Radu Tudor Ionescu; Ozan Irsoy; Alexei V. Ivanov; Mohit Iyyer; Guillaume Jacquet; Peter Jansen; Yangfeng Ji; Ping Jian; Wenbin Jiang; Anders Johannsen; Michael Johnston; Kristiina Jokinen; Doug Jones; Shafiq Joty; Marcin Junczys-Dowmunt; Dan Jurafsky; David Jurgens; Nobuhiro Kaji; Pallika Kanani; Hiroshi Kanayama; Dimitri Kartsaklis; Arzoo Katiyar; Daisuke Kawahara; Aniruddha Kembhavi; Ruth Kempson; Casey Kennington; Mitesh M. Khapra; Tushar Khot; Bernd Kiefer; Douwe Kiela; Yuta Kikuchi; Jin-Dong Kim; Seokhwan Kim; Tracy Holloway King; Brian Kingsbury; Svetlana Kiritchenko; Chunyu Kit; Roman Klinger; Julien Kloetzer; Kevin Knight; Simon Kocbek; Ekaterina Kochmar; Thomas Kollar; Kazunori Komatani; Rik Koncel-Kedziorski; Lingpeng Kong; Ioannis Konstas; Parisa Kordjamshidi; Alexander Kotov; Zornitsa Kozareva; Mikhail Kozhevnikov; Martin Krallinger; Jayant Krishnamurthy; Canasai Kruengkrai; Lun-Wei Ku; Sandra Kübler; Marco Kuhlmann; Roland Kuhn; Shankar Kumar; Jonathan K. Kummerfeld; Tsung-Ting Kuo; Sadao Kurohashi; Nate Kushman; Polina Kuznetsova; Igor Labutov; Mathias Lambert; Patrik Lambert; Vasileios Lampos; Ian Lane; Ni Lao; Mirella Lapata; Jey Han Lau; Alon Lavie; Joseph Le Roux; John Lee; Kenton Lee; Sungjin Lee; Els Lefever; Tao Lei; Alessandro Lenci; Omer Levy; Roger Levy; Mike Lewis; Fangtao Li; Haibo Li; Jing Li; Junyi Jessy Li; Qi Li; Sujian Li; Wenjie Li; Yanen Li; Zhenghua Li; Maria Liakata; Constantine Lignos; Chuan-Jie Lin; Victoria Lin; Wang Ling; Xiao Ling; Tal Linzen; Pierre Lison; Diane Litman; Changsong Liu; Fei Liu; Fei Liu; Jiangming Liu; Jing Liu; Kang Liu; Qian Liu; Qun Liu; Ting Liu; Yang Liu; Yang Liu; Yiqun Liu; Zhanyi Liu; Zhiyuan Liu; Adam Lopez; Oier Lopez de Lacalle; Adrian Pastor López Monroy; Annie Louis; Ryan Lowe; Bin Lu; Wei Lu; Yi Luan; Marco Lui; Minh-Thang Luong; Franco M. Luque; Anh Tuan Luu; Teresa Lynn; Ji Ma; Xuezhe Ma; Klaus Macherey; Wolfgang Macherey; Pierre Magistry; Suraj Maharjan; Wolfgang Maier; Igor Malioutov; Shervin Malmasi; Suresh Manandhar; Gideon Mann; Christopher D. Manning; Saab Mansour; Amin Mantrach; Diego Marcheggiani; Daniel Marcu; Anna Margolis; Alex Marin; Héctor Martínez Alonso; André F. T. Martins; Bruno Martins; Yuichiroh Matsubayashi; Yuji Matsumoto; Takuya Matsuzaki; Austin Matthews; Arne Mauser; Jonathan May; Diana Maynard; Andrew McCallum; Diana McCarthy; David McClosky; Yashar Mehdad; Yelena Mejova; Pablo Mendes; Helen Meng; Haitao Mi; Yishu Miao; Claudiu Mihăilă; Timothy Miller; Tristan Miller; Bonan Min; Paramita Mirza; Dipendra Misra; Dipendra Misra; Makoto Miwa; Daichi Mochihashi; Ashutosh Modi; Karo Moilanen; Manuel Montes; Christof Monz; Taesun Moon; Raymond Mooney; Roser Morante; Véronique MORICEAU; Alessandro Moschitti; Nasrin Mostafazadeh; Roozbeh Mottaghi; Animesh Mukherjee; Dragos Munteanu; Yugo Murawaki; Smaranda Muresan; Kenton Murray; Maria Nadejde; Ajay Nagesh; Mikio Nakano; Ndapandula Nakashole; Preslav Nakov; Courtney Napoles; Jason Naradowsky; Karthik Narasimhan; Shashi Narayan; Alexis Nasr; Vivi Nastase; Borja

Navarro; Roberto Navigli; Adeline Nazarenko; Mark-Jan Nederhof; Arvind Neelakantan; Sapna Negi; Aida Nematzadeh; Graham Neubig; Hwee Tou Ng; Jun-Ping Ng; Vincent Ng; Axel-Cyrille Ngonga Ngomo; Dong Nguyen; Thien Huu Nguyen; Toan Nguyen; Jian-Yun NIE; Nobal Bikram Niraula; Joakim Nivre; Hiroshi Noji; Joel Nothman; Pierre Nugues; Diarmuid Ó Séaghdha; Brendan O'Connor; Stephan Oepen; Kemal Oflazer; Alice Oh; Jong-Hoon Oh; Naoaki Okazaki; Manabu Okumura; Constantin Orasan; Vicente Ordonez; Myle Ott; Muntsa Padró; Alexis Palmer; Martha Palmer; Shimei Pan; Sinno Jialin Pan; Alexander Panchenko; Bo Pang; Denis Paperno; Aasish Pappu; Natalie Parde; Ankur Parikh; Niki Parmar; Patrick Paroubek; Rebecca J. Passonneau; Panupong Pasupat; John K Pate; Siddharth Patwardhan; Michael J. Paul; Adam Pauls; Umashanthi Pavalanathan; Ellie Pavlick; Adam Pease; Viktor Pekar; Anselmo Peñas; Xiaochang Peng; Sergio Penkale; Gerald Penn; Julien PEREZ; Verónica Pérez-Rosas; Johann Petrak; Slav Petrov; Nghia The Pham; Peter Phandi; Olivier Pietquin; Daniele Pighin; Mohammad Taher Pilehvar; Joelle Pineau; Yuval Pinter; Emily Pitler; Barbara Plank; Julien Plu; Massimo Poesio; Tamara Polajnar; Heather Pon-Barry; Simone Paolo Ponzetto; Ana-Maria Popescu; Andrei Popescu-Belis; Maja Popović; Fred Popowich; Soujanya Poria; Matt Post; Christopher Potts; Vinodkumar Prabhakaran; Daniel Preoţiuc-Pietro; Prokopis Prokopidis; Emily Prud'hommeaux; Jay Pujara; Matthew Purver; Ashequl Qadir; Longhua Qian; Xian Qian; Long Qiu; Minghui Qiu; Ella Rabinovich; Will Radford; Alessandro Raganato; Preethi Raghavan; Altaf Rahman; Nazneen Fatema Rajani; Rafal Rak; Bhuvana Ramabhadran; Carlos Ramisch; Delip Rao; Ari Rappoport; Mohammad Sadegh Rasooli; Sravana Reddy; Ines Rehbein; Roi Reichart; Nils Reimers; David Reitter; Steffen Remus; Xiang Ren; Adithya Renduchintala; Corentin Ribeyre; Giuseppe Riccardi; Matthew Richardson; Martin Riedl; Verena Rieser; Stefan Riezler; German Rigau; Ellen Riloff; Laura Rimell; Fabio Rinaldi; Eric Ringger; Brian Riordan; Alan Ritter; Brian Roark; Kirk Roberts; Molly Roberts; Tim Rocktäschel; Anna Rohrbach; Marcus Rohrbach; Lina M. Rojas Barahona; Stephen Roller; Salvatore Romeo; Paolo Rosso; Mihai Rotaru; Benjamin Roth; Michael Roth; Alla Rozovskaya; Vasile Rus; Attapol Rutherford; Derek Ruths; Fatiha Sadat; Mehrnoosh Sadrzadeh; Markus Saers; Kenji Sagae; Horacio Saggion; Rishiraj Saha Roy; Magnus Sahlgren; Keisuke Sakaguchi; Mohammad Salameh; Shimi Salant; Yunita Sari; Ruhi Sarikaya; Ryohei Sasano; Hassan Sawaf; Asad Sayeed; David Schlangen; Jonathan Schler; Natalie Schluter; Helmut Schmid; Sebastian Schuster; H. Andrew Schwartz; Lane Schwartz; Roy Schwartz; Stephanie Schwartz; Holger Schwenk; Djamé Seddah; Satoshi Sekine; Ethan Selfridge; Jean Senellart; Rico Sennrich; Minjoon Seo; BURR SETTLES; Izhak Shafran; Kashif Shah; Samira Shaikh; Amr Sharaf; Rebecca Sharp; Wade Shen; xiaodong shi; Chaitanya Shivade; Prasha Shrestha; Avirup Sil; Fabrizio Silvestri; Yanchuan Sim; Dan Simonson; Kiril Simov; Steve Skiena; Kevin Small; Sharon Small; Noah A. Smith; Jan Šnajder; Parinaz Sobhani; Richard Socher; Anders Søgaard; Swapna Somasundaran; Hyun-Je Song; Linfeng Song; Sandeep Soni; Alessandro Sordoni; Aitor Soroa; Victor Soto; Rohini Srihari; Somayajulu Sripada; Christian Stab; Sanja Štajner; Miloš Stanojević; Gabriel Stanovsky; Manfred Stede; Mark Steedman; Pontus Stenetorp; Lucas Sterckx; Mark Stevenson; Brandon Stewart; Ian Stewart; Svetlana Stoyanchev; Veselin Stoyanov; Carlo Strapparava; Karl Stratos; Keh-Yih Su; Qi Su; Yu Su; Fabian Suchanek; Alane Suhr; Fei Sun; Huan Sun; Le Sun; Ming Sun; Jun Suzuki; Swabha Swayamdipta; Idan Szpektor; Oscar Täckström; Shabnam Tafreshi; Hiroya Takamura; Partha Talukdar; Chenhao Tan; Hristo Tanev; Duyu Tang; Jian Tang; Xavier Tannier; Makarand Tapaswi; Christoph Teichmann; TestSeb TestRiedel; Joel Tetreault; Simone Teufel; Kapil Thadani; Sam Thomson; Jörg Tiedemann; Christoph Tillmann; Ivan Titov; Takenobu Tokunaga; Gaurav Singh Tomar; Nadi Tomeh; Marc Tomlinson; Sara Tonelli; Kentaro Torisawa; Lamia Tounsi; Isabel Trancoso; Adam Trischler; Yuta Tsuboi; Oren Tsur; Yoshimasa Tsuruoka; Yulia Tsvetkov; Marco Turchi; Ferhan Ture; Kateryna Tymoshenko; Raghavendra Udupa; Stefan Ultes; Lyle Ungar; Shyam Upadhyay; L. Alfonso Urena Lopez; Dmitry Ustalov; Jakob Uszkoreit; Masao Utiyama; Naushad UzZaman; Sowmya Vajjala;

Marco A. Valenzuela-Escárcega; Tim Van de Cruys; Kees van Deemter; Ielka van der Sluis; Benjamin Van Durme; Gertjan van Noord; Marten van Schijndel; David Vandyke; Tony Veale; Eva Maria Vecchi; sriram venkatapathy; Giulia Venturi; Ashish Venugopal; Patrick Verga; Marc Verhagen; Yannick Versley; Guido Vetere; Paul Vicol; David Vilar; Aline Villavicencio; Rob Voigt; Svitlana Volkova; Ivan Vulić; Yogarshi Vyas; V.G.Vinod Vydiswaran; Henning Wachsmuth; Joachim Wagner; Byron Wallace; Matthew Walter; Stephen Wan; Xiaojun Wan; Chuan Wang; Dingquan Wang; Hai Wang; Houfeng WANG; Lu Wang; Qin Iris Wang; Shuai Wang; Shuohang Wang; William Yang Wang; Zhiguo Wang; Zhongqing Wang; Leo Wanner; Taro Watanabe; Nick Webb; Bonnie Webber; Ingmar Weber; Kellie Webster; Julie Weeds; Furu Wei; Gerhard Weikum; David Weir; Ralph Weischedel; David Weiss; Dirk Weissenborn; Robert West; Michael White; Michael Wick; Michael Wiegand; John Wieting; Jason D Williams; Steven Wilson; Guillaume Wisniewski; Michael Wojatzki; Kam-Fai Wong; Hua Wu; Stephen Wu; Joern Wuebker; Rui Xia; Deyi Xiong; Wei Xu; Wenduan Xu; Yadollah Yaghoobzadeh; Adam Yala; Rui Yan; Bishan Yang; Diyi Yang; Yi Yang; Roman Yangarber; Berrin Yanikoglu; Helen Yannakoudakis; tae yano; Mark Yatskar; Seid Muhie Yimam; Wenpeng Yin; Dani Yogatama; Naoki Yoshinaga; Jianfei Yu; Mo Yu; Ning Yu; Zhuoran Yu; François Yvon; Taras Zagibalov; Marcos Zampieri; Fabio Massimo Zanzotto; Sina Zarrieß; Victoria Zayats; Richard Zens; Torsten Zesch; Luke Zettlemoyer; Congle Zhang; Hao Zhang; Justine Zhang; Lei Zhang; Meishan Zhang; Min Zhang; Qi Zhang; Yuan Zhang; Hai Zhao; Jun Zhao; Shiqi Zhao; Wayne Xin Zhao; Bowen Zhou; Muhua Zhu; Xiaodan Zhu; Yukun Zhu; Michael Zock; Chengqing Zong; Ingrid Zukerman; Pierre Zweigenbaum

Invited Speaker: Dan Jurafsky, Stanford University
"Does This Vehicle Belong to You”? Processing the Language of Policing for Improving Police-Community Relations"

Abstract: Police body-cameras have the potential to play an important role in understanding and improving police-community relations. In this talk I describe a series of studies conducted by our large interdisciplinary team at Stanford that use speech and natural language processing on body-camera recordings to model the interactions between police officers and community members in traffic stops. We use text and speech features to automatically measure linguistic aspects of the interaction, from discourse factors like conversational structure to social factors like respect. I describe the differences we find in the language directed toward black versus white community members, and offer suggestions for how these findings can be used to help improve the fraught relations between police officers and the communities they serve.

Bio: Dan Jurafsky is Professor and Chair of Linguistics and Professor of Computer Science, at Stanford University. His research has focused on the extraction of meaning, intention, and affect from text and speech, on the processing of Chinese, and on applying natural language processing to the cognitive and social sciences. Dan’s deep interest in NLP education led him to co-write with Jim Martin the widely-used textbook "Speech and Language Processing” (whose 3rd edition is in (slow) progress) and co-teach with Chris Manning the first massive open online class on natural language processing. Dan was the recipient of the 2002 MacArthur Fellowship and is a 2015 James Beard Award Nominee for his book, "The Language of Food: A Linguist Reads the Menu".

Invited Speaker: Sharon Goldwater, University of Edinburgh
Towards more universal language technology: unsupervised learning from speech

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

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

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

Invited Speaker: Nando de Freitas, Google Deepmind
Physical simulation, learning and language

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

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

Table of Contents

Conference Program

Friday, September 8, 2017

19:00–22:00 Welcome Reception

Saturday, September 9, 2017

07:30–17:30 Registration Day 1

08:00–08:30 ***Morning Coffee***

08:30–09:00 Plenary Session. Opening Remarks

08:30–09:00 *Opening Remarks*
General Chair, PC Co-Chairs

09:00–10:00 Plenary Session. Invited Talk by Nando de Freitas

09:00–10:00 *Physical simulation, learning and language*
Nando de Freitas

10:00–10:30 ***Coffee Break***

10:30–12:10 Session 1A: Syntax 1

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

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

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

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

10:30–12:10 Session 1B: Information Extraction 1

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

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

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

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

Saturday, September 9, 2017 (continued)

10:30–12:10 Session 1C: Multilingual NLP

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

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

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

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

10:30–12:10 Session 1D: Poster Session: Demo

12:10–13:40 *Lunch*

13:40–15:20 Session 2A: Machine Translation 1

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

14:05–14:30 *Neural Machine Translation with Word Predictions*
Rongxiang Weng, Shujian Huang, Zaixiang Zheng, XIN-YU DAI and Jiajun CHEN

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

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

Saturday, September 9, 2017 (continued)

13:40–15:20 Session 2B: Language Grounding

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

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

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

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

13:40–15:20 Session 2C: Discourse and Summarization

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

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

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

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

Saturday, September 9, 2017 (continued)

13:40–15:20 Session 2D: Poster Session. Embeddings

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

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

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

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

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

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

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

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

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

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

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

Saturday, September 9, 2017 (continued)

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

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

13:40–15:20 Session 2E: Poster Session. Machine Learning 1

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

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

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

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

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

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

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

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

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

Saturday, September 9, 2017 (continued)

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

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

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

13:40–15:20 Session 2F: Poster Session. Sentiment Analysis 1

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

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

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

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

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

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

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

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

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

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

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

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

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

15:20–15:50 ***Coffee Break***

15:50–17:30 **Session 3A: Machine Learning 2**

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

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

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

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

Saturday, September 9, 2017 (continued)

15:50–17:30 Session 3B: Generation

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

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

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

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

15:50–17:30 Session 3C: Semantics 1

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

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

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

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

Saturday, September 9, 2017 (continued)

15:50–17:30 Session 3D: Poster Session. Syntax 2

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

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

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

A Sub-Character Architecture for Korean Language Processing
Karl Stratos

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

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

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

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

Segmentation-Free Word Embedding for Unsegmented Languages
Takamasa Oshikiri

Saturday, September 9, 2017 (continued)

15:50–17:30 Session 3E: Poster Session. Question Answering and Machine Comprehension

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

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

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

15:50–17:30 Session 3F: Poster Session. Multimodal NLP 1

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

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

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

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

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

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

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

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

Saturday, September 9, 2017 (continued)

Reinforced Video Captioning with Entailment Rewards
Ramakanth Pasunuru and Mohit Bansal

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

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

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

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

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

Sunday, September 10, 2017

07:30–17:30 **Registration Day 2**

08:00–09:00 ***Morning Coffee***

09:00–10:00 Plenary Session. Invited Talk by Sharon Goldwater

09:00–10:00 *Towards more universal language technology: unsupervised learning from speech*
Sharon Goldwater

10:00–10:30 *Coffee Break*

10:30–12:10 Session 4A: Reading and Retrieving

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

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

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

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

10:30–12:10 Session 4B: Multimodal NLP 2

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

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

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

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

10:30–12:10 Session 4C: Human Centered NLP and Linguistic Theory

10:30–12:10 Session 4D: Poster Session. Semantics 2

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

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

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

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

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

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

Predicting Word Association Strengths
Andrew Cattle and Xiaojuan Ma

10:30–12:10 Session 4E: Poster Session. Discourse

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

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

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

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

Sunday, September 10, 2017 (continued)

10:30–12:10 Session 4F: Poster Session. Machine Translation and Multilingual NLP 1

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

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

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

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

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

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

Massive Exploration of Neural Machine Translation Architectures
Denny Britz, Anna Goldie, Minh-Thang Luong and Quoc Le

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

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

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

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

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

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

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

12:10–13:40 ***Lunch***

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

13:40–15:20 **Session 5A: Semantics 3**

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

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

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

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

Sunday, September 10, 2017 (continued)

13:40–15:20 Session 5B: Computational Social Science 1

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

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

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

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

13:40–15:20 Session 5C: Sentiment Analysis 2

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

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

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

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

13:40–15:20 Session 5D: Poster Session. Syntax 3

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

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

Incremental Graph-based Neural Dependency Parsing
Xiaoqing Zheng

Neural Discontinuous Constituency Parsing
Miloš Stanojević and Raquel Garrido Alhama

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

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

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

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

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

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

13:40–15:20 Session 5E: Poster Session. Relations

Global Normalization of Convolutional Neural Networks for Joint Entity and Relation Classification
Heike Adel and Hinrich Schütze

End-to-End Neural Relation Extraction with Global Optimization
Meishan Zhang, Yue Zhang and Guohong Fu

KGEval: Accuracy Estimation of Automatically Constructed Knowledge Graphs
Prakhar Ojha and Partha Talukdar

Sparsity and Noise: Where Knowledge Graph Embeddings Fall Short
Jay Pujara, Eriq Augustine and Lise Getoor

Dual Tensor Model for Detecting Asymmetric Lexico-Semantic Relations
Goran Glavaš and Simone Paolo Ponzetto

Incorporating Relation Paths in Neural Relation Extraction
Wenyuan Zeng, Yankai Lin, Zhiyuan Liu and Maosong Sun

Adversarial Training for Relation Extraction
Yi Wu, David Bamman and Stuart Russell

Context-Aware Representations for Knowledge Base Relation Extraction
Daniil Sorokin and Iryna Gurevych

A Soft-label Method for Noise-tolerant Distantly Supervised Relation Extraction
Tianyu Liu, Kexiang Wang, Baobao Chang and Zhifang Sui

A Sequential Model for Classifying Temporal Relations between Intra-Sentence Events
Prafulla Kumar Choubey and Ruihong Huang

Deep Residual Learning for Weakly-Supervised Relation Extraction
YiYao Huang and William Yang Wang

Noise-Clustered Distant Supervision for Relation Extraction: A Nonparametric Bayesian Perspective
Qing Zhang and Houfeng Wang

Exploring Vector Spaces for Semantic Relations
Kata Gábor, Haifa Zargayouna, Isabelle Tellier, Davide Buscaldi and Thierry Charnois

Temporal dynamics of semantic relations in word embeddings: an application to predicting armed conflict participants
Andrey Kutuzov, Erik Velldal and Lilja Øvrelid

13:40–15:20 Session 5F: Poster Session. Language Models, Text Mining, and Crowd Sourcing

Dynamic Entity Representations in Neural Language Models
Yangfeng Ji, Chenhao Tan, Sebastian Martschat, Yejin Choi and Noah A. Smith

Towards Quantum Language Models
Ivano Basile and Fabio Tamburini

Reference-Aware Language Models
Zichao Yang, Phil Blunsom, Chris Dyer and Wang Ling

A Simple Language Model based on PMI Matrix Approximations
Oren Melamud, Ido Dagan and Jacob Goldberger

Syllable-aware Neural Language Models: A Failure to Beat Character-aware Ones
Zhenisbek Assylbekov, Rustem Takhanov, Bagdat Myrzakhmetov and Jonathan N. Washington

Inducing Semantic Micro-Clusters from Deep Multi-View Representations of Novels
Lea Frermann and György Szarvas

Initializing Convolutional Filters with Semantic Features for Text Classification
Shen Li, Zhe Zhao, Tao Liu, Renfen Hu and Xiaoyong Du

Shortest-Path Graph Kernels for Document Similarity
Giannis Nikolentzos, Polykarpos Meladianos, Francois Rousseau, Yannis Stavrakas and Michalis Vazirgiannis

Adapting Topic Models using Lexical Associations with Tree Priors
Weiwei Yang, Jordan Boyd-Graber and Philip Resnik

Finding Patterns in Noisy Crowds: Regression-based Annotation Aggregation for Crowdsourced Data
Natalie Parde and Rodney Nielsen

CROWD-IN-THE-LOOP: A Hybrid Approach for Annotating Semantic Roles
Chenguang Wang, Alan Akbik, laura chiticariu, Yunyao Li, Fei Xia and Anbang Xu

A Joint Many-Task Model: Growing a Neural Network for Multiple NLP Tasks
Kazuma Hashimoto, caiming xiong, Yoshimasa Tsuruoka and Richard Socher

15:20–15:50 ***Coffee Break***

15:50–17:30 **Session 6A: Machine Translation 2**

15:50–16:15 *Earth Mover's Distance Minimization for Unsupervised Bilingual Lexicon Induction*
Meng Zhang, Yang Liu, Huanbo Luan and Maosong Sun

16:15–16:40 *Unfolding and Shrinking Neural Machine Translation Ensembles*
Felix Stahlberg and Bill Byrne

16:40–17:05 *Graph Convolutional Encoders for Syntax-aware Neural Machine Translation*
Joost Bastings, Ivan Titov, Wilker Aziz, Diego Marcheggiani and Khalil Simaan

17:05–17:30 *Trainable Greedy Decoding for Neural Machine Translation*
Jiatao Gu, Kyunghyun Cho and Victor O.K. Li

Sunday, September 10, 2017 (continued)

15:50–17:30 Session 6B: Text Mining and NLP applications

15:50–17:30 Session 6C: Machine Comprehension

15:50–17:30 Session 6D: Poster Session. Summarization, Generation, Dialog, and Discourse 1

15:50–17:30 **Session 6E: Poster Session. Summarization, Generation, Dialog, and Discourse 2**

Preserving Distributional Information in Dialogue Act Classification
Quan Hung Tran, Ingrid Zukerman and Gholamreza Haffari

Adversarial Learning for Neural Dialogue Generation
Jiwei Li, Will Monroe, Tianlin Shi, Sébastien Jean, Alan Ritter and Dan Jurafsky

Using Context Information for Dialog Act Classification in DNN Framework
Yang Liu, Kun Han, Zhao Tan and Yun Lei

Modeling Dialogue Acts with Content Word Filtering and Speaker Preferences
Yohan Jo, Michael Yoder, Hyeju Jang and Carolyn Rose

Towards Implicit Content-Introducing for Generative Short-Text Conversation Systems
Lili Yao, Yaoyuan Zhang, Yansong Feng, Dongyan Zhao and Rui Yan

Affordable On-line Dialogue Policy Learning
Cheng Chang, Runzhe Yang, Lu Chen, Xiang Zhou and Kai Yu

Generating High-Quality and Informative Conversation Responses with Sequence-to-Sequence Models
Yuanlong Shao, Stephan Gouws, Denny Britz, Anna Goldie, Brian Strope and Ray Kurzweil

Bootstrapping incremental dialogue systems from minimal data: the generalisation power of dialogue grammars
Arash Eshghi, Igor Shalyminov and Oliver Lemon

Composite Task-Completion Dialogue Policy Learning via Hierarchical Deep Reinforcement Learning
Baolin Peng, Xiujun Li, Lihong Li, Jianfeng Gao, Asli Celikyilmaz, Sungjin Lee and Kam-Fai Wong

Why We Need New Evaluation Metrics for NLG
Jekaterina Novikova, Ondřej Dušek, Amanda Cercas Curry and Verena Rieser

Challenges in Data-to-Document Generation
Sam Wiseman, Stuart Shieber and Alexander Rush

15:50–17:30 Session 6F: Poster Session. Computational Social Science 2

All that is English may be Hindi: Enhancing language identification through automatic ranking of the likeliness of word borrowing in social media
Jasabanta Patro, Bidisha Samanta, Saurabh Singh, Abhipsa Basu, Prithwish Mukherjee, Monojit Choudhury and Animesh Mukherjee

Multi-View Unsupervised User Feature Embedding for Social Media-based Substance Use Prediction
Tao Ding, Warren K. Bickel and Shimei Pan

Demographic-aware word associations
Aparna Garimella, Carmen Banea and Rada Mihalcea

A Factored Neural Network Model for Characterizing Online Discussions in Vector Space
Hao Cheng, Hao Fang and Mari Ostendorf

Dimensions of Interpersonal Relationships: Corpus and Experiments
Farzana Rashid and Eduardo Blanco

Argument Mining on Twitter: Arguments, Facts and Sources
Mihai Dusmanu, Elena Cabrio and Serena Villata

Distinguishing Japanese Non-standard Usages from Standard Ones
Tatsuya Aoki, Ryohei Sasano, Hiroya Takamura and Manabu Okumura

Connotation Frames of Power and Agency in Modern Films
Maarten Sap, Marcella Cindy Prasettio, Ari Holtzman, Hannah Rashkin and Yejin Choi

Controlling Human Perception of Basic User Traits
Daniel Preoţiuc-Pietro, Sharath Chandra Guntuku and Lyle Ungar

Topic Signatures in Political Campaign Speeches
Clément Gautrais, Peggy Cellier, René Quiniou and Alexandre Termier

Assessing Objective Recommendation Quality through Political Forecasting
H. Andrew Schwartz, Masoud Rouhizadeh, Michael Bishop, Philip Tetlock, Barbara Mellers and Lyle Ungar

Sunday, September 10, 2017 (continued)

Never Abandon Minorities: Exhaustive Extraction of Bursty Phrases on Microblogs Using Set Cover Problem
Masumi Shirakawa, Takahiro Hara and Takuya Maekawa

18:00–22:00 Social Event

Monday, September 11, 2017

07:30–17:30 Registration Day 3

08:00–09:00 ***Morning Coffee***

09:00–10:00 Plenary Session. Invited Talk by Dan Jurafsky

09:00–10:00 *"Does This Vehicle Belong to You"? Processing the Language of Policing for Improving Police-Community Relations*
Dan Jurafsky

10:00–10:30 ***Coffee Break***

10:30–12:10 Session 7A: Machine Learning 3

10:30–10:55 *Maximum Margin Reward Networks for Learning from Explicit and Implicit Supervision*
Haoruo Peng, Ming-Wei Chang and Wen-tau Yih

10:55–11:20 *The Impact of Modeling Overall Argumentation with Tree Kernels*
Henning Wachsmuth, Giovanni Da San Martino, Dora Kiesel and Benno Stein

11:20–11:45 *Learning Generic Sentence Representations Using Convolutional Neural Networks*
Zhe Gan, Yunchen Pu, Ricardo Henao, Chunyuan Li, Xiaodong He and Lawrence Carin

11:45–12:10 *Repeat before Forgetting: Spaced Repetition for Efficient and Effective Training of Neural Networks*
Hadi Amiri, Timothy Miller and Guergana Savova

10:30–12:10 Session 7B: Syntax 4

10:30–10:55 *Part-of-Speech Tagging for Twitter with Adversarial Neural Networks*
Tao Gui, Qi Zhang, Haoran Huang, Minlong Peng and Xuanjing Huang

10:55–11:20 *Investigating Different Syntactic Context Types and Context Representations for Learning Word Embeddings*
Bofang Li, Tao Liu, Zhe Zhao, Buzhou Tang, Aleksandr Drozd, Anna Rogers and Xiaoyong Du

11:20–11:45 *Does syntax help discourse segmentation? Not so much*
Chloé Braud, Ophélie Lacroix and Anders Søgaard

11:45–12:10 *Nonparametric Bayesian Semi-supervised Word Segmentation*
Ryo Fujii, Ryo Domoto and Daichi Mochihashi

10:30–12:10 Session 7C: Dialogue

10:30–10:55 *Deal or No Deal? End-to-End Learning of Negotiation Dialogues*
Mike Lewis, Denis Yarats, Yann Dauphin, Devi Parikh and Dhruv Batra

10:55–11:20 *Agent-Aware Dropout DQN for Safe and Efficient On-line Dialogue Policy Learning*
Lu Chen, Xiang Zhou, Cheng Chang, Runzhe Yang and Kai Yu

11:20–11:45 *Towards Debate Automation: a Recurrent Model for Predicting Debate Winners*
Peter Potash and Anna Rumshisky

11:45–12:10 *Conversation Modeling on Reddit Using a Graph-Structured LSTM*
Victoria Zayats and Mari Ostendorf

Monday, September 11, 2017 (continued)

10:30–12:10 Session 7D: Poster Session. Machine Translation and Multilingual NLP 2

Joint Prediction of Word Alignment with Alignment Types
Anahita Mansouri Bigvand, Te Bu and Anoop Sarkar

Further Investigation into Reference Bias in Monolingual Evaluation of Machine Translation
Qingsong Ma, Yvette Graham, Timothy Baldwin and Qun Liu

A Challenge Set Approach to Evaluating Machine Translation
Pierre Isabelle, Colin Cherry and George Foster

Knowledge Distillation for Bilingual Dictionary Induction
Ndapandula Nakashole and Raphael Flauger

Machine Translation, it's a question of style, innit? The case of English tag questions
Rachel Bawden

Deciphering Related Languages
Nima Pourdamghani and Kevin Knight

Identifying Cognate Sets Across Dictionaries of Related Languages
Adam St Arnaud, David Beck and Grzegorz Kondrak

Learning Language Representations for Typology Prediction
Chaitanya Malaviya, Graham Neubig and Patrick Littell

Cheap Translation for Cross-Lingual Named Entity Recognition
Stephen Mayhew, Chen-Tse Tsai and Dan Roth

Cross-Lingual Induction and Transfer of Verb Classes Based on Word Vector Space Specialisation
Ivan Vulić, Nikola Mrkšić and Anna Korhonen

Classification of telicity using cross-linguistic annotation projection
Annemarie Friedrich and Damyana Gateva

Semantic Specialisation of Distributional Word Vector Spaces using Monolingual and Cross-Lingual Constraints
Nikola Mrkšić, Ivan Vulić, Diarmuid Ó Séaghdha, Ira Leviant, Roi Reichart, Milica Gašić, Anna Korhonen and Steve Young

Counterfactual Learning from Bandit Feedback under Deterministic Logging : A Case Study in Statistical Machine Translation
Carolin Lawrence, Artem Sokolov and Stefan Riezler

10:30–12:10 Session 7E: Poster Session. Information Extraction 2

Learning Fine-grained Relations from Chinese User Generated Categories
Chengyu Wang, Yan Fan, Xiaofeng He and Aoying Zhou

Improving Slot Filling Performance with Attentive Neural Networks on Dependency Structures
Lifu Huang, Avirup Sil, Heng Ji and Radu Florian

Identifying Products in Online Cybercrime Marketplaces: A Dataset for Fine-grained Domain Adaptation
Greg Durrett, Jonathan K. Kummerfeld, Taylor Berg-Kirkpatrick, Rebecca Portnoff, Sadia Afroz, Damon McCoy, Kirill Levchenko and Vern Paxson

Labeling Gaps Between Words: Recognizing Overlapping Mentions with Mention Separators
Aldrian Obaja Muis and Wei Lu

Deep Joint Entity Disambiguation with Local Neural Attention
Octavian-Eugen Ganea and Thomas Hofmann

MinIE: Minimizing Facts in Open Information Extraction
Kiril Gashteovski, Rainer Gemulla and Luciano Del Corro

Scientific Information Extraction with Semi-supervised Neural Tagging
Yi Luan, Mari Ostendorf and Hannaneh Hajishirzi

NITE: A Neural Inductive Teaching Framework for Domain Specific NER
Siliang Tang, Ning Zhang, Jinjiang Zhang, Fei Wu and Yueting Zhuang

Speeding up Reinforcement Learning-based Information Extraction Training using Asynchronous Methods
Aditya Sharma, Zarana Parekh and Partha Talukdar

Leveraging Linguistic Structures for Named Entity Recognition with Bidirectional Recursive Neural Networks
Peng-Hsuan Li, Ruo-Ping Dong, Yu-Siang Wang, Ju-Chieh Chou and Wei-Yun Ma

Fast and Accurate Entity Recognition with Iterated Dilated Convolutions
Emma Strubell, Patrick Verga, David Belanger and Andrew McCallum

Entity Linking via Joint Encoding of Types, Descriptions, and Context
Nitish Gupta, Sameer Singh and Dan Roth

An Insight Extraction System on BioMedical Literature with Deep Neural Networks
Hua He, Kris Ganjam, Navendu Jain, Jessica Lundin, Ryen White and Jimmy Lin

10:30–12:10 Session 7F: Poster Session. NLP Applications

Word Etymology as Native Language Interference
Vivi Nastase and Carlo Strapparava

A Simpler and More Generalizable Story Detector using Verb and Character Features
Joshua Eisenberg and Mark Finlayson

Multi-modular domain-tailored OCR post-correction
Sarah Schulz and Jonas Kuhn

Learning to Predict Charges for Criminal Cases with Legal Basis
Bingfeng Luo, Yansong Feng, Jianbo Xu, Xiang Zhang and Dongyan Zhao

Quantifying the Effects of Text Duplication on Semantic Models
Alexandra Schofield, Laure Thompson and David Mimno

Identifying Semantically Deviating Outlier Documents
Honglei Zhuang, Chi Wang, Fangbo Tao, Lance Kaplan and Jiawei Han

Detecting and Explaining Causes From Text For a Time Series Event
Dongyeop Kang, Varun Gangal, Ang Lu, Zheng Chen and Eduard Hovy

A Novel Cascade Model for Learning Latent Similarity from Heterogeneous Sequential Data of MOOC
Zhuoxuan Jiang, Shanshan Feng, Gao Cong, Chunyan Miao and Xiaoming Li

Identifying the Provision of Choices in Privacy Policy Text
Kanthashree Mysore Sathyendra, Shomir Wilson, Florian Schaub, Sebastian Zimmeck and Norman Sadeh

An Empirical Analysis of Edit Importance between Document Versions
Tanya Goyal, Sachin Kelkar, Manas Agarwal and Jeenu Grover

Transition-Based Disfluency Detection using LSTMs
Shaolei Wang, Wanxiang Che, Yue Zhang, Meishan Zhang and Ting Liu

Neural Sequence-Labelling Models for Grammatical Error Correction
Helen Yannakoudakis, Marek Rei, Øistein E. Andersen and Zheng Yuan

Adapting Sequence Models for Sentence Correction
Allen Schmaltz, Yoon Kim, Alexander Rush and Stuart Shieber

12:10–13:40 ***Lunch***

13:40–15:25 **Session 8A: Machine Translation and Multilingual/Multimodal NLP (Short)**

13:40–13:55 *A Study of Style in Machine Translation: Controlling the Formality of Machine Translation Output*
Xing Niu, Marianna Martindale and Marine Carpuat

13:55–14:10 *Sharp Models on Dull Hardware: Fast and Accurate Neural Machine Translation Decoding on the CPU*
Jacob Devlin

14:10–14:25 *Exploiting Cross-Sentence Context for Neural Machine Translation*
Longyue Wang, Zhaopeng Tu, Andy Way and Qun Liu

14:25–14:40 *Cross-Lingual Transfer Learning for POS Tagging without Cross-Lingual Resources*
Joo-Kyung Kim, Young-Bum Kim, Ruhi Sarikaya and Eric Fosler-Lussier

13:40–15:25 Session 8C: NLP Applications (Short)

15:25–15:50 ***Coffee Break***

Monday, September 11, 2017 (continued)

15:50–17:25 Plenary Session. Best Paper

17:00–17:25 *Bringing Structure into Summaries: Crowdsourcing a Benchmark Corpus of Concept Maps*
Tobias Falke and Iryna Gurevych

16:20–16:35 *Natural Language Does Not Emerge 'Naturally' in Multi-Agent Dialog*
Satwik Kottur, José Moura, Stefan Lee and Dhruv Batra

16:35–17:00 *Depression and Self-Harm Risk Assessment in Online Forums*
Andrew Yates, Arman Cohan and Nazli Goharian

15:55–16:20 *Men Also Like Shopping: Reducing Gender Bias Amplification using Corpus-level Constraints*
Jieyu Zhao, Tianlu Wang, Mark Yatskar, Vicente Ordonez and Kai-Wei Chang

17:25–17:45 Plenary Session. Closing Remarks

17:25–17:45 *Closing Remarks*
General Chair

Train-O-Matic: Large-Scale Supervised Word Sense Disambiguation in Multiple Languages without Manual Training Data

Tommaso Pasini and **Roberto Navigli**
Department of Computer Science
Sapienza University of Rome
{pasini,navigli}@di.uniroma1.it

Abstract

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

1 Introduction

Word Sense Disambiguation (WSD) is a key task in computational lexical semantics, inasmuch as it addresses the lexical ambiguity of text by making explicit the meaning of words occurring in a given context (Navigli, 2009). Anyone who has struggled with frustratingly unintelligible translations from an automatic system, or with the meaning bias of search engines, can understand the importance for an intelligent system to go beyond the surface appearance of text.

There are two mainstream lines of research in WSD: supervised and knowledge-based WSD. Supervised WSD frames the problem as a classical machine learning task in which, first a training phase occurs aimed at learning a classification model from sentences annotated with word senses and, second the model is applied to previously-unseen sentences focused on a target word. A key difference from many other problems, however, is that the classes to choose from (i.e., the senses of a target word) vary for each word, therefore requiring a separate training process to be performed on a word by word basis. As a result, hundreds of training instances are needed for each ambiguous word in the vocabulary. This would necessitate a million-item training set to be manually created for each language of interest, an endeavour that is currently beyond reach even in resource-rich languages like English.

The second paradigm, i.e., knowledge-based WSD, takes a radically different approach: the idea is to exploit a general-purpose knowledge resource like WordNet (Fellbaum, 1998) to develop an algorithm which can take advantage of the structural and lexical-semantic information in the resource to choose among the possible senses of a target word occurring in context. For example, a PageRank-based algorithm can be developed to determine the probability of a given sense being reached starting from the senses of its context words. Recent approaches of this kind have been shown to obtain competitive results (Agirre et al., 2014; Moro et al., 2014). However, due to its inherent nature, knowledge-based WSD tends to adopt bag-of-word approaches which do not exploit the local lexical context of a target word, including function and collocation words, which limits this approach in some cases.

In this paper we get the best of both worlds and present Train-O-Matic, a novel method for generating huge high-quality training sets for all the words in a language's vocabulary. The approach is language-independent, thanks to its use of a multilingual knowledge resource, BabelNet (Navigli and Ponzetto, 2012), and it can be applied to any kind of corpus. The training sets produced with Train-O-Matic are shown to provide competitive performance with those of manually and semi-

Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pages 78–88
Copenhagen, Denmark, September 7–11, 2017.

automatically tagged corpora. Moreover, state-of-the-art performance is also reported for low resourced languages (i.e., Italian and Spanish) and domains, where manual training data is not available.

2 Building a Training Set from Scratch

In this Section we present Train-O-Matic, a language-independent approach to the automatic construction of a sense-tagged training set. Train-O-Matic takes as input a corpus C (e.g., Wikipedia) and a semantic network $G = (V, E)$. We assume a WordNet-like structure of G, i.e., V is the set of concepts (i.e., synsets) such that, for each word w in the vocabulary, $Senses(w)$ is the set of vertices in V that are expressed by w, e.g., the WordNet synsets that include w as one of their senses.

Train-O-Matic consists of three steps:

- **Lexical profiling:** for each vertex in the semantic network, we compute its Personalized PageRank vector, which provides its lexical-semantic profile (Section 2.1).

- **Sentence scoring:** For each sentence containing a word w, we compute a probability distribution over all the senses of w based on its context (Section 2.2).

- **Sentence ranking and selection:** for each sense s of a word w in the vocabulary, we select those sentences that are most likely to use w in the sense of s (Section 2.3).

2.1 Lexical profiling

In terms of semantic networks the probability of reaching a node v' starting from v can be interpreted as a measure of relatedness between the synsets v and v'. Thus we define the lexical profile of a vertex v in a graph $G = (V, E)$ as the probability distribution over all the vertices v' in the graph. Such distribution is computed by applying the Personalized PagaRank algorithm, a variant of the traditional PageRank (Brin and Page, 1998). While the latter is equivalent to performing random walks with uniform restart probability on every vertex at each step, PPR, on the other hand, makes the restart probability non-uniform, thereby concentrating more probability mass in the surroundings of those vertices having higher restart probability. Formally, (P)PR is computed as follows:

$$v^{(t+1)} = (1 - \alpha)v^{(0)} + \alpha M v^{(t)} \qquad (1)$$

where M is the row-normalized adjacency matrix of the semantic network, the restart probability distribution is encoded by vector $v^{(0)}$, and α is the well-known damping factor usually set to 0.85 (Brin and Page, 1998). If we set $v^{(0)}$ to a unit probability vector $(0, \ldots, 0, 1, 0, \ldots, 0)$, i.e., restart is always on a given vertex, PPR outputs the probability of reaching every vertex starting from the restart vertex after a certain number of steps. This approach has been used in the literature to create semantic signatures (i.e., profiles) of individual concepts, i.e., vertices of the semantic network (Pilehvar et al., 2013), and then to determine the semantic similarity of concepts. As also done by Pilehvar and Collier (2016), we instead use the PPR vector as an estimate of the conditional probability of a word w' given the target sense[1] $s \in V$ of word w:

$$P(w'|s, w) = \frac{\max_{s' \in Senses(w')} v_s(s')}{Z} \qquad (2)$$

where $Z = \sum_{w"} P(w"|s, w)$ is a normalization constant, v_s is the vector resulting from an adequate number of random walks used to calculate PPR, and $v_s(s')$ is the vector component corresponding to sense s'. To fix the number of iterations needed to have a sufficiently accurate vector, we follow Lofgren et al. (2014) and set the error $\delta = 0.00001$ and the number of iterations to $\frac{1}{\delta} = 100,000$.

As a result of this lexical profiling step we have a probability distribution over vocabulary words for each given word sense of interest.

2.2 Sentence scoring

The objective of the second step is to score the importance of word senses for each of the corpus sentences which contain the word of interest. Given a sentence $\sigma = w_1, w_2, \ldots, w_n$, for a given target word w in the sentence ($w \in \sigma$), and for each of its senses $s \in Senses(w)$, we compute the probability $P(s|\sigma, w)$. Thanks to Bayes' theorem we can determine the probability of sense s of w given the

[1]Note that we use senses and concepts (synsets) interchangeably, because – given a word – a word sense unambiguously determines a concept (i.e., the synset it is contained in) and vice versa.

sentence as follows:

$$P(s|\sigma, w) = \frac{P(\sigma|s, w)P(s|w)}{P(\sigma|w)} \quad (3)$$

$$= \frac{P(w_1, \dots, w_n|s, w)P(s|w)}{P(w_1, \dots, w_n|w)}$$

$$\propto P(w_1, \dots, w_n|s, w)P(s|w) \quad (4)$$

$$\approx P(w_1|s, w) \dots P(w_n|s, w)P(s|w) \quad (5)$$

where Formula 4 is proportional to the original probability (due to removing the constant in the denominator) and is approximated with Formula 5 due to the assumption of independence of the words in the sentence. $P(w_i|s, w)$ is calculated as in Formula 2 and $P(s|w)$ is set to $1/|Senses(w)|$ (recall that s is a sense of w). For example, given the sentence σ = "A match is a tool for starting a fire", the target word w = *match* and its set of senses $S_{match} = \{s^1_{match}, s^2_{match}\}$, where s^1_{match} is the sense of *lighter* and s^2_{match} is the sense of *game match*, we want to calculate the probability of each $s^i_{match} \in S_{match}$ of being the correct sense of *match* in the sentence σ. Following Formula 5 we have:

$$\begin{aligned} P(s^1_{match}|\sigma, match) \approx & \\ & P(\text{tool}|s^1_{match}, match) \\ & \cdot P(\text{start}|s^1_{match}, match) \\ & \cdot P(\text{fire}|s^1_{match}, match) \\ & \cdot P(s^1_{match}|match) \\ & = 2.1 \cdot 10^{-4} \cdot 2 \cdot 10^{-3} \cdot 10^{-2} \cdot 5 \cdot 10^{-1} \\ & = 2.1 \cdot 10^{-9} \end{aligned}$$

$$\begin{aligned} P(s^2_{match}|\sigma, match) \approx & \\ & P(\text{tool}|s^2_{match}, match) \\ & \cdot P(\text{start}|s^2_{match}, match) \\ & \cdot P(\text{fire}|s^2_{match}, match) \\ & \cdot P(s^2_{match}|match) \\ & = 10^{-5} \cdot 2.9 \cdot 10^{-4} \cdot 10^{-6} \cdot 5 \cdot 10^{-1} \\ & = 1.45 \cdot 10^{-15} \end{aligned}$$

As can be seen, the first sense of *match* has a much higher probability due to its stronger relatedness to the other words in the context (i.e. *start*, *fire* and *tool*). Note also that all the probabilities for the second sense are at least one magnitude less than the probability of the first sense.

2.3 Sense-based sentence ranking and selection

Finally, for a given word w and a given sense $s_1 \in Senses(w)$, we score each sentence σ in which w appears and s_1 is its most likely sense according to a formula that takes into account the difference between the first (i.e., s_1) and the second most likely sense of w in σ:

$$\Delta_{s_1}(\sigma) = P(s_1|\sigma, w) - P(s_2|\sigma, w) \quad (6)$$

where $s_1 = \arg\max_{s \in Senses(w)} P(s|\sigma, w)$, and $s_2 = \arg\max_{s \in Senses(w) \setminus \{s_1\}} P(s|\sigma, w)$. We then sort all sentences based on $\Delta_{s_1}(\cdot)$ and return a ranked list of sentences where word w is most likely to be sense-annotated with s_1. Although we recognize that other scoring strategies could have been used, this was experimentally the most effective one when compared to alternative strategies, i.e., the sense probability, the number of words related to the target word w, the sentence length or a combination thereof.

3 Creating a Denser and Multilingual Semantic Network

In the previous Section we assumed that WordNet was our semantic network, with synsets as vertices and edges represented by its semantic relations. However, while its lexical coverage is high, with a rich set of fine-grained synsets, at the relation level WordNet provides mainly paradigmatic information, i.e., relations like hypernymy (is-a) and meronymy (part-of). It lacks, on the other hand, syntagmatic relations, such as those that connect verb synsets to their arguments (e.g., the appropriate senses of eat_v and $food_n$), or pairs of noun synsets (e.g., the appropriate senses of bus_n and $driver_n$).

Intuitively, Train-O-Matic would suffer from such a lack of syntagmatic relations, as the relevance of a sense for a given word in a sentence depends directly on the possibility of visiting senses of the other words in the same sentence (cf. Formula 5) via random walks as calculated with Formula 1. Such reachability depends on the connections available between synsets. Because syntagmatic relations are sparse in WordNet, if it was used on its own, we would end up with a poor ranking of sentences for any given word sense. Moreover, even though the methodology presented in Section 2 is language-independent, Train-O-Matic would lack informa-

mouse (animal)		mouse (device)	
WordNet	WordNet$_{BN}$	WordNet	WordNet$_{BN}$
mouse^1_n	mouse^1_n	mouse^4_n	mouse^4_n
tail^1_n	little^1_a	wheel^1_n	computer^1_n
hairless^1_a	rodent^1_n	electronic_device^1_n	pad^4_n
rodent^1_n	cheese^1_n	ball^3_n	cursor^1_n
trunk^3_n	cat^1_n	hand_operated^1_n	operating_system^1_n
elongate^2_a	rat^1_n	mouse_button^1_n	trackball^1_n
house_mouse^1_n	elephant^1_n	cursor^1_n	wheel^1_n
minuteness^1_n	pet^1_n	operate^3_v	joystick^1_n
nude_mouse^1_n	experiment^1_n	object^1_n	Windows^1_n

Table 1: Top-ranking synsets of the PPR vectors computed on WordNet (first and third columns) and WordNet$_{BN}$ (second and fourth columns) for *mouse* as animal (left) and as device (right).

tion (e.g. senses for a word in an arbitrary vocabulary) for languages other than English.

To cope with these issues, we exploit BabelNet,[2] a huge multilingual semantic network obtained from the automatic integration of WordNet, Wikipedia, Wiktionary and other resources (Navigli and Ponzetto, 2012), and create the BabelNet subgraph induced by the WordNet vertices. The result is a graph whose vertices are BabelNet synsets that contain at least one WordNet synset and whose edge set includes all those relations in BabelNet coming either from WordNet itself or from links in other resources mapped to WordNet (such as hyperlinks in a Wikipedia article connecting it to other articles). The greatest contribution of syntagmatic relations comes, indeed, from Wikipedia, as its articles are linked to related articles (e.g., the English Wikipedia *Bus* article[3] is linked to *Passenger*, *Tourism*, *Bus lane*, *Timetable*, *School*, and many more).

Because not all Wikipedia (and other resources') pages are connected with the same degree of relatedness (e.g., countries are often linked, but they are not necessarily closely related to the source article in which the link occurs), we apply the following weighting strategy to each edge $(s, s') \in E$ of our WordNet-induced subgraph of BabelNet $G = (V, E)$:

$$w(s, s') = \begin{cases} 1 & (s, s') \in E(\text{WordNet}) \\ WO(s, s') & \text{otherwise} \end{cases} \quad (7)$$

where E(WordNet) is the edge set of the original WordNet graph and $WO(s, s')$ is the weighted overlap measure which calculates the similarity between two synsets:

$$WO(s, s') = \frac{\sum_{i=1}^{|S|} (r_i^1 + r_i^2)^{-1}}{\sum_{i=1}^{|S|} (2i)^{-1}}$$

where r_i^1 and r_i^2 are the rankings of the i-th synsets in the set S of the components in common between the vectors associated with s and s', respectively. Because at this stage we still have to calculate our synset vector representation, we use the pre-computed NASARI vectors (Camacho-Collados et al., 2015) to calculate WO. This choice is due to WO's higher performance over cosine similarity for vectors with explicit dimensions (Pilehvar et al., 2013).

As a result, each row of the original adjacency matrix M of G will be replaced with the weights calculated in Formula 7 and then normalized in order to be ready for PPR calculation (see Formula 1). An idea of why a denser semantic network has more useful connections and thus leads to better results is provided by the example in Table 1[4], where we show the highest-probability synsets in the PPR vectors calculated with Formula 1 for two different senses of *mouse* (its animal and device senses) when WordNet (left) and our WordNet-induced subgraph of BabelNet (WordNet$_{BN}$, right) are used as the underlying semantic network for PPR computation. Note that WordNet's top synsets are related to the target synset via paradigmatic (i.e., hypernymy and meronymy) relations, while WordNet$_{BN}$ includes many syntagmatically-related synsets (e.g., *exper-*

[2] http://babelnet.org

[3] Retrieved on February 3rd, 2017.

[4] We use the notation w_p^k introduced in (Navigli, 2009) to denote the k-th sense of word w with part-of-speech tag p.

iment for the animal, and *operating system* and *Windows* for the device sense, among others).

4 Experimental Setup

Corpora for sense annotation We used two different corpora to extract sentences: Wikipedia and the United Nations Parallel Corpus (Ziemski et al., 2016). The first is the largest and most up-to-date encyclopedic resource, containing definitional information, the second, on the other hand, is a public collection of parliamentary documents of the United Nations. The application of Train-O-Matic to the two corpora produced two sense-annotated datasets, which we named T-O-M$_{Wiki}$ and T-O-M$_{UN}$, respectively.

Semantic Network We created sense-annotated corpora with Train-O-Matic both when using PPR vectors computed from vanilla WordNet and when using WordNet$_{BN}$, our denser network obtained from the WordNet-induced subgraph of BabelNet (see Section 3).

Gold standard datasets We performed our evaluations using the framework made available by Raganato et al. (2017a) on five different all-words datasets, namely: the Senseval-2 (Edmonds and Cotton, 2001), Senseval-3 (Snyder and Palmer, 2004), SemEval-2007 (Pradhan et al., 2007), SemEval-2013 (Navigli et al., 2013) and SemEval-2015 (Moro and Navigli, 2015) WSD datasets. We focused on nouns only, given the fact that Wikipedia provides connections between nominal synsets only, and therefore contributes mainly to syntagmatic relations between nouns.

Comparison sense-annotated corpora To show the impact of our T-O-M corpora in WSD, we compared its performance on the above gold standard datasets, against training with:

- **SemCor** (Miller et al., 1993), a corpus containing about 226,000 words annotated manually with WordNet senses.
- **One Million Sense-Tagged Instances** (Taghipour and Ng, 2015, OMSTI), a sense-annotated dataset obtained via a semi-automatic approach based on the disambiguation of a parallel corpus, i.e., the United Nations Parallel Corpus, performed by exploiting manually translated word senses. Because OMSTI integrates SemCor to increase coverage, to keep a level playing field we excluded the latter from the corpus.

We note that T-O-M, instead, is fully automatic and does not require any WSD-specific human intervention nor any aligned corpus.

Reference system In all our experiments, we used It Makes Sense (Zhong and Ng, 2010, IMS), a state-of-the-art WSD system based on linear Support Vector Machines, as our reference system for comparing its performance when trained on T-O-M, against the same WSD system trained on other sense-annotated corpora (i.e., SemCor and OMSTI). Following the WSD literature, unless stated otherwise, we report performance in terms of F1, i.e., the harmonic mean of precision and recall.

We note that it is not the purpose of this paper to show that T-O-M, when integrated into IMS, beats all other configurations or alternative systems, but rather to fully automatize the WSD pipeline with performances which are competitive with the state of the art.

Baseline As a traditional baseline in WSD, we used the Most Frequent Sense (MFS) baseline given by the first sense in WordNet. The MFS is a very competitive baseline, due to the sense skewness phenomenon in language (Navigli, 2009).

Number of training sentences per sense Given a target word w, we sorted its senses $Senses(w)$ following the WordNet ordering and selected the top k_i training sentences for the i-th sense according to Formula 6, where:

$$k_i = \frac{1}{i^z} * K \quad (8)$$

with $K = 500$ and $z = 2$ which were tuned on a separate small in-house development dataset[5].

5 Results

5.1 Impact of syntagmatic relations

The first result we report regards the impact of vanilla WordNet vs. our WordNet-induced subgraph of BabelNet (WordNet$_{BN}$) when calculating PPR vectors. As can be seen from Table 2 – which shows the performance of the T-O-M$_{Wiki}$ corpora generated with the two semantic networks – using WordNet for PPR computation decreases

[5]50 word-sense pairs annotated manually.

Dataset	T-O-M$_{Wiki}$ BN	T-O-M$_{Wiki}$ WN
Senseval-2	**70.5**	70.0
Senseval-3	**67.4**	63.1
SemEval-07	**59.8**	57.9
SemEval-13	**65.5**	63.7
SemEval-15	68.6	**69.5**
ALL	**67.3**	65.7

Table 2: F1 of IMS trained on T-O-M when PPR is obtained from the WordNet graph (WN) and from the WordNet-induced subgraph of BabelNet (BN).

the overall performance of IMS from 0.5 to around 4 points across the five datasets, with an overall loss of 1.6 F1 points. Similar performance losses were observed when using T-O-M$_{UN}$ (see Table 3). This corroborates our hunch discussed in Section 3 that a resource like BabelNet can contribute important syntagmatic relations that are beneficial for identifying (and ranking high) sentences which are semantically relevant for the target word sense. In the following experiments, we report only results using WordNet$_{BN}$.

5.2 Comparison against sense-annotated corpora

We now move to comparing the performance of T-O-M, which is fully automatic, against corpora which are annotated manually (SemCor) and semi-automatically (OMSTI). In Table 3 we show the F1-score of IMS on each gold standard dataset in the evaluation framework and on all datasets merged together (last row), when it is trained with the various corpora described above.

As can be seen, T-O-M$_{Wiki}$ and T-O-M$_{UN}$ obtain higher performance than OMSTI (up to 5.5 points above) on 3 out of 5 datasets, and, overall, T-O-M$_{Wiki}$ scores 1 point above OMSTI. The MFS is in the same ballpark as T-O-M$_{Wiki}$, performing better on some datasets and worse on others. We note that IMS trained on T-O-M$_{Wiki}$ succeeds in surpassing or obtaining the same results as IMS trained on SemCor on SemEval-15 and SemEval-13. We view this as a significant achievement given the total absence of manual effort involved in T-O-M. Because overall T-O-M$_{Wiki}$ outperforms T-O-M$_{UN}$, in what follows we report all the results with T-O-M$_{Wiki}$, except for the domain-oriented evaluation (see Section 5.4).

5.3 Performance without backoff strategy

IMS uses the MFS as a backoff strategy when no sense can be output for a target word in context (Zhong and Ng, 2010). Consequently, the performance of the MFS is mixed up with that of the SVM classifier. As shown in Table 4, OMSTI is able to provide annotated sentences for roughly half of the tokens in the datasets. Train-O-Matic, on the other hand, is able to cover almost all words in each dataset with at least one training sentence. This means that in around 50% of cases OMSTI gives an answer based on the IMS backoff strategy.

To determine the real impact of the different training data, we therefore decided to perform an additional analysis of the IMS performance when the MFS backoff strategy is disabled. Because we suspected the system would not always return a sense for each target word, in this experiment we measured precision, recall and their harmonic mean, i.e., F1. The results in Table 5 confirm our hunch, showing that OMSTI's recall drops heavily, thereby affecting F1 considerably. T-O-M performances, instead, remain high in terms of precision, recall and F1. This confirms that OMSTI relies heavily on data (those obtained for the MFS and from SemCor) that are produced manually, rather than semi-automatically.

5.4 Domain-oriented WSD

To further inspect the ability of T-O-M to enable disambiguation in different domains, we decided to evaluate on specific documents from the various gold standard datasets which could be clearly assigned a domain label. Specifically, we tested on 13 SemEval-13 documents from various domains[6] and 2 SemEval-15 documents (namely, maths & computers, and biomedicine) and carried out two separate tests and evaluations of T-O-M on each domain: once using the MFS backoff strategy, and once not using it. In Tables 6 and 7 we report the results of both T-O-M$_{Wiki}$ and T-O-M$_{UN}$ to determine the impact of the corpus type.

As can be seen in the tables, T-O-M$_{Wiki}$ systematically attains higher scores than OMSTI (except for the biology domain), and, in most cases, attains higher scores than MFS when the backoff is used, with a drastic, systematic increase over OMSTI with both Train-O-Matic configurations

[6]Namely biology, climate, finance, health care, politics, social issues and sport.

Dataset	Train-O-Matic$_{Wiki}$	Train-O-Matic$_{UN}$	OMSTI	SemCor	MFS
Senseval-2	70.5	69.0	74.1	**76.8**	72.1
Senseval-3	67.4	68.3	67.2	**73.8**	72.0
SemEval-07	59.8	57.9	62.3	**67.3**	65.4
SemEval-13	**65.5**	62.5	62.8	**65.5**	63.0
SemEval-15	**68.6**	63.5	63.1	66.1	66.3
ALL	67.3	65.3	66.4	**70.4**	67.6

Table 3: F1 of IMS trained on Train-O-Matic, OMSTI and SemCor, and MFS for the Senseval-2, Senseval-3, SemEval-07, SemEval-13 and SemEval-15 datasets.

Dataset	OMSTI	Train-O-Matic	Total
Senseval-2	469	1005	1066
Senseval-3	494	860	900
Semeval-07	89	159	159
Semeval-13	757	1428	1644
Semeval-15	249	494	531
ALL	2058	3946	4300

Table 4: Number of nominal tokens for which at least one training example is provided by OMSTI or Train-O-Matic for each dataset.

Dataset	OMSTI			Train-O-Matic		
	P	R	F1	P	R	F1
Senseval-2	64.8	28.5	39.6	69.5	65.5	**67.4**
Senseval-3	55.7	31.0	39.8	66.1	63.1	**64.6**
SemEval-07	64.1	35.9	46.0	59.8	59.8	**59.8**
SemEval-13	50.7	23.4	32.0	61.3	53.3	**57.0**
SemEval-15	57.0	26.7	36.4	67.0	62.3	**64.6**
ALL	56.5	27.0	36.5	65.1	59.7	**62.3**

Table 5: Precision, Recall and F1 of IMS trained on OMSTI and Train-O-Matic corpus without MFS backoff strategy for Senseval-2, Senseval-3, SemEval-07, SemEval-13 and SemEval-15.

in recall and F1 when the backoff strategy is disabled. This demonstrates the usefulness of the corpora annotated by Train-O-Matic not only on open text, but also on specific domains. We note that T-O-M$_{UN}$ obtains the best results in the politics domain, which is the closest domain to the UN corpus from which its training sentences are obtained.

6 Scaling up to Multiple Languages

Experimental Setup In this section we investigate the ability of Train-O-Matic to scale to low-resourced languages, such as Italian and Spanish, for which training data for WSD is not available.

Thanks to BabelNet, in fact, Train-O-Matic can be used to generate sense-annotated data for any language supported by the knowledge base. Thus, in order to build new training datasets for the two languages, we ran Train-O-Matic on their corresponding versions of Wikipedia, then we tuned the two parameters K and z on an in-house development dataset[7]. In contrast to the English setting, in order to calculate Formula 8 we sorted the senses of each word by vertex degree. Finally we used the output data to train IMS.

Results To perform our evaluation we chose the most recent multilingual task (SemEval 2015 task 13) which includes gold data for Italian and Spanish. As can be seen from Table 8 Train-O-Matic enabled IMS to perform better than the best participating system (Manion and Sainudiin, 2014, SUDOKU) in all three settings (All domains, Maths & Computer and Biomedicine). Its performance was in fact, 1 to 3 points higher, with a 6-point peak on Maths & Computer in Spanish and on Biomedicine in Italian. This demonstrates the ability of Train-O-Matic to enable supervised WSD systems to surpass state-of-the-art knowledge-based WSD approaches in low-resourced languages without relying on manually curated data for training.

7 Related Work

There are two mainstream approaches to Word Sense Disambiguation: supervised and knowledge-based approaches. Both suffer in different ways from the so-called knowledge acquisition bottleneck, that is, the difficulty in obtaining an adequate amount of lexical-semantic data: for training in the case of supervised systems, and for enriching semantic networks in the case of knowledge-based ones (Pilehvar and

[7]We set $K = 100$ and $z = 2.3$ for Spanish and $K = 100$ and $z = 2.5$ for Italian.

Domain	Backoff	T-O-M$_{Wiki}$			T-O-M$_{UN}$			OMSTI			SemCor	MFS	Size
		P	R	F1	P	R	F1	P	R	F1	F1	F1	
Biology	MFS	63.0	63.0	63.0	65.9	65.9	**65.9**	65.9	65.9	**65.9**	66.3	64.4	135
	-	59.0	53.3	56.0	62.3	56.3	**59.2**	48.1	18.5	26.7	-		
Climate	MFS	68.1	68.1	**68.1**	63.4	63.4	63.4	68.0	68.0	68.0	70.1	67.5	194
	-	63.4	50.0	**55.9**	57.5	45.4	50.7	58.0	24.2	34.2	-		
Finance	MFS	68.0	68.0	**68.0**	56.6	56.6	56.6	64.4	64.4	64.4	63.7	56.2	219
	-	62.1	51.6	**56.4**	48.4	40.2	43.9	57.4	28.3	37.9	-		
Health Care	MFS	65.2	65.2	**65.2**	60.1	60.1	60.1	52.9	52.9	52.9	62.7	56.5	138
	-	61.3	55.1	**58.0**	55.6	50.0	52.6	34.6	18.4	24.0	-		
Politics	MFS	65.2	65.2	65.2	66.3	66.3	**66.3**	63.4	63.4	63.4	69.5	67.7	279
	-	62.5	54.8	58.4	63.9	55.9	**59.6**	54.1	21.5	30.8	-		
Social Issues	MFS	68.5	68.5	**68.5**	63.6	63.6	63.6	65.6	65.6	65.6	66.8	67.6	349
	-	63.1	53.0	**57.6**	57.2	47.9	52.1	54.7	25.2	34.5	-		
Sport	MFS	60.3	60.3	60.3	60.9	60.9	**60.9**	58.8	58.8	58.8	60.4	57.6	330
	-	58.3	54.6	**56.4**	58.1	53.3	55.5	45.0	23.0	30.4	-		

Table 6: Performance comparison over SemEval-2013 domain-specific datasets.

Domain	Backoff	T-O-M$_{Wiki}$			T-O-M$_{UN}$			OMSTI			SemCor	MFS	Size
		P	R	F1	P	R	F1	P	R	F1	F1	F1	
Biomedicine	MFS	76.3	76.3	**76.3**	66.0	66.0	66.0	64.9	64.9	64.9	70.3	71.1	100
	-	76.1	72.2	**74.1**	64.4	59.8	62.0	60.5	26.8	37.2	-		
Maths & Computer	MFS	50.0	50.0	**50.0**	48.0	48.0	48.0	36.0	36.0	36.0	40.6	40.9	97
	-	50.0	47.0	**48.5**	47.8	44.0	45.8	21.2	11.0	14.5	-		

Table 7: Performance comparison over the Biomedical and Maths & Computer domains in SemEval-15.

Language	Dataset	Best System	Train-O-Matic		
		F1	P	R	F1
Italian	ALL	56.6	65.1	55.6	**59.9**
	Computers & Math	46.6	52.7	43.3	**47.6**
	Biomedicine	65.9	76.6	67.6	**71.8**
Spanish	ALL	56.3	61.3	54.8	**57.9**
	Computers & Math	42.4	53.3	44.4	**48.5**
	Biomedicine	65.5	71.8	65.5	**68.5**

Table 8: Performance comparison between T-O-M and SemEval-2015's best SUDOKU Run.

Navigli, 2014; Navigli, 2009).

State-of-the-art supervised systems include Support Vector Machines such as IMS (Zhong and Ng, 2010) and, more recently, LSTM neural networks with attention and multitask learning (Raganato et al., 2017b) as well as LSTMs paired with nearest neighbours classification (Melamud et al., 2016; Yuan et al., 2016). The latter also integrates a label propagation algorithm in order to enrich the sense annotated dataset. The main difference from our approach is its need for a manually annotated dataset to start the label propagation algorithm, whereas Train-O-Matic is fully automatic. An evaluation against this system would have been interesting, but neither the proprietary training data nor the code are available at the time of writing.

In order to generalize effectively, these supervised systems require large numbers of training instances annotated with senses for each target word occurrence. Overall, this amounts to millions of training instances for each language of interest, a number that is not within reach for any language. In fact, no supervised system has been submitted in major multilingual WSD competitions for languages other than English (Navigli et al., 2013; Moro and Navigli, 2015). To overcome this problem, new methodologies have recently been developed which aim to create sense-tagged corpora automatically. Raganato et al. (2016) developed 7 heuristics to grow the number of hyperlinks in Wikipedia pages. Otegi et al. (2016) applied a different disambiguation pipeline for each language to parallel text in Europarl (Koehn, 2005) and QTLeap (Agirre et al., 2015) in order to enrich them with semantic annotations. Taghipour and Ng (2015), the work closest to ours, exploits the alignment from English to Chinese sentences of the United Nation Parallel Corpus (Ziemski et al., 2016) to reduce the ambiguity of English words and sense-tag English sentences. The assumption is that the second language is less ambiguous than the first one and that hand-made translations of senses are available for each WordNet synset. This approach is, therefore, semi-automatic and relies on certain assumptions, in contrast to Train-O-Matic which is, instead, fully automatic and can be applied to any kind of corpus (and language) depending on the specific need. Earlier attempts at the automatic extraction of training samples were made by Agirre and De Lacalle (2004) and Fernández et al. (2004). Both exploited the monosemous relatives method (Leacock et al., 1998) in order to retrieve sentences from the Web which contained a given monosemous noun or a relative monosemous word (e.g., a synonym, a hypernym, etc.). As can be seen in (Fernández et al., 2004) this approach can lead to the retrieval of very accurate examples, but its main drawback lies in the number of senses covered. In fact, for all those synsets that do not have any monosemous relative, the system is unable to retrieve examples, thus heavily affecting the performance in terms of recall and F1.

Knowledge-based WSD, instead, bypasses the heavy requirement of sense-annotated corpora by applying algorithms that exploit a general-purpose semantic network, such as WordNet, which encodes the relational information that interconnects synsets via different kinds of relation. Approaches include variants of Personalized PageRank (Agirre et al., 2014) and densest subgraph approximation algorithms (Moro et al., 2014) which, thanks to the availability of multilingual resources such as BabelNet, can easily be extended to perform WSD in arbitrary languages. Other approaches to knowledge-based WSD exploit the definitional knowledge contained in a dictionary. The Lesk algorithm (Lesk, 1986) and its variants (Banerjee and Pedersen, 2002; Kilgarriff and Rosenzweig, 2000; Vasilescu et al., 2004) aim to determine the correct sense of a word by comparing each word-sense definition with the context in which the target word appears. The limit of knowledge-based WSD, however, lies in the absence of mechanisms that can take into account the very local context of a target word occurrence, including non-content words such as prepositions and articles. Furthermore, recent studies seem to suggest that such approaches are barely able to surpass supervised WSD systems when they enrich their networks starting from a comparable amount of annotated data (Pilehvar and Navigli, 2014). With T-O-M, rather than further enriching an existing semantic network, we exploit the information available in the network to annotate raw sentences with sense information and train a state-of-the-art supervised WSD system without task-specific human annotations.

8 Conclusion

In this paper we presented Train-O-Matic, a novel approach to the automatic construction of large training sets for supervised WSD in an arbitrary language. Train-O-Matic removes the burden of manual intervention by leveraging the structural semantic information available in the WordNet graph enriched with additional relational information from BabelNet, and achieves performance competitive to that of semi-automatic approaches and, in some cases, of manually-curated training data. T-O-M was shown to provide training data for virtually all the target ambiguous nouns, in marked contrast to alternatives like OMSTI, which covers in many cases around half of the tokens, resorting to the MFS otherwise. Moreover Train-O-Matic has proven to scale well to low-resourced languages, for which no manually annotated dataset exists, surpassing the current state of the art of knowledge-based systems.

We believe that the ability of T-O-M to overcome the current paucity of annotated data for WSD, coupled with video games with a purpose for validation purposes (Jurgens and Navigli, 2014; Vannella et al., 2014), paves the way for high-quality multilingual supervised WSD. All the training corpora, including approximately one million sentences which cover English, Italian and Spanish, are made available to the community at `http://trainomatic.org`.

As future work we plan to extend our approach to verbs, adjectives and adverbs. Following Bennett et al. (2016) we will also experiment on more realistic estimates of $P(s|w)$ in Formula 5 as well as other assumptions made in our work.

Acknowledgments

The authors gratefully acknowledge the support of the ERC Consolidator Grant MOUSSE No. 726487.

References

Eneko Agirre, António Branco, Martin Popel, and Kiril Simov. 2015. Europarl QTLeap WSD/NED corpus. LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics, Charles University in Prague.

Eneko Agirre and Oier Lopez De Lacalle. 2004. Publicly available topic signatures for all wordnet nominal senses. In *LREC*.

Eneko Agirre, Oier Lopez de Lacalle, and Aitor Soroa. 2014. Random walks for knowledge-based word sense disambiguation. *Computational Linguistics*, 40(1):57–84.

Satanjeev Banerjee and Ted Pedersen. 2002. An adapted Lesk algorithm for word sense disambiguation using wordnet. In *International Conference on Intelligent Text Processing and Computational Linguistics*, pages 136–145. Springer.

Andrew Bennett, Timothy Baldwin, Jey Han Lau, Diana McCarthy, and Francis Bond. 2016. Lexsemtm: A semantic dataset based on all-words unsupervised sense distribution learning. In *Proceedings of the 54nd Annual Meeting of the Association for Computational Linguistics (ACL 2016)*, pages 1513 – 1524, Berlin, Germany.

Sergey Brin and Lawrence Page. 1998. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1–7):107–117.

José Camacho-Collados, Mohammad Taher Pilehvar, and Roberto Navigli. 2015. Nasari: a novel approach to a semantically-aware representation of items. In *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 567–577, Denver, Colorado. Association for Computational Linguistics.

Philip Edmonds and Scott Cotton. 2001. Senseval-2: overview. In *The Proceedings of the Second International Workshop on Evaluating Word Sense Disambiguation Systems*, pages 1–5. Association for Computational Linguistics.

Christiane Fellbaum, editor. 1998. *WordNet: An Electronic Database*. MIT Press, Cambridge, MA.

Juan Fernández, Mauro Castillo Valdés, German Rigau Claramunt, Jordi Atserias Batalla, and Jordi Tormo. 2004. Automatic acquisition of sense examples using exretriever. In *IBERAMIA Workshop on Lexical Resources and The Web for Word Sense Disambiguation*, pages 97–104.

David Jurgens and Roberto Navigli. 2014. It's All Fun and Games until Someone Annotates: Video Games with a Purpose for Linguistic Annotation. *Transactions of the Association for Computational Linguistics (TACL)*, 2:449–464.

Adam Kilgarriff and Joseph Rosenzweig. 2000. Framework and results for english SENSEVAL. *Computers and the Humanities*, 34(1–2):15–48.

Philipp Koehn. 2005. Europarl: A parallel corpus for statistical machine translation. In *MT summit*, volume 5, pages 79–86.

Claudia Leacock, George A Miller, and Martin Chodorow. 1998. Using corpus statistics and wordnet relations for sense identification. *Computational Linguistics*, 24(1):147–165.

Michael Lesk. 1986. Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone. In *Proceedings of the 5th Annual Conference on Systems Documentation, Toronto, Ontario, Canada*, pages 24–26.

Peter A Lofgren, Siddhartha Banerjee, Ashish Goel, and C Seshadhri. 2014. Fast-ppr: Scaling personalized pagerank estimation for large graphs. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1436–1445. ACM.

Steve L Manion and Raazesh Sainudiin. 2014. An iterative "sudoku style" approach to subgraph-based word sense disambiguation. In *In Proceedings of the Third Joint Conference on Lexical and Computational Se- mantics (*SEM 2014)*, pages 40–50, Dublin, Ireland.

Oren Melamud, Jacob Goldberger, and Ido Dagan. 2016. context2vec: Learning generic context embedding with bidirectional lstm. In *Proceedings of CONLL*, pages 51–61.

George A. Miller, Claudia Leacock, Randee Tengi, and Ross Bunker. 1993. A semantic concordance. In *Proceedings of the 3rd DARPA Workshop on Human Language Technology*, pages 303–308, Plainsboro, N.J.

Andrea Moro and Roberto Navigli. 2015. Semeval-2015 task 13: Multilingual all-words sense disambiguation and entity linking. *Proc. of SemEval-2015*.

Andrea Moro, Alessandro Raganato, and Roberto Navigli. 2014. Entity Linking meets Word Sense Disambiguation: a Unified Approach. *Transactions of the Association for Computational Linguistics (TACL)*, 2:231–244.

Roberto Navigli. 2009. Word Sense Disambiguation: A survey. *ACM Computing Surveys*, 41(2):1–69.

Roberto Navigli, David Jurgens, and Daniele Vannella. 2013. Semeval-2013 task 12: Multilingual word sense disambiguation. In *Proceedings of the* 7^{th} *International Workshop on Semantic Evaluation (SemEval 2013), in conjunction with the Second Joint Conference on Lexical and Computational Semantics (*SEM 2013)*, pages 222–231, Atlanta, USA.

Roberto Navigli and Simone Paolo Ponzetto. 2012. BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193:217–250.

Arantxa Otegi, Nora Aranberri, Antonio Branco, Jan Hajic, Steven Neale, Petya Osenova, Rita Pereira, Martin Popel, Joao Silva, Kiril Simov, et al. 2016. Qtleap wsd/ned corpora: Semantic annotation of parallel corpora in six languages. In *Proceedings of the 10th Language Resources and Evaluation Conference, LREC*, pages 3023–3030.

Mohammad Taher Pilehvar and Nigel Collier. 2016. De-conflated semantic representations. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1680–1690, Austin, TX.

Mohammad Taher Pilehvar, David Jurgens, and Roberto Navigli. 2013. Align, disambiguate and walk: A unified approach for measuring semantic similarity. In *Proceedings of ACL*, pages 1341–1351.

Mohammad Taher Pilehvar and Roberto Navigli. 2014. A large-scale pseudoword-based evaluation framework for state-of-the-art word sense disambiguation. *Computational Linguistics*, 40(4):837–881.

Sameer S Pradhan, Edward Loper, Dmitriy Dligach, and Martha Palmer. 2007. Semeval-2007 task 17: English lexical sample, srl and all words. In *Proceedings of the 4th International Workshop on Semantic Evaluations*, pages 87–92.

Alessandro Raganato, Jose Camacho-Collados, and Roberto Navigli. 2017a. Word Sense Disambiguation: A Unified Evaluation Framework and Empirical Comparison. In *Proceedings of EACL*, pages 99–110, Valencia, Spain.

Alessandro Raganato, Claudio Delli Bovi, and Roberto Navigli. 2016. Automatic Construction and Evaluation of a Large Semantically Enriched Wikipedia. In *Proceedings of IJCAI*, pages 2894–2900, New York City, NY, USA.

Alessandro Raganato, Claudio Delli Bovi, and Roberto Navigli. 2017b. Neural sequence learning models for word sense disambiguation. In *Proceedings of Empirical Methods in Natural Language Processing*, Copenhagen, Denmark.

Benjamin Snyder and Martha Palmer. 2004. The english all-words task. In *Senseval-3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text*, pages 41–43, Barcelona, Spain.

Kaveh Taghipour and Hwee Tou Ng. 2015. One million sense-tagged instances for word sense disambiguation and induction. *CoNLL 2015*, page 338.

Daniele Vannella, David Jurgens, Daniele Scarfini, Domenico Toscani, and Roberto Navigli. 2014. Validating and extending semantic knowledge bases using video games with a purpose. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL 2014)*, pages 1294–1304, Baltimore, Maryland. Association for Computational Linguistics.

Florentina Vasilescu, Philippe Langlais, and Guy Lapalme. 2004. Evaluating variants of the lesk approach for disambiguating words. In *Proceedings of LREC*, Lisbon, Portugal.

Dayu Yuan, Julian Richardson, Ryan Doherty, Colin Evans, and Eric Altendorf. 2016. Semi-supervised word sense disambiguation with neural models. *Proceedings of COLING*, pages 1374–1385.

Zhi Zhong and Hwee Tou Ng. 2010. It makes sense: A wide-coverage word sense disambiguation system for free text. In *Proceedings of the ACL 2010 System Demonstrations*, pages 78–83, Uppsala, Sweden. Association for Computational Linguistics.

Micha Ziemski, Marcin Junczys-Dowmunt, and Bruno Pouliquen. 2016. The United Nations parallel corpus v1.0. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, Portoroz, Slovenia. European Language Resources Association (ELRA).