LREC 2020 Workshop
Language Resources and Evaluation Conference
11–16 May 2020

**People in language, vision and the mind
(ONION 2020)**

# PROCEEDINGS

Editors: Patrizia Paggio, Albert Gatt and Roman Klinger

# Proceedings of the LREC 2020 Workshop on People in language, vision and the mind (ONION 2020)

Edited by: Patrizia Paggio, Albert Gatt and Roman Klinger

# Introduction

## 1  Motivation and Aims of the Workshop

The ability to adequately model and describe people in terms of their body and face is interesting for a variety of language technology applications, e.g., conversational agents and interactive narrative generation, as well as forensic applications in which people need to be identified or their images generated from textual or spoken descriptions. Such systems need resources and models where images associated with human bodies and faces are coupled with linguistic descriptions. Thus, the research needed to develop such datasets and models is placed at the interface between vision and language research, a cross-disciplinary area which has received considerable attention in recent years, e.g., through the activities of the European Network on Integrating Vision and Language (iV&L Net), the 2015–2018 Language and Vision Workshops, the 2018–2019 Workshops on Shortcomings in Vision and Language and the ongoing Multi-Task, Multilingual, Multimodal (Multi3Generation) Generation COST Action.

The aim of this first edition of the ONION workshop was to provide a forum to present and discuss current research focusing on multimodal resources as well as computational and cognitive models aiming to describe people in terms of their bodies and faces, including their affective state as it is reflected physically. Such models might either generate textual descriptions of people, generate images corresponding to descriptions of people, or in general exploit multimodal representations for different purposes and applications. Knowledge of the way human bodies and faces are perceived, understood and described by humans is key to the creation of such resources and models, therefore the workshop also invited contributions where the human body and face are studied from a cognitive, neurocognitive or multimodal communication perspective.

Recent research on the analysis of images and text or the generation of image descriptions focused on datasets which might contain people as a subset; however, we argue that such general multimodal resources are not adequate for the specific challenges posed by applications based on the modelling of human bodies and faces. Descriptions of people are frequent in human communication, for example when one seeks to identify an individual or distinguish one person from another, or in the course of conveying a person's affective state on the basis of facial expression, posture etc. These descriptions are also pervasive in descriptive or narrative text. Depending on the context, they may focus on physical attributes, or incorporate inferred characteristics and emotional elements.

Human body postures and faces are being studied by researchers from different research communities, including those working with vision and language modeling, natural language generation, cognitive science, cognitive psychology, multimodal communication and embodied conversational agents. The workshop aimed to reach out to all these communities to explore the many different aspects of research on the human body and face, including the resources that such research needs, and to foster cross-disciplinary synergy.

## 2  Contributions

Five papers were accepted for the workshop and are included in this publication. Although this is a small collection, it reflects well the cross-disciplinary nature of the research area which the workshop is targeting.

The paper by Lembke, Folgerø, Andresen and Johansson presents the results of an experimental study which investigates the effect of prototypicality and self-recognition in participants' perception of the attractiveness of facial images. The stimuli used are depictions of Christ that were adapted into more human, gender-specific images and then morphed with individual photos. The results of the study have general implications for the psychological perception of faces that go beyond the study of Christian iconography.

Moving from the study of facial images to analyses of gestural behaviours, the study by Mori, Jokinen and Den deals with the different ways in which hand gestures, head movements and body posture are used in human-robot interaction as opposed to human-human interaction, and also identifies interesting differences between English and Japanese users in the way they use gestural behaviour when interacting with robots.

The paper by Paggio, Agirrezabal, Jongejan and Navarretta reports state-of-the-art results from machine learning classification experiments aimed at the automatic detection of different types of head movement from video-recorded face-to-face dialogues involving twelve different speakers. A number of models are trained using a combination of visual, acoustic and word features in a leave-one-out cross-validation scenario where classifiers are repeatedly trained on data from eleven speakers and tested on the remaining one.

The contribution by Anastasiou, Afkari and Maquil reports the results of a user study on collaborative problem-solving using an interactive tabletop. The focus of the authors is on the role of pointing gestures in low awareness situations, i.e., situations in which a user involved in a task might employ exaggerated manual actions to draw attention and thus raise awareness. The paper argues that the way in which a problem-solving scenario is designed has an effect on the type and frequency of occurring gestures.

Finally, the paper by Schlör, Zehe, Kobs, Veseli, Westermeier, Brübach, Roth, Latoschik and Hotho presents a novel approach to the automatic classification of sentiment in text relying on physiological signals to improve the performance of lexicon-based sentiment classifiers. The physiological signals considered are the heart rate and brain activity of readers recorded while they read short texts that have been annotated with sentiment labels. In addition to reporting the results of the sentiment analysis experiments, the authors make available a dataset that includes sentiment annotations, as well as two types of biofeedback data, namely heart rate and EEG data.

As can be seen from this summary, the papers include a variety of topics, from image perception to the use of gestural behaviour in different scenarios; they employ a range of methods, from experimental analysis to machine learning; they investigate the potential of different kinds of signal from visual and acoustic features to biofeedback data. In conclusion, they touch upon many different aspects of an area of research which we hope future editions of the ONION workshop will contribute to showcase and develop even further.

## 3   Online presentations

Unfortunately the workshop, which was originally planned to take place on 16 May 2020 in conjunction with the LREC 2020 conference, could not be held as a face-to-face meeting due to the ongoing coronavirus pandemic. Therefore, authors were asked to produce online presentations of the papers. All the presentations will be available from the workshop website at `https://onion2020.github.io/`.

**Organizers:**

Patrizia Paggio, University of Copenhagen, Denmark and University of Malta, Malta
Albert Gatt, University of Malta, Malta
Roman Klinger, University of Stuttgart, Germany

**Program Committee:**

Adrian Muscat, University of Malta
Andreas Hotho, University of Würzburg
Andrew Hendrickson, University of Tilburg
Catherine Pelachaud, Institute for Intelligent Systems and Robotics, UPMC and CNRS
Costanza Navarretta, CST, University of Copenhagen
David Hogg, University of Leeds
Diego Frassinelli, University of Konstanz
Isabella Poggi, Roma Tre University
Jordi Gonzalez, Universitat Autònoma de Barcelona
Kristiina Jokinen, National Institute of Advanced Industrial Science and Technology (AIST)
Mihael Arcan, National University of Ireland, Galway
Raffaella Bernardi, CiMEC Trento
Sebastian Padó, University of Stuttgart

# Table of Contents

# Workshop Program