



LREC 2022 Workshop
Language Resources and Evaluation Conference
20-25 June 2022

**People in Vision, Language, and the Mind
P-VLAM**

PROCEEDINGS

Editors:
Patrizia Paggio, Albert Gatt, Marc Tanti

Proceedings of the LREC 2022 workshop on People in Vision, Language, and the Mind (P-VLAM 2022)

Edited by:

Patrizia Paggio, Albert Gatt, Marc Tanti

ISBN: 979-10-95546-79-5

EAN: 9791095546795

For more information:

European Language Resources Association (ELRA)

9 rue des Cordelières

75013, Paris

France

<http://www.elra.info>

Email: lrec@elda.org



© European Language Resources Association (ELRA)

These workshop proceedings are licensed under a Creative Commons
Attribution-NonCommercial 4.0 International License

Message from the Program Chairs

This volume documents the proceedings of the second workshop on People in Vision, Language, and the Mind (formerly ONION 2020), held on June 2022 in Marseille, France, as part of the LREC 2022 conference (International Conference on Language Resources and Evaluation). This workshop focuses on how people, their bodies and faces as well as mental states are described in text with associated images, and modelled in computational and cognitive terms. Our goal is to build bridges between researchers from the cognitive science, natural language processing, and vision communities who have an interest in the representation of people. We have accepted six papers this year, three short and three long, with topics varying from automatically generating descriptions of human faces to the ambiguity of emotions to the meaning of nods. We hope that future P-VLAM workshops will continue to have such a variety of topics in this interesting area of research.

Organizers

Patrizia Paggio – University of Copenhagen and University of Malta
Albert Gatt – Utrecht University
Marc Tanti – University of Malta

Program Committee:

Manex Aguirrezabal, CST, University of Copenhagen
Francesca D'Errico, Roma Tre University
Diego Frassinelli, University of Konstanz
Jordi Gonzalez, Universitat Autònoma de Barcelona
David Hogg, University of Leeds
Christer Johansson, University of Bergen
Kristiina Jokinen, National Institute of Advanced Industrial Science and Technology (AIST), Japan
Roman Klinger, University of Stuttgart
Adrian Muscat, University of Malta
Costanza Navarretta, CST, University of Copenhagen
Catherine Pelachaud, Institute for Intelligent Systems and Robotics, UPMC and CNRS
Isabella Poggi, Roma Tre University

Table of Contents

<i>Exploring the GLIDE model for Human Action Effect Prediction</i> Fangjun Li, David C. Hogg and Anthony G. Cohn	1
<i>Do Multimodal Emotion Recognition Models Tackle Ambiguity?</i> Hélène Tran, Issam Falih, Xavier Goblet and Engelbert Mephu Nguifo	6
<i>Development of a MultiModal Annotation Framework and Dataset for Deep Video Understanding</i> Erika Loc, Keith Curtis, George Awad, Shahzad Rajput and Ian Soboroff	12
<i>Cognitive States and Types of Nods</i> Taiga Mori, Kristiina Jokinen and Yasuharu Den	17
<i>Examining the Effects of Language-and-Vision Data Augmentation for Generation of Descriptions of Human Faces</i> Nikolai Ilinykh, Rafal Černiavski, Eva Elžbieta Sventickaitė, Viktorija Buzaitė and Simon Dobnik	26
<i>Face2Text revisited: Improved data set and baseline results</i> Marc Tanti, Shaun Abdilla, Adrian Muscat, Claudia Borg, Reuben A. Farrugia and Albert Gatt .	41

Conference Program

09:00–09:15 *Welcome*

09:15–10:10 *Keynote - Your face says it all: Automated analysis and synthesis of facial actions*
İtir Önal Ertuğrul

10:10–10:30 *Exploring the GLIDE model for Human Action Effect Prediction*
Fangjun Li, David C. Hogg and Anthony G. Cohn

10:30–11:00 *Coffee break*

11:00–11:20 *Do Multimodal Emotion Recognition Models Tackle Ambiguity?*
Hélène Tran, Issam Falih, Xavier Goblet and Engelbert Mephu Nguifo

11:20–11:40 *Development of a MultiModal Annotation Framework and Dataset for Deep Video Understanding*
Erika Loc, Keith Curtis, George Awad, Shahzad Rajput and Ian Soboroff

11:40–12:00 *Cognitive States and Types of Nods*
Taiga Mori, Kristiina Jokinen and Yasuharu Den

12:00–12:20 *Examining the Effects of Language-and-Vision Data Augmentation for Generation of Descriptions of Human Faces*
Nikolai Ilinykh, Rafal Černiavski, Eva Elžbieta Sventickaitė, Viktorija Buzaitė and Simon Dobnik

12:20–12:40 *Face2Text revisited: Improved data set and baseline results*
Marc Tanti, Shaun Abdilla, Adrian Muscat, Claudia Borg, Reuben A. Farrugia and Albert Gatt

12:40–13:00 *Closing*