Proceedings of the Fourth Joint Conference on Lexical and Computational Semantics (*SEM 2015)

June 4–5, 2015
Denver, Colorado, USA

Acknowledgements

We thank SIGLEX for sponsoring the joint *SEM / SemEval keynote. We also thank NAACL for handling all the local arrangements.

Gemma Boleda acknowledges funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 655577. Paolo Rosso acknowledges the EC FP 7's IRSES PEOPLE research programme under the Marie Curie grant agreement No 269180 (WIQ-EI).

©2015 The *SEM 2015 Organizing Committee.

All papers ©2015 their respective authors.

This proceedings volume and all papers therein are licensed under a Creative Commons Attribution 4.0 International License.

License details: http://creativecommons.org/licenses/by/4.0/

ISBN 978-1-941643-39-6

*SEM 2015: Joint Conference on Lexical and Computational Semantics

The **Joint Conference on Lexical and Computational Semantics (*SEM)** provides a forum of exchange for the growing number of NLP researchers working on different aspects of semantic processing. After the previous editions of *SEM in Montreal (2012), Atlanta (2013), and Dublin (2014), the 2015 edition will take place in Denver on June 4 and 5 and is colocated with SemEval and NAACL. As in 2014 at COLING, also on this occasion *SEM and SemEval chose to coordinate their programs by featuring a joint invited talk. In this way, *SEM aims to bring together the ACL SIGLEX and ACL SIGSEM communities.

The acceptance rate of *SEM 2015 was quite competitive: out of 98 submissions, we accepted 36 papers for an overall acceptance of 37%. The acceptance rate of long paper that were accepted for oral presentation (18 out of 62) is 29%. The papers cover a wide range of topics including distributional semantics; lexical semantics and lexical acquisition; formal and linguistic semantics; discourse semantics; lexical resources, linked data and ontologies; semantics for applications; and extra-propositional semantics: sentiment and figurative meaning.

The *SEM 2015 program consists of oral presentations for selected long papers and a poster session for long and short papers.

Day One, June 4th:

- Joint *SEM SemEval keynote talk by **Marco Baroni**;
- Oral presentation sessions on distributional semantics, lexical semantics, and extra-propositional semantics;
- Poster session.

Day Two, June 5th:

- Keynote talk by **Preslav Natkov**;
- Oral presentation sessions on semantics for applications, lexical resources and ontologies, formal semantics, and discourse semantics;
- *SEM Best Paper Award.

We cannot finish without saying that *SEM 2015 would not have been possible without the considerable efforts of our area chairs, their reviewers, and the computational semantics community in general.

We hope you will enjoy *SEM 2015,

Martha Palmer, University of Colorado Boulder, General Chair
Gemma Boleda, University of Trento, Program Co-Chair
Paolo Rosso, Universitat Politècnica de València, Program Co-Chair

*SEM 2015 Chairs and Reviewers

General Chair

Martha Palmer, University of Colorado Boulder

Program Co-Chairs

Gemma Boleda, University of Trento
Paolo Rosso, Universitat Politècnica de València

Publication Co-Chairs

Marc Franco-Salvador, Universitat Politècnica de València
Enrique Flores, Universitat Politècnica de València

Area Chairs

Distributional semantics:
Edward Grefenstette, Google DeepMind

Lexical semantics, lexical acquisition, WSD:
Hwee Tou Ng, National University of Singapore

Lexical resources, linked data, ontologies:
Paul Buitelaar, National University of Ireland, Galway

Formal and linguistic semantics:
Aurelie Herbelot, University of Cambridge

Discourse semantics:
Graeme Hirst, University of Toronto

Semantics for applications (textual entailment, IE, QA, summarization, social media):
Alessandro Moschitti, Qatar Computing Research Institute

Extra-propositional semantics (sentiment, metaphor, irony, figurative meanings, etc.):
Tony Veale, University College Dublin

Program Committee

Apoorv Agarwal, Eneko Agirre, Guadalupe Aguado-de-Cea, Nicholas Asher, Timothy Baldwin, Francesco Barbieri, John Barnden, Roberto Basili, Núria Bel, Islam Beltagy, Emily M. Bender, Igor Boguslavsky, Johan Bos, Antal van den Bosch, Paul Buitelaar, Marine Carpuat, Fabio Celli, Grzegorz Chrupała, Yee Seng Chan, Christian Chiarcos, Robin Cooper, Bonaventura Coppola, Anna Corazza, Danilo Croce, Walter Daelemans, Daniel Dahlmeier, Thierry Declerck, Rodolfo Delmonte, Georgiana Dinu, Judith Eckle-Kohler, Katrin Erk, James Fan, Manaal Faruqui, Vanessa Wei Feng, Claire Gardent, Andrew Gargett, Aniruddha Ghosh, Stephan Gouws, Jorge Gracia, Edward Grefenstette, Gregory Grefenstette, Weiwei Guo, Sebastian Hellmann, Iris

Hendrickx, Aurélie Herbelot, Delia Irazú Hernández Farías, Graeme Hirst, Dirk Hovy, Richard Johansson, Shafiq Joty, Dimitri Kartsaklis, Douwe Kiela, Adam Kilgarriff, Valia Kordoni, Anna Korhonen, Jayant Krishnamurthy, Germán Kruszewski, Angeliki Lazaridou, Yoong Keok Lee, Alessandro Lenci, Guofu Li, Patricia Lichtenstein, Annie Louis, Wei Lu, Marie-Catherine de Marneffe, Lluís Màrquez, David Martinez, Sebastian Martschat, Yashar Mehdad, Diana McCarthy, John Philip McCrae, Louise McNally, Dmitrijs Milajevs, Shachar Mirkin, Saif Mohammad, Andrea Moro, Alessandro Moschitti, Philippe Muller, Preslav Nakov, Roberto Navigli, Malvina Nissim, Hwee Tou Ng, Vincent Ng, Nghia The Pham, Siddharth Patwardhan, Sameer Pradhan, Laurette Pretorius, Tamara Polajnar, John Prager, Stephen Pulman, Antonio Reyes, Sebastian Riedel, German Rigau, Tim Rocktäschel, Horacio Rodriguez, Michael Roth, Mehrnoosh Sadrzadeh, Gilles Sérasset, Eyal Shnarch, Jennifer Spenader, Christian Stab, Armando Stellato, Mark Stevenson, Carlo Strapparava, Stan Szpakowicz, Stefan Thater, Ivan Titov, Sara Tonelli, Kateryna Tymoshenko, Christina Unger, Olga Uryupina, Alessandro Valitutti, Tony Veale, Eva Maria Vecchi, Laure Vieu, Andreas Vlachos, Carl Vogel, Gabriela Vulcu, Joachim Wagner, Janyce Wiebe, Annie Zaenen, Roberto Zamparelli, Fabio Massimo Zanzotto, Heike Zinsmeister.

Table of Contents

Neural Networks for Integrating Compositional and Non-compositional Sentiment in Sentiment Composition
Xiaodan Zhu, Hongyu Guo and Parinaz Sobhani 1

Compositional Distributional Semantics with Long Short Term Memory
Phong Le and Willem Zuidema 10

A Hybrid Distributional and Knowledge-based Model of Lexical Semantics
Nikolaos Aletras and Mark Stevenson 20

Distributional semantics for ontology verification
Julien Corman, Laure Vieu and Nathalie Aussenac-Gilles 30

Combining Seemingly Incompatible Corpora for Implicit Semantic Role Labeling
Parvin Sadat Feizabadi and Sebastian Padó 40

Identification of Caused Motion Construction
Jena D. Hwang and Martha Palmer 51

A Methodology for Word Sense Disambiguation at 90% based on large-scale CrowdSourcing
Oier Lopez de Lacalle and Eneko Agirre 61

Learning Structures of Negations from Flat Annotations
Vinodkumar Prabhakaran and Branimir Boguraev 71

A New Dataset and Evaluation for Belief/Factuality
Vinodkumar Prabhakaran, Tomas By, Julia Hirschberg, Owen Rambow, Samira Shaikh, Tomek Strzalkowski, Jennifer Tracey, Michael Arrigo, Rupayan Basu, Micah Clark, Adam Dalton, Mona Diab, Louise Guthrie, Anna Prokofieva, Stephanie Strassel, Gregory Werner, Yorick Wilks and Janyce Wiebe 82

Non-Orthogonal Explicit Semantic Analysis
Nitish Aggarwal, Kartik Asooja, Georgeta Bordea and Paul Buitelaar 92

Combining Mention Context and Hyperlinks from Wikipedia for Named Entity Disambiguation
Ander Barrena, Aitor Soroa and Eneko Agirre 101

Collective Document Classification with Implicit Inter-document Semantic Relationships
Clint Burford, Steven Bird and Timothy Baldwin 106

SGRank: Combining Statistical and Graphical Methods to Improve the State of the Art in Unsupervised Keyphrase Extraction
Soheil Danesh, Tamara Sumner and James H. Martin 117

Towards Semantic Language Classification: Inducing and Clustering Semantic Association Networks from Europarl
Steffen Eger, Niko Schenk and Alexander Mehler 127

Ideological Perspective Detection Using Semantic Features
Heba Elfardy, Mona Diab and Chris Callison-Burch 137

Mapping Different Rhetorical Relation Annotations: A Proposal
Farah Benamara and Maite Taboada 147

Dissecting the Practical Lexical Function Model for Compositional Distributional Semantics
Abhijeet Gupta, Jason Utt and Sebastian Padó 153

Event Extraction as Frame-Semantic Parsing
Alex Judea and Michael Strube 159

Competence in lexical semantics
Andras Kornai, Judit Ács, Márton Makrai, Dávid Márk Nemeskey, Katalin Pajkossy and Gábor Recski 165

Extending a Single-Document Summarizer to Multi-Document: a Hierarchical Approach
Luis Marujo, Ricardo Ribeiro, David Martins de Matos, João Neto, Anatole Gershman and Jaime Carbonell 176

Reading Between the Lines: Overcoming Data Sparsity for Accurate Classification of Lexical Relationships
Silvia Necsulescu, Sara Mendes, David Jurgens, Núria Bel and Roberto Navigli 182

Multi-Level Alignments As An Extensible Representation Basis for Textual Entailment Algorithms
Tae-Gil Noh, Sebastian Padó, Vered Shwartz, Ido Dagan, Vivi Nastase, Kathrin Eichler, Lili Kotlerman and Meni Adler 193

Leveraging Preposition Ambiguity to Assess Compositional Distributional Models of Semantics
Samuel Ritter, Cotie Long, Denis Paperno, Marco Baroni, Matthew Botvinick and Adele Goldberg 199

Learning to predict script events from domain-specific text
Rachel Rudinger, Vera Demberg, Ashutosh Modi, Benjamin Van Durme and Manfred Pinkal 205

Combining Open Source Annotators for Entity Linking through Weighted Voting
Pablo Ruiz and Thierry Poibeau 211

Automatic Generation of a Lexical Resource to support Semantic Role Labeling in Portuguese
Magali Sanches Duran and Sandra Aluísio 216

Can Selectional Preferences Help Automatic Semantic Role Labeling?
Shumin Wu and Martha Palmer 222

Implicit Entity Recognition in Clinical Documents
Sujan Perera, Pablo Mendes, Amit Sheth, Krishnaprasad Thirunarayan, Adarsh Alex, Christopher Heid and Greg Mott 228

A Distant Supervision Approach to Semantic Role Labeling
Peter Exner, Marcus Klang and Pierre Nugues 239

Discovering Hypernymy Relations using Text Layout
Jean-Philippe Fauconnier and Mouna Kamel 249

The complexity of finding the maximum spanning DAG and other restrictions for DAG parsing of natural language
Natalie Schluter 259

Incremental Semantic Construction Using Normal Form CCG Derivation
Yoshihide Kato and Shigeki Matsubara 269

Dependency-Based Semantic Role Labeling using Convolutional Neural Networks
William Foland and James Martin 279

A State-of-the-Art Mention-Pair Model for Coreference Resolution
Olga Uryupina and Alessandro Moschitti 290

Resolving Discourse-Deictic Pronouns: A Two-Stage Approach to Do It
Sujay Kumar Jauhar, Raul Guerra, Edgar Gonzàlez Pellicer and Marta Recasens 300

Graph-based Coherence Modeling For Assessing Readability
Mohsen Mesgar and Michael Strube 310

Conference Program

Thursday, June 4

08:50–10:30 **Block 1 - Distributional semantics**
Session Chair: Martha Palmer

08:50–09:00 ***Welcome***

09:00–10:00 Joint *SEM / SemEval keynote talk by **Marco Baroni**
Playing ficles and running with the corbons: What (multimodal) distributional semantic models learn during their childhood

10:00–10:30 *Neural Networks for Integrating Compositional and Non-compositional Sentiment in Sentiment Composition*
Xiaodan Zhu, Hongyu Guo and Parinaz Sobhani

10:30–11:00 ***Coffee break***

11:00–12:30 **Block 2 - Distributional semantics**
Session Chair: Ed Grefenstette

11:00–11:30 *Compositional Distributional Semantics with Long Short Term Memory*
Phong Le and Willem Zuidema

11:30–12:00 *A Hybrid Distributional and Knowledge-based Model of Lexical Semantics*
Nikolaos Aletras and Mark Stevenson

12:00–12:30 *Distributional semantics for ontology verification*
Julien Corman, Laure Vieu and Nathalie Aussenac-Gilles

12:30–14:00 ***Lunch break***

Thursday, June 4 (continued)

14:00–15:30 **Block 3 - Lexical semantics**
Session Chair: Gemma Boleda

14:00–14:30 *Combining Seemingly Incompatible Corpora for Implicit Semantic Role Labeling*
Parvin Sadat Feizabadi and Sebastian Padó

14:30–15:00 *Identification of Caused Motion Construction*
Jena D. Hwang and Martha Palmer

15:00–15:30 *A Methodology for Word Sense Disambiguation at 90% based on large-scale CrowdSourcing*
Oier Lopez de Lacalle and Eneko Agirre

15:30–16:00 ***Coffee break***

16:00–17:00 **Block 4 - Extra-propositional semantics**
Session Chair: Tony Veale

16:00–16:30 *Learning Structures of Negations from Flat Annotations*
Vinodkumar Prabhakaran and Branimir Boguraev

16:30–17:00 *A New Dataset and Evaluation for Belief/Factuality*
Vinodkumar Prabhakaran, Tomas By, Julia Hirschberg, Owen Rambow, Samira Shaikh, Tomek Strzalkowski, Jennifer Tracey, Michael Arrigo, Rupayan Basu, Micah Clark, Adam Dalton, Mona Diab, Louise Guthrie, Anna Prokofieva, Stephanie Strassel, Gregory Werner, Yorick Wilks and Janyce Wiebe

17:00–19:00 **Poster session with lightning talks intro**

Non-Orthogonal Explicit Semantic Analysis
Nitish Aggarwal, Kartik Asooja, Georgeta Bordea and Paul Buitelaar

Combining Mention Context and Hyperlinks from Wikipedia for Named Entity Disambiguation
Ander Barrena, Aitor Soroa and Eneko Agirre

Collective Document Classification with Implicit Inter-document Semantic Relationships
Clint Burford, Steven Bird and Timothy Baldwin

Thursday, June 4 (continued)

SGRank: Combining Statistical and Graphical Methods to Improve the State of the Art in Unsupervised Keyphrase Extraction
Soheil Danesh, Tamara Sumner and James H. Martin

Towards Semantic Language Classification: Inducing and Clustering Semantic Association Networks from Europarl
Steffen Eger, Niko Schenk and Alexander Mehler

Ideological Perspective Detection Using Semantic Features
Heba Elfardy, Mona Diab and Chris Callison-Burch

Mapping Different Rhetorical Relation Annotations: A Proposal
Farah Benamara and Maite Taboada

Dissecting the Practical Lexical Function Model for Compositional Distributional Semantics
Abhijeet Gupta, Jason Utt and Sebastian Padó

Event Extraction as Frame-Semantic Parsing
Alex Judea and Michael Strube

Competence in lexical semantics
Andras Kornai, Judit Ács, Márton Makrai, Dávid Márk Nemeskey, Katalin Pajkossy and Gábor Recski

Extending a Single-Document Summarizer to Multi-Document: a Hierarchical Approach
Luis Marujo, Ricardo Ribeiro, David Martins de Matos, João Neto, Anatole Gershman and Jaime Carbonell

Reading Between the Lines: Overcoming Data Sparsity for Accurate Classification of Lexical Relationships
Silvia Necsulescu, Sara Mendes, David Jurgens, Núria Bel and Roberto Navigli

Multi-Level Alignments As An Extensible Representation Basis for Textual Entailment Algorithms
Tae-Gil Noh, Sebastian Padó, Vered Shwartz, Ido Dagan, Vivi Nastase, Kathrin Eichler, Lili Kotlerman and Meni Adler

Leveraging Preposition Ambiguity to Assess Compositional Distributional Models of Semantics
Samuel Ritter, Cotie Long, Denis Paperno, Marco Baroni, Matthew Botvinick and Adele Goldberg

Thursday, June 4 (continued)

Learning to predict script events from domain-specific text
Rachel Rudinger, Vera Demberg, Ashutosh Modi, Benjamin Van Durme and Manfred Pinkal

Combining Open Source Annotators for Entity Linking through Weighted Voting
Pablo Ruiz and Thierry Poibeau

Automatic Generation of a Lexical Resource to support Semantic Role Labeling in Portuguese
Magali Sanches Duran and Sandra Aluísio

Can Selectional Preferences Help Automatic Semantic Role Labeling?
Shumin Wu and Martha Palmer

Friday, June 5

09:00–10:30 **Block 1 - Semantics for applications**
Session Chair: Paolo Rosso

09:00–10:00 *60 Years Ago People Dreamed of Talking with a Machine. Are We Any Closer?*
Keynote by Preslav Nakov

10:00–10:30 *Implicit Entity Recognition in Clinical Documents*
Sujan Perera, Pablo Mendes, Amit Sheth, Krishnaprasad Thirunarayan, Adarsh Alex, Christopher Heid and Greg Mott

10:30–11:00 ***Coffee break***

11:00–12:00 **Block 2 - Semantics for applications; Lexical resources and ontologies**
Session Chair: Alessandro Moschitti

11:00–11:30 *A Distant Supervision Approach to Semantic Role Labeling*
Peter Exner, Marcus Klang and Pierre Nugues

11:30–12:00 *Discovering Hypernymy Relations using Text Layout*
Jean-Philippe Fauconnier and Mouna Kamel

12:00–13:30 ***Lunch break***

Friday, June 5 (continued)

13:30–15:00 **Block 3 - Formal semantics**
Session Chair: TBD

13:30–14:00 *The complexity of finding the maximum spanning DAG and other restrictions for DAG parsing of natural language*
Natalie Schluter

14:00–14:30 *Incremental Semantic Construction Using Normal Form CCG Derivation*
Yoshihide Kato and Shigeki Matsubara

14:30–15:00 *Dependency-Based Semantic Role Labeling using Convolutional Neural Networks*
William Foland and James Martin

15:00–15:30 ***Coffee break***

15:30–17:00 **Block 4 - Discourse semantics**
Session Chair: TBD

15:30–16:00 *A State-of-the-Art Mention-Pair Model for Coreference Resolution*
Olga Uryupina and Alessandro Moschitti

16:00–16:30 *Resolving Discourse-Deictic Pronouns: A Two-Stage Approach to Do It*
Sujay Kumar Jauhar, Raul Guerra, Edgar Gonzàlez Pellicer and Marta Recasens

16:30–17:00 *Graph-based Coherence Modeling For Assessing Readability*
Mohsen Mesgar and Michael Strube

17:00 **Best Paper Award and closing**

Invited Talks

Playing ficles and running with the corbons:
What (multimodal) distributional semantic models learn during their childhood
Marco Baroni, University of Trento

Joint work with: Angeliki Lazaridou, Marco Marelli (University of Trento),
Raquel Fernandez (University of Amsterdam), Grzegorz Chrupała (Tilburg University)

Distributional semantic methods have some a priori appeal as models of human meaning acquisition, because they induce word representations from contextual distributions naturally occurring in corpus data without need for supervision. However, learning the meaning of a (concrete) word also involves establishing a link between the word and its typical visual referents, which is beyond the scope of classic, text-based distributional semantics. Since recently several proposals have been put forward about how to induce multimodal word representations from linguistic and visual contexts, it is natural to ask if this line of work, besides its practical implications, can help us to develop more realistic, grounded models of human word learning within the distributional semantics framework.

In my talk, I will report about two studies in which we used multimodal distributional semantics (MDS) to simulate human word learning. In one study, we first measured the ability of subjects to link a nonce word to relevant linguistic and visual associates when prompted only by exposure to minimal corpus evidence about it. We then simulated the same task with an MDS model, finding its behavior remarkably similar to that of subjects. In the second study, we constructed a corpus in which child-directed speech is aligned with real-life pictures of the objects mentioned by care-givers. We then trained our MDS model on these data, and inspected the generalizations it learned about the words in the corpus and the objects they might denote.

The results highlight interesting issues not only for distributional semantics (can we build meaningful word representations from very limited contexts? are such representations reasonably human-like?), but also for the study of human language acquisition (are we "done" with learning a word once we associate it to a referent? do we incrementally refine our word representations? is an explicit cross-situational mechanism really necessary?).

60 Years Ago People Dreamed of Talking with a Machine. Are We Any Closer?
Preslav Nakov, Qatar Computing Research Institute (QCRI)

Joint work with Marti Hearst (UC Berkeley)

The 60-year-old dream of computational linguistics is to make computers capable of communicating with humans in natural language. This has proven hard, and thus research has focused on sub-problems. Even so, the field was stuck with manual rules until the early 90s, when computers became powerful enough to enable the rise of statistical approaches. Eventually, this shifted the main research attention to machine learning from text corpora, thus triggering a revolution in the field.

Today, the Web is the biggest available corpus, providing access to quadrillions of words; and, in corpus-based natural language processing, size does matter. Unfortunately, while there has been substantial research on the Web as a corpus, it has typically been restricted to using page hit counts as an estimate for n-gram word frequencies; this has led some researchers to conclude that the Web should be only used as a baseline.

In this talk, I will reveal some of the hidden potential of the Web that lies beyond the n-gram, with focus on the syntax and semantics of English noun compounds. I will further show how these ideas apply to a number of NLP problems, including syntactic parsing and machine translation, among others. Finally, I will share some thoughts about the future of lexical semantics and machine translation, in view of the ongoing deep learning revolution.