1 General information
Activity: Course
Title: Language, action and perception (APL-ESSLLI)
Area: Language and computation
Level: Introductory

2 Instructors
Simon Dobnik is an Associate Professor of Computational Linguistics and a member of the Centre for Language Technology (CLT) and the Centre for Linguistic Theory and Studies in Probability (CLASP), both at the University of Gothenburg. His research interests include spatial cognition, computational models of language and perception, human-robot interaction, situated spoken dialogue systems, and computational representations of meaning (semantics).

Address: Department of Philosophy, Linguistics and Theory of Science (FLOV), University of Gothenburg, Box 200, SE 405 30 Göteborg, Sweden • Phone: +46 31 786 6917

John D. Kelleher is a Professor in Computer Science at the Dublin Institute of Technology, where he is the Academic Leader of the ICE research institute. John is also the head of the ADAPT Research Centre at DIT which is funded by the Science Foundation Ireland grant (RC/13/2106). John has published a number of books with MIT Press including an introduction to Data Science (Kelleher and Tierney, 2018), and a textbook on machine learning (Kelleher et al., 2015). John’s primary interest in the area of spatial cognition relates to developing cognitively inspired models that enable computational systems (such as robots) to ground spatial language in sensor data. In particular, John is interested in the role of perceptual phenomenon (such as visual attention, object occlusions and viewer perspective) on the semantics of spatial terms.

Address: Directorate of Research, Enterprise, & Innovation Services, Greenway Hub, DIT Grangegorman, Dublin Institute of Technology, Ireland • Phone: +353 1 402 4789

3 Course contents
3.1 Abstract
The course gives a survey of theory and practical computational implementations of how natural language interacts with the physical world through action and perception. We will look at topics such as semantic theories and computational approaches to modeling natural language, action and perception (grounding), situated dialogue systems, integrated robotic systems, grounding of language in action and perception, generation and interpretation of scene descriptions from images and videos, spatial cognition, and others.
As the course studies how humans structure and interact with the physical world and express it in language, it bridges into the domains of cognitive science, computer vision, robotics and therefore more broadly belongs to the field of cognitive artificial intelligence. Typical applications of computational models of language, action, and perception are image search and retrieval on the web, navigation systems that provide more natural, human-like instructions, and personal robots and situated conversational agents that interact with us in our home environment through language.

The learning outcomes of the course are based on covering 3 topics: (i) the relation between language and perception in human interaction, (ii) how language and perception is modelled with formal and computational models and methods and how these are integrated with different applications, and (iii) how research in the field is communicated scientifically.

3.2 Motivation and description
3.2.1 Motivation
Humans engage into two kinds of linguistic interaction: with their physical world that they are situated in and other agents with whom they engage in conversations. The ability to identify objects and actions and (spatial) relations between them is a part of basic survival strategies of living organisms. Language give humans the ability to refer to objects and actions and exchange this information with each other in an efficient way through language and dialogue. In order for the robotic situated agents to be useful, however, they must be able to engage in natural dialogue with humans about the environment in the same way. Consequently, the ability to process and use descriptions (such as “bring me the red book on the table beside my bed” or “put the cheese on the bottom shelf in the fridge”) will be crucial for these situated agents.

Computational modelling of language grounded in action and perception requires the integration of different sources of knowledge that affect the semantics of descriptions: scene geometry and object motion, knowledge about dynamic kinematic routines of objects, and language coordination with dialogue partners. Furthermore, because situated agents are located within dynamic linguistic and perceptual environments they need to continuously adapt their understanding and representations of the environment.

3.2.2 Topics
The course gives a survey of approaches in linguistics, psychology, computer science and robotics related to computational interpretation and generation of language that is grounded in action and perception and takes into account geometric, cognitive, functional, or embodiment criteria in modelling. Topics that are included are:

- multi-modal and grounded representations (Harnad, 1990; Roy, 2002, 2005; Matuszek et al., 2012; Kennington et al., 2014; Skantze and Al Moubayed, 2012; Ghanimifard and Dobnik, 2017),
- image classification (Lowe, 1999; Muja and Lowe, 2009; Lowe, 2004; Bradski and Kaehler, 2008; Krizhevsky et al., 2012; Bruni et al., 2014),
- generating referring expressions (Clark and Wilkes-Gibbs, 1986; Dale and Reiter, 1995; Van Deemter et al., 2005; Kelleher and Kruijff, 2005a; Viethen et al., 2011; Krahmer and van Deemter, 2011; Kelleher and Kruijff, 2006; Fernández, 2013; McMahan and Stone, 2015; Monro et al., 2017; Schütte et al., 2017),
- image and scene descriptions (Mitchell et al., 2013; Elliott and Keller, 2013; Karpathy and Fei-Fei, 2015; Xu et al., 2015; Lu et al., 2017),
- visual question answering (Malinowski et al., 2015; Andreas et al., 2016),
- route instructions (Tversky, 1993; Stoia et al., 2006; Cuayahuitl et al., 2010; Janarthanam et al., 2013; Mei et al., 2016),
- learning language with robots (Lauria et al., 2001, 2002; Kruijff et al., 2007; Zender et al., 2008; Dobnik, 2009),
- robotic systems (Quigley et al., 2009; Goebel, 2013; O’Kane, 2013),
- modelling action, perception and dialogue interaction (Steels and Baillie, 2003; Steels and Loetzsch, 2009; Skočaj et al., 2010, 2011; Lazaridou et al., 2016; Chai et al., 2016; Pustejovsky and
computational models of spatial descriptions: (Herskovits, 1986; Logan and Sadler, 1996; Regier, 1996; Mukerjee, 1998; Kelleher and Kruijff, 2005b; Kelleher and Costello, 2009; Dobnik and Åstbom, 2017; Coventry et al., 2005; Coventry and Garrod, 2005; Dobnik and Kelleher, 2013, 2014; Dobnik et al., 2018, 2015),

- attention and salience (Kelleher et al., 2005; Dobnik and Kelleher, 2016),
- and other.

3.2.3 Learning outcomes
On successful completion of the course the participants are able to:

Knowledge and understanding
- Define geometric and cognitive features required for modelling of language grounded in action and perception.
- Describe existing resources and approaches in this field: corpora, experimental methods and computational models.
- Read and understand a scientific article in this area.

Competence and skills
- Describe how geometric and cognitive features interact with each other.
- Explain how their interaction is captured in existing computational resources and approaches.
- Explain how current approaches model their interaction in computational models.
- Critically reflect upon scientific literature in this field.
- Apply the knowledge about language, action and perception to a new problem.

Judgement and approach
- Analyse the interaction of geometric and cognitive features.
- Critically discuss practical programming and experimental examples.
- Recognise and formulate problems related to modelling language, action and perception.
- Propose solutions for conducting experiments or implementing computational models related to language, action and perception.
- Assess the importance of different knowledge sources (geometric and cognitive features) for language, action and perception.
- Evaluate how and to what extent experiments or computational implementations capture human cognition and behaviour.

3.3 Schedule
- Day 1: Grounding language in action and perception
- Day 2: Learning language with robots
- Day 3: Generating image descriptions
- Day 4: Language and space
- Day 5: Referring to what it matters (attention)

3.4 Course pre-requisites
In order to follow the course, the participants should at least have experience with one or several of the following fields at undergraduate (BA) level:

- Formal semantics and pragmatics
- Experimental psychology
- Natural language processing
- Computational semantics
- Machine learning
- Robotics
- or equivalent skills and knowledge.
3.5 Literature
Please see the references below.

4 Practical information

4.1 Preceding meetings
A similar and longer version of the course has been given at the University of Gothenburg as a part of Masters in Language Technology (MLT) as LT2308 Embodied and situated language processing or LT2318 Artificial intelligence: Cognitive systems.

4.2 External funding
The instructors would be able to supplement any funding they receive from the ESSLLI organisers with their own funding.

References


Patrick Goebel. 2013. **ROS by example.** Lulu.


