

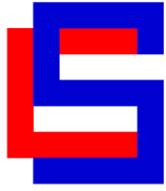
(Dis)embodiment 2022

Proceedings of the 2022 CLASP Conference on (Dis)embodiment

Simon Dobnik, Julian Grove and Asad Sayeed (eds.)



**Gothenburg and online
15–16 September 2022**



CLASP centre for
linguistic theory
and studies in probability

CLASP Papers in Computational Linguistics, Volume 4
©2022 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 978-1-955917-67-4
ISSN 2002-9764

Introduction

This volume contains the papers presented at the CLASP Conference on (Dis)embodiment at the Department of Philosophy, Linguistics and Theory of Science (FLoV), University of Gothenburg, held on September 15–16, 2022.

(Dis)embodiment brings together researchers from several areas examining the role of grounding and embodiment in modelling human language and behaviour – or limits thereof. The conference covers areas such as machine learning, computational linguistics, theoretical linguistics and philosophy, cognitive science and psycholinguistics, as well as artificial intelligence ethics and policy.

Papers were invited on topics from these and closely related areas, including (but not limited to) large-scale neural language modelling, both text-only and multimodal; training corpus and test task development; visual, dialogue and multi-modal inference systems; neurolinguistic and psycholinguistic experimental approaches to human language processing; philosophical discussions of linguistic groundedness and embodiment (or limits thereof) as it pertains to computational modelling; semantics and pragmatics in neural models; dialogue modelling and linguistic interaction; formal and theoretical approaches to language production and comprehension; statistical, machine learning and information theoretic approaches that either avoid or embrace groundedness and/or embodiment; methodologies and practices for annotating dialogue and multi-modal datasets; visual, dialogue and multi-modal generation; text generation in both the dialogue and monologue settings; multimodal and grounded approaches to computing meaning; semantics-pragmatics interface; social and ethical implications of the development and application of neural language models, as well as relevant policy implications and debates.

This conference aims to initiate a genuine discussion between these related topics and to examine different approaches and how they can inform each other. It features 3 invited talks by leading researchers, 9 peer-reviewed archival papers and 7 non-archival presentations.

We would like to thank all our contributors and programme committee members, with special thanks to CLASP for organising the hybrid conference and our sponsors SIGSEM <http://sigsem.org>, the ACL special interest group on semantics, and the Swedish Research Council for funding CLASP.

Simon Dobnik, Julian Grove and Asad Sayeed

Gothenburg

September 2022

Programme Committee:

Kathrein Abu Kwaik	University of Gothenburg
Afra Alishahi	Tillburg University
Alexander Berman	University of Gothenburg
Raffaella Bernardi	University of Trento
Jean-Philippe Bernardy	University of Gothenburg
Yonatan Bisk	Carnegie Mellon University
Ellen Breitholtz	University of Gothenburg
Harry Bunt	Tillburg University
Stergios Chatzikyriakidis	University of Crete
Alexander Clark	University of Gothenburg
Robin Cooper	University of Gothenburg
Ryan Cotterell	Eidgenössische Technische Hochschule Zürich
Simon Dobnik	University of Gothenburg
Markus Egg	Humboldt University
Adam Ek	University of Gothenburg
Katrin Erk	University of Texas at Austin
Chris Fox	University of Gothenburg
Jonathan Ginzburg	Université Paris-Diderot
Dimitra Gkatzia	Edinburgh Napier University
Eleni Gregoromichelaki	University of Gothenburg
Julian Grove	University of Gothenburg
Xudong Hong	Saarland University
Christine Howes	University of Gothenburg
Nikolai Ilinykh	University of Gothenburg
Cassandra Jacobs	State University of New York at Buffalo
Elisabetta Jezek	University of Pavia
Richard Johansson	Chalmers Technical University
Aram Karimi	University of Gothenburg
Ruth Kempson	King's College London
Nikhil Krishnaswamy	Colorado State University
Shalom Lappin	University of Gothenburg
Staffan Larsson	University of Gothenburg
Tal Linzen	New York University
Sharid Loáiciga	University of Gothenburg
Vladislav Maraev	University of Gothenburg
Yuval Marton	University of Washington
Elin McCreedy	Aoyama Gakuin University
Paul McKevitt	Ulster University
Louise McNally	Universitat Pompeu Fabra
Gregory Mills	University of Groningen
Joakim Nivre	Uppsala University
Bill Noble	University of Gothenburg
Manfred Pinkal	Saarland University
Violaine Prince	Université de Montpellier 2
James Pustejovsky	Brandeis University
Christian Retoré	Université de Montpellier
German Rigau	University of the Basque Country

Hannah Rohde	University of Edinburgh
David Schlangen	University of Potsdam
William Schuler	The Ohio State University
Sabine Schulte im Walde	University of Stuttgart
Gabriel Skantze	KTH Royal Institute of Technology
Vidya Somashekarappa	University of Gothenburg
Tim Van de Cruys	KU Leuven
Marten van Schijndel	Cornell University
Eva Maria Vecchi	University of Stuttgart
Carl Vogel	Trinity College Dublin
Alessandra Zarcone	Augsburg University of Applied Sciences
Sina Zarrieß	University of Bielefeld

Invited Speakers:

Afra Alishahi, Tilburg University
Felix Hill, DeepMind
Magnus Sahlgren, AI Sweden

Invited talk 1: Afra Alishahi

Getting closer to reality: Grounding and interaction in models of human language acquisition

Humans learn to understand speech from weak and noisy supervision: they manage to extract structure and meaning from speech by simply being exposed to utterances situated and grounded in their daily sensory experience. Emulating this remarkable skill has been the goal of numerous studies; however researchers have often used severely simplified settings where either the language input or the extralinguistic sensory input, or both, are small-scale and symbolically represented. I present a series of studies on modelling visually grounded language understanding.

Invited talk 2: Felix Hill

Three studies that show that artificial models of general intelligence learn better with language

Having and using language makes humans as a species better learners and better able to solve hard problems. I'll present three studies that demonstrate how this is also the case for artificial models of general intelligence. In the first, I show that agents with access to visual and linguistic semantic knowledge explore their environment more effectively than non-linguistic agents, enabling them to learn more about the world around them. In the second, I demonstrate how an agent embodied in a simulated 3D world can be enhanced by learning from explanations – answers to the question “why?” expressed in language. Agents that learn from explanations solve harder cognitive challenges than those trained from reinforcement learning alone, and can also better learn to make interventions in order to uncover the causal structure of their world. Finally, I'll present evidence that the skewed and bursty distribution of natural language may explain how large language models can be prompted to rapidly acquire new skills or behaviours. Together with other recent literature, this suggests that modelling language may make a neural network better able to acquire new cognitive capacities quickly, even when those capacities are not necessarily explicitly linguistic.

Invited talk 3: Magnus Sahlgren

The Singleton Fallacy: why current critiques of language models miss the point

There is currently a lively debate about the semantic (in)capabilities of current language models: do language models really understand language or are they simply stochastic parrots? Are we wasting our time in the pursuit of bigger and bigger models, and should we instead be climbing some other hill in the NLP landscape? This talk provides an overview over the different positions in the debate, and attempts to disentangle the debate by pointing out an argumentation error that is referred to as the singleton fallacy.

Table of Contents

<i>A Small but Informed and Diverse Model: The Case of the Multimodal GuessWhat!? Guessing Game</i> Claudio Greco, Alberto Testoni, Raffaella Bernardi and Stella Frank	1
<i>A Cross-lingual Comparison of Human and Model Relative Word Importance</i> Felix Morger, Stephanie Brandl, Lisa Beinborn and Nora Hollenstein	11
<i>Dispatcher: A Message-Passing Approach to Language Modelling</i> Alberto Cetoli	24
<i>In Search of Meaning and Its Representations for Computational Linguistics</i> Simon Dobnik, Robin Cooper, Adam Ek, Bill Noble, Staffan Larsson, Nikolai Ilinykh, Vladislav Maraev and Vidya Somashekarappa	30
<i>Can We Use Small Models to Investigate Multimodal Fusion Methods?</i> Lovisa Hagström, Tobias Norlund and Richard Johansson	45
<i>Embodied Interaction in Mental Health Consultations: Some Observations on Grounding and Repair</i> Jing Hui Law, Patrick Healey and Rosella Galindo Esparza	51
<i>Norm Participation Grounds Language</i> David Schlangen	62
<i>Where Am I and Where Should I Go? Grounding Positional and Directional Labels in a Disoriented Human Balancing Task</i> Sheikh Mannan and Nikhil Krishnaswamy	70
<i>From Speed to Car and Back: An Exploratory Study about Associations between Abstract Nouns and Images</i> Ludovica Cerini, Eliana Di Palma and Alessandro Lenci	80

