

A WORKSHOP OF THE 2015 CONFERENCE ON  
EMPIRICAL METHODS IN NATURAL LANGUAGE PROCESSING  
(EMNLP 2015)

**The Workshop on  
Vision and Language 2015  
(VL'15)**

VISION AND LANGUAGE MEET COGNITIVE SYSTEMS

Proceedings

September 18, 2015  
Lisbon, Portugal

This workshop is supported by ICT COST Action IC1307, the European Network on Integrating Vision and Language (iV&L Net): Combining Computer Vision and Language Processing For Advanced Search, Retrieval, Annotation and Description of Visual Data.

<http://ivl-net.eu/>



COST is supported by the EU Framework Programme Horizon 2020



©2015 The Association for Computational Linguistics  
ISBN: 978-1-941643-32-7

## Preface

The Workshop on Vision and Language 2015 (VL'15) took place in the beautiful city of Lisbon, Portugal on September 18th 2015, as part of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP 2015). The workshop was organized by the new European Network on Integrating Vision and Language, an initiative funded as a European COST Action under the Horizon 2020 programme supported by the European Commission.

The 2015 edition of the VL workshop is a successful continuation of the previous VL editions, where the VL workshops have the following general aims:

1. to provide a venue for reporting and discussing planned, ongoing and completed research that involves both language and vision; and
2. to enable NLP and computer vision researchers to meet, exchange ideas, expertise and technology, and form new research partnerships.

The flagship workshop's main purpose is to establish a strong inter-disciplinary forum which will ignite fertilizing discussions and ideas on how to combine and integrate established and novel techniques from different (but related) fields into new unified modeling approaches, as well as how to approach the problem of multi-modal data processing for NLP and vision from a completely new angle.

The call for papers for VL'15 soliciting both full research papers and short abstracts was issued in May 2015 and elicited a good number of high-quality submissions (23 in total), each of which was peer-reviewed by three members of the program committee. The interest in the workshop from leading NLP and computer vision researchers and the quality of submissions was high, so we aimed to be as inclusive as possible within the practical constraints of the workshop. In the end we accepted 13 full research papers, and 5 short abstracts.

The resulting workshop program packed a lot of exciting and diverse content into one day. We were delighted to be able to include in the program two great keynote speakers: Krystian Mikolajczyk and Marco Baroni. Our technical program combined 7 oral papers, and 11 poster presentations accompanied by short 5-minute poster spotlights.

The program also included a discussion session on future directions for the VL community and workshops, including plans for shared task competitions, summer schools, and expansions towards other related fields and research communities (e.g., information retrieval, data mining, digital humanities, Web search, cognitive science).

We would like to thank all the people who contributed to the organization and delivery of this workshop: the authors who submitted such high-quality papers; the program committee for their prompt and effective reviewing; our keynote speakers; the EMNLP 2015 organising committee, especially the workshops chairs, Zornitsa Kozareva and Jörg Tiedemann, and the publication chairs, Yuval Marton and Daniele Pighin; the participants in the workshop; and future readers of these proceedings for your shared interest in this exciting new area of research.

*September 2015*

*VL'15 Organizers*



**Organizing Committee:**

Anja Belz, *University of Brighton, UK*  
Luísa Coheur, *INESC-ID, Portugal*  
Vittorio Ferrari, *University of Edinburgh, UK*  
Marie-Francine Moens, *KU Leuven, Belgium*  
Katerina Pastra, *CSRI, Greece*  
Ivan Vulić, *KU Leuven, Belgium*

**Program Committee:**

Ahmet Aker, *University of Sheffield, UK*  
Yiannis Aloimonos, *University of Maryland, USA*  
Marco Baroni, *University of Trento, Italy*  
Raffaella Bernardi, *University of Trento, Italy*  
Gemma Boleda, *University Pompeu Fabra, Spain*  
Antoine Bordes, *Facebook Inc., USA*  
Léon Bottou, *Microsoft Research, USA*  
Elia Bruni, *Free University of Bolzano, Italy*  
Yejin Choi, *University of Washington, USA*  
Darren Cosker, *University of Bath, UK*  
Simon Dobnik, *University of Gothenburg, Sweden*  
Desmond Elliott, *CWI Amsterdam, The Netherlands*  
Erkut Erdem, *Hacettepe University, Turkey*  
Sergio Escalera, *Autonomous University of Barcelona, Spain*  
Michel Galley, *Microsoft Research, USA*  
Kristen Grauman, *University of Texas at Austin, USA*  
Lewis Griffin, *University College London, UK*  
Julia Hockenmaier, *University of Illinois at Urbana-Champaign, USA*  
Yangqing Jia, *Google, USA*  
Henry Kautz, *University of Rochester, USA*  
Frank Keller, *University of Edinburgh, UK*  
Douwe Kiela, *University of Cambridge, UK*  
Polina Kuznetsova, *Stony Brook University, USA*  
Pierre Lison, *University of Oslo, Norway*  
Rebecca Mason, *Brown University, USA*  
Cynthia Matuszek, *University of Maryland, USA*  
John Philip McCrae, *University of Bielefeld, Germany*  
Florian Metze, *Carnegie Mellon University, USA*  
Rada Mihalcea, *University of Michigan, USA*  
Margaret Mitchell, *Microsoft Research, USA*  
Ray Mooney, *University of Texas at Austin, USA*

Vicente Ordonez, *University of North Carolina at Chapel Hill, USA*  
Simone Paolo Ponzetto, *University of Mannheim, Germany*  
Kate Saenko, *UMass Lowell, USA*  
Carina Silberer, *University of Edinburgh, UK*  
Alan Smeaton, *Dublin City University, Ireland*  
Richard Socher, *MetaMind, USA*  
Richard Sproat, *Google, USA*  
Stefanie Tellex, *Brown University, USA*  
Isabel Trancoso, *INESC-ID, Portugal*  
Lucy Vanderwende, *Microsoft Research, USA*

**Invited Speakers:**

Marco Baroni, *University of Trento, Italy*  
Krystian Mikolajczyk, *University of Surrey, UK*

## Table of Contents

<i>Visually-Verifiable Textual Entailment: A Challenge Task for Combining Language and Vision</i> Jayant Krishnamurthy .....	1
<i>Computational Integration of Human Vision and Natural Language through Bitext Alignment</i> Preethi Vaidyanathan, Emily Prud'hommeaux, Cecilia O. Alm, Jeff B. Pelz and Anne R. Haake .	4
<i>Towards Reliable Automatic Multimodal Content Analysis</i> Olli Philippe Lautenbacher, Liisa Tiittula, Maija Hirvonen, Jorma Laaksonen and Mikko Kurimo	6
<i>Linguistic Analysis of Multi-Modal Recurrent Neural Networks</i> Ákos Kádár, Grzegorz Chrupała and Afra Alishahi .....	8
<i>Defining Visually Descriptive Language</i> Robert Gaizauskas, Josiah Wang and Arnau Ramisa .....	10
<i>Semantic Tuples for Evaluation of Image to Sentence Generation</i> Lily D. Ellebracht, Arnau Ramisa, Pranava Swaroop Madhyastha, Jose Cordero-Rama, Francesc Moreno-Noguer and Ariadna Quattoni .....	18
<i>Image Representations and New Domains in Neural Image Captioning</i> Jack Hessel, Nicolas Savva and Michael Wilber .....	29
<i>Image with a Message: Towards Detecting Non-Literal Image Usages by Visual Linking</i> Lydia Weiland, Laura Dietz and Simone Paolo Ponzetto .....	40
<i>Visual Classifier Prediction by Distributional Semantic Embedding of Text Descriptions</i> Mohamed Elhoseiny and Ahmed Elgammal .....	48
<i>Understanding Urban Land Use through the Visualization of Points of Interest</i> Evgheni Polisciuc, Ana Alves and Penousal Machado .....	51
<i>Comparing Attribute Classifiers for Interactive Language Grounding</i> Yanchao Yu, Arash Eshghi and Oliver Lemon .....	60
<i>Generating Semantically Precise Scene Graphs from Textual Descriptions for Improved Image Retrieval</i> Sebastian Schuster, Ranjay Krishna, Angel Chang, Li Fei-Fei and Christopher D. Manning . . . .	70
<i>Do Distributed Semantic Models Dream of Electric Sheep? Visualizing Word Representations through Image Synthesis</i> Angeliki Lazaridou, Dat Tien Nguyen and Marco Baroni .....	81
<i>A Weighted Combination of Text and Image Classifiers for User Gender Inference</i> Tomoki Taniguchi, Shigeyuki Sakaki, Ryosuke Shigenaka, Yukihiro Tsuboshita and Tomoko Ohkuma	87
<i>Coupling Natural Language Processing and Animation Synthesis in Portuguese Sign Language Translation</i> Inês Almeida, Luísa Coheur and Sara Candeias .....	94
<i>Describing Spatial Relationships between Objects in Images in English and French</i> Anja Belz, Adrian Muscat, Maxime Aberton and Sami Benjelloun .....	104





# Conference Program

**Friday, September 18, 2015**

**08:45–10:30 Session 1**

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

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

**10:00–10:30 Poster Spotlights 1**

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

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

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

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

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

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

**10:30–11:00 Coffee Break**

**Friday, September 18, 2015 (continued)**

**11:00–12:15 Session 2**

11:00–11:25 *Image Representations and New Domains in Neural Image Captioning*

Jack Hessel, Nicolas Savva and Michael Wilber

11:25–11:50 *Image with a Message: Towards Detecting Non-Literal Image Usages by Visual Linking*

Lydia Weiland, Laura Dietz and Simone Paolo Ponzetto

**11:50–12:15 Poster Spotlights 2**

11:50–11:55 *Visual Classifier Prediction by Distributional Semantic Embedding of Text Descriptions*

Mohamed Elhoseiny and Ahmed Elgammal

11:55–12:00 *Understanding Urban Land Use through the Visualization of Points of Interest*

Evgheni Polisciuc, Ana Alves and Penousal Machado

12:00–12:05 *Comparing Attribute Classifiers for Interactive Language Grounding*

Yanchao Yu, Arash Eshghi and Oliver Lemon

12:05–12:10 *Combining Geometric, Textual and Visual Features for Generating Prepositions in Image Descriptions (to appear in EMNLP 2015)*

Arnau Ramisa, Josiah Wang, Ying Lu, Emmanuel Dellandrea, Francesc Moreno-Noguer and Robert Gaizauskas

12:10–12:15 *From the Virtual to the Real World: Referring to Visible Objects Under Uncertainty in Real-World Spatial Scenes (to appear in EMNLP 2015)*

Dimitra Gkatzia and Verena Rieser

**12:15–14:00 Poster Session and Lunch**

**Friday, September 18, 2015 (continued)**

**14:00–15:30 Session 3**

14:00–15:00 *Invited Talk 2: The ImageCLEF 2015 Task on Scalable Image Annotation, Localization and Sentence Generation*  
Krystian Mikolajczyk

15:00–15:30 *Generating Semantically Precise Scene Graphs from Textual Descriptions for Improved Image Retrieval*  
Sebastian Schuster, Ranjay Krishna, Angel Chang, Li Fei-Fei and Christopher D. Manning

**15:30–16:00 Coffee Break**

**16:00–18:00 Session 4**

16:00–16:25 *Do Distributed Semantic Models Dream of Electric Sheep? Visualizing Word Representations through Image Synthesis*  
Angeliki Lazaridou, Dat Tien Nguyen and Marco Baroni

16:25–16:50 *A Weighted Combination of Text and Image Classifiers for User Gender Inference*  
Tomoki Taniguchi, Shigeyuki Sakaki, Ryosuke Shigenaka, Yukihiro Tsuboshita and Tomoko Ohkuma

16:50–17:15 *Coupling Natural Language Processing and Animation Synthesis in Portuguese Sign Language Translation*  
Inês Almeida, Luísa Coheur and Sara Candeias

17:15–17:40 *Describing Spatial Relationships between Objects in Images in English and French*  
Anja Belz, Adrian Muscat, Maxime Aberton and Sami Benjelloun

**Friday, September 18, 2015 (continued)**

**17:40–18:00 Closing Remarks and Discussion**