

ACL HLT 2011



**Distributional Semantics and Compositionality (DiSCo'2011)
Workshop at ACL HLT 2011**

Proceedings of the Workshop

24 June 2011
Portland, Oregon, USA

Production and Manufacturing by
Omnipress, Inc.
2600 Anderson Street
Madison, WI 53704 USA

supported by



©2011 The Association for Computational Linguistics

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-13 9781937284022

Introduction

Any NLP system that does semantic processing relies on the assumption of semantic compositionality: the meaning of a phrase is determined by the meanings of its parts and their combination. For this, it is necessary to have automatic methods that are capable to reproduce the compositionality of language.

Recent years have shown the renaissance of interest in distributional semantics. While distributional methods in semantics have proven to be very efficient in tackling a wide range of tasks in natural language processing, e.g., document retrieval, clustering and classification, question answering, query expansion, word similarity, synonym extraction, relation extraction, and many others, they are still strongly limited by being inherently word-based. The main hurdle for vector space models to further progress is the ability to handle compositionality.

The workshop is of potential interest to the researchers working on distributional semantics and compositionality as well as for those interested in extracting non-compositional phrases from large corpora by applying distributional methods that assign a graded compositionality score to a phrase. This score denotes the extent to which the compositionality assumption holds for a given expression. The latter can be used, for example, to decide whether the phrase should be treated as a single unit in applications or included in a dictionary. We have emphasized that the focus is on automatically acquiring semantic compositionality, thereby explicitly avoiding approaches that employ prefabricated lists of non-compositional phrases.

This volume contains papers accepted for publication at DiSCo'2011 Workshop on Distributional Semantics and Compositionality, collocated with ACL-HLT 2011, the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies.

The workshop consists of a main session and a shared task. To the best of our knowledge, this has been the first attempt in the community to offer a dataset and a shared task that allows to explicitly evaluate the models of graded compositionality for phrases per se that occur in three types of grammatical relations: adjective-noun pairs, subject-verb and verb-object pairs in English and German.

For the main session, one long and two short papers have been accepted for publication. Further, seven teams with 19 systems have taken part in the shared task. We consider this a success, taking into consideration that the task is new and difficult.

The description of the task and the results of evaluation are part of these proceedings. In short, approaches ranging from pure statistical association measures to various variations of word space models have been applied to solve the DiSCo task. Six system description papers have been accepted for publication.

Both regular and system description papers have been carefully reviewed by the program committee. We would like to thank the committee for insightful and timely reviews (in spite of the Easter holidays).

The accepted regular articles address a rather wide spectrum of issues within distributional semantics, such as:

- automatic detection of semantic deviance in attributive Adjective-Noun (AN) expressions with

four compositional methods for distributional vectors (Vecchi, Baroni and Zamparelli, 2011);

- encoding syntactic trees in distributed vectors and the application of those for recognizing textual entailment (RTE) (Zanzotto and Dell’Arciprete, 2011);
- two possible generalizations of pointwise mutual information for three-way distributional models (Van de Cruys, 2011).

Last but not least, we would like to thank Dominic Widdows for agreeing to give an invited talk about the theory and practice behind some of recent developments in semantic vectors.

Enjoy the workshop!

The organizers:

- Chris Biemann, UKP lab, TU Darmstadt, Germany
- Eugenie Giesbrecht, FZI Forschungszentrum Informatik¹ at the University of Karlsruhe, Germany

¹Research Center for Information Technology

Organizers:

Chris Biemann, UKP lab, TU Darmstadt, Germany
Eugenie Giesbrecht, FZI Forschungszentrum Informatik, Karlsruhe, Germany

Program Committee:

Enrique Alfonseca, Google Research, Switzerland
Tim Baldwin, University of Melbourne, Australia
Marco Baroni, University of Trento, Italy
Paul Buitelaar, National University of Ireland, Ireland
Chris Brockett, Microsoft Research, Redmond, US
Tim van de Cruys, University of Cambridge, UK
Stefan Evert, University of Osnabrück, Germany
Antske Fokkens, Saarland University, Germany
Silvana Hartmann, TU Darmstadt, Germany
Alfio Massimiliano Gliozzo, IBM, Hawthorne, NY, USA
Mirella Lapata, University of Edinburgh, UK
Ted Pedersen, University of Minnesota, Duluth, USA
Yves Peirsman, Stanford University, USA
Sebastian Rudolph, Karlsruhe Institute of Technology, Germany
Peter D. Turney, National Research Council Canada, Canada
Magnus Sahlgren, Gavagai, Sweden
Serge Sharoff, University of Leeds, UK
Anders Søgaard, University of Copenhagen, Denmark
Daniel Sonntag, German Research Center for AI, Germany
Diana McCarthy, Lexical Computing Ltd., UK
Dominic Widdows, Google, USA

Invited Speaker:

Dominic Widdows, Google, USA

Table of Contents

<i>(Linear) Maps of the Impossible: Capturing Semantic Anomalies in Distributional Space</i> Eva Maria Vecchi, Marco Baroni and Roberto Zamparelli	1
<i>Distributed Structures and Distributional Meaning</i> Fabio Massimo Zanzotto and Lorenzo Dell’ Arciprete	10
<i>Two Multivariate Generalizations of Pointwise Mutual Information</i> Tim Van de Cruys	16
<i>Distributional Semantics and Compositionality 2011: Shared Task Description and Results</i> Chris Biemann and Eugenie Giesbrecht	21
<i>Shared Task System Description: Frustratingly Hard Compositionality Prediction</i> Anders Johannsen, Hector Martinez, Christian Rishøj and Anders Søgaard	29
<i>Identifying Collocations to Measure Compositionality: Shared Task System Description</i> Ted Pedersen	33
<i>Shared Task System Description: Measuring the Compositionality of Bigrams using Statistical Methodologies</i> Tanmoy Chakraborty, Santanu Pal, Tapabrata Mondal, Tanik Saikh and Sivaju Bandyopadhyay	38
<i>Detecting Compositionality Using Semantic Vector Space Models Based on Syntactic Context. Shared Task System Description</i> Guillermo Garrido and Anselmo Peñas	43
<i>Measuring the Compositionality of Collocations via Word Co-occurrence Vectors: Shared Task System Description</i> Alfredo Maldonado-Guerra and Martin Emms	48
<i>Exemplar-Based Word-Space Model for Compositionality Detection: Shared Task System Description</i> Siva Reddy, Diana McCarthy, Suresh Manandhar and Spandana Gella	54

Conference Program

Friday June 24, 2011

- 9:20–9:30 Opening
- 09:30–10:30 Invited Talk by Dominic Widdows
- 10:30–11:00 Morning break
- 11:00–11:40 *(Linear) Maps of the Impossible: Capturing Semantic Anomalies in Distributional Space*
Eva Maria Vecchi, Marco Baroni and Roberto Zamparelli
- 11:40–12:05 *Distributed Structures and Distributional Meaning*
Fabio Massimo Zanzotto and Lorenzo Dell’Arciprete
- 12:05–12:30 *Two Multivariate Generalizations of Pointwise Mutual Information*
Tim Van de Cruys
- 12:30–14:00 Lunch break
- 14:00–14:30 *Distributional Semantics and Compositionality 2011: Shared Task Description and Results*
Chris Biemann and Eugenie Giesbrecht
- 14:30–14:50 *Shared Task System Description: Frustratingly Hard Compositionality Prediction*
Anders Johannsen, Hector Martinez, Christian Rishøj and Anders Søgaard
- 14:50–15:10 *Identifying Collocations to Measure Compositionality: Shared Task System Description*
Ted Pedersen
- 15:10–15:30 *Shared Task System Description: Measuring the Compositionality of Bigrams using Statistical Methodologies*
Tanmoy Chakraborty, Santanu Pal, Tapabrata Mondal, Tanik Saikh and Sivaju Bandyopadhyay
- 15:30–16:00 Afternoon break
- 16:00–16:20 *Detecting Compositionality Using Semantic Vector Space Models Based on Syntactic Context. Shared Task System Description*
Guillermo Garrido and Anselmo Peñas

Friday June 24, 2011 (continued)

- 16:20–16:40 *Measuring the Compositionality of Collocations via Word Co-occurrence Vectors: Shared Task System Description*
Alfredo Maldonado-Guerra and Martin Emms
- 16:40–17:00 *Exemplar-Based Word-Space Model for Compositionality Detection: Shared Task System Description*
Siva Reddy, Diana McCarthy, Suresh Manandhar and Spandana Gella
- 17:00–17:30 Wrap-Up Discussion