Learning language with robots

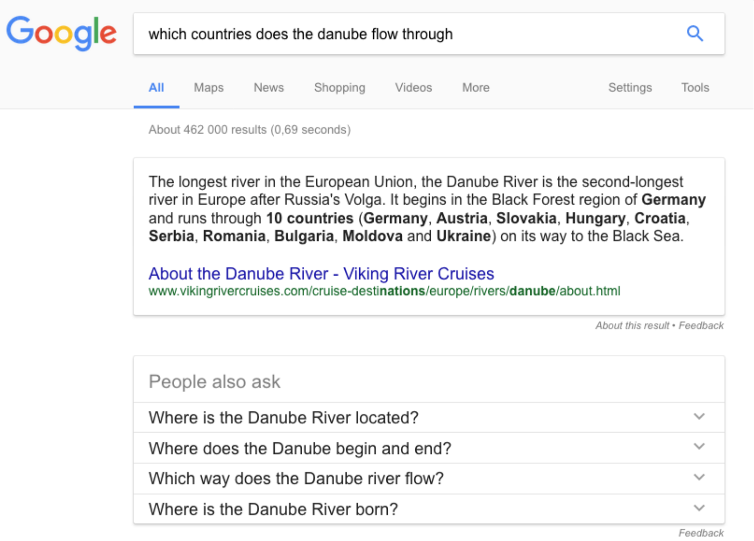
Simon Dobnik

31 European Summer School in Logic, Language and Information
12 August 2019, Riga Latvia

UNIVERSITY OF GOTHENBURG

CLASP centre for linguistic theory and studies in probability

Several conversational interfaces



Several conversational interfaces

UNIVERSITY OF GOTHENBURG

Why robotics?

- A robot that can make sense of the world and interact with humans is very useful: assistants to people with disabilities, robots on rescue missions, just for fun, etc.
- Spatial cognition and action represent the core of human cognition and behaviour.
- Having access to robot' sensors and actuators can give us a theoretical insight into language, spatial perception and action.

Properties of the robot's world

The nature of robot's world and language:

- Partially observable (sensory data is noisy and incomplete)
- Dynamic (changes over time)
- Continuous (real valued sensory data)
- Sequential (current decisions affect future actions)
- Contains other interacting agents
- Stochastic (outcomes of actions are non-deterministic)

Learning from environment

SLAM, (Newman and Durrant-Whyte, 2001)
http://www.youtube.com/watch?v=6afrMnEmXFI

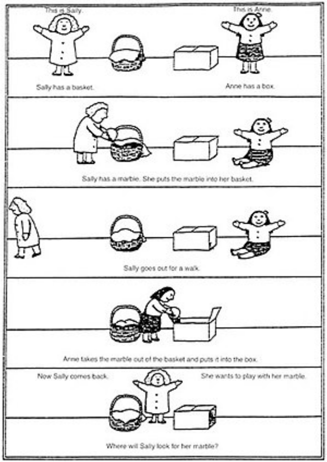
Embodiment

- Agents with different bodies (sensors and actuators) perceive an interact with the world differently.
- Consequently, they also structure the world differently: the representations they learn will be different ("embodied mind") (Maurice Merleau-Ponty and George Lakoff).
- Is human-robot communication possible at all?

Situatedness

- Human and robot are situated in the same environment which imposes identical constraints on both kinds of representations.
- They can also interact with each other: see each other, jointly attend to each other and refer to the same situations (socialness).
- Perhaps the fact that they may internally operate with different representations is not that important.

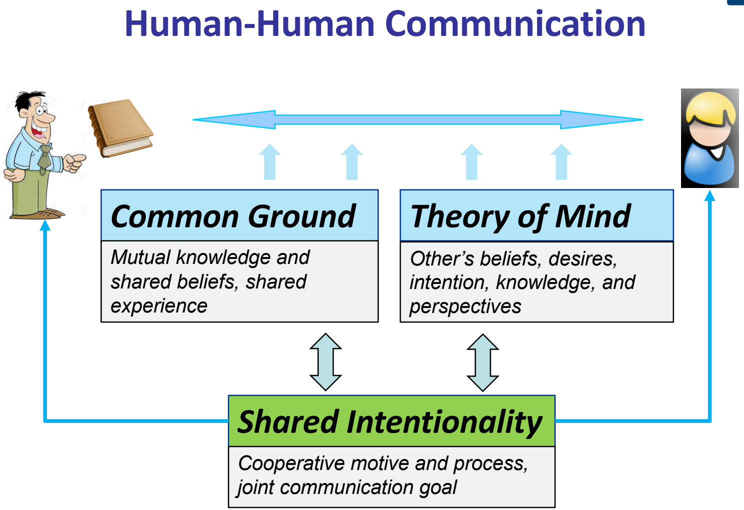
Theory of mind



(Baron-Cohen et al., 1985)

The beliefs, desires and intentions of the other

Theory of mind and common ground



Slide from (Chai, 2019)

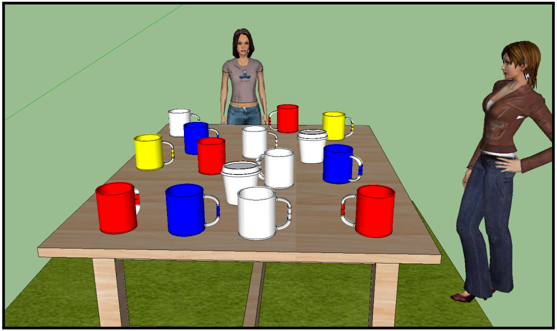
Learning from linguistic interaction

Referring as a collaborative process (Clark and Wilkes-Gibbs, 1986)

- Speakers and addressees work together in the making of a definite reference
- Speaker proposes/invites a noun phrase
- Participants iteratively repair, expand, replace the referring expression until they reach mutual agreement
- Minimise joint effort

Situated conversation #1

UNIVERSITY OF GOTHENBURG



(Dobnik, Howes, and Kelleher, 2015; Dobnik, Howes, Demaret, and Kelleher, 2016)

Situated conversation #1

UNIVERSITY OF GOTHENBURG

P2: 123: ok, so i see a red mug directly behind the red one on your left
P2: 124: probably next to the white with "funny top" that i cant see
P1: 125: it is just behind that and to my left/your right
P1: 126 : behind from my perspective
P2: 127 : and the red i can't see is it to the left of the yellow?
P1: 128 : yes, as you se it its left
P2: 129 : ok, i mark it, and you mark the other red
P1: 130 : yup
P1: 131 : and the blue ones are one on the second row from you, to the right from you
P1: 132 : one slightly to my left
P1: 133 : and one in front of katie in the first row
P2: 134 : yes, that's the same
P1: 135 : and the yellow are on between us to your far right
P1: 136 : and one quite close to the corner on your left and katies right?
P2: 137 : yes the same

CLASP centre for linguistic theory and studies in probability

Situated conversation #1 I

From SCARE corpus (Stoia et al., 2008), 2.txt, l.38

DG: SIL AND uh WHAT WE GOTTA DO IS MOVE THE PICTURE TO THE OTHER WALL SIL [pause]
DF: SIL WHAT'S OTHER
DG: I
DF: OPPOSITE
DG: D-
DF: [pause]
DG: I DON'T KNOW the DEFINITION OF OTHER [pause]
DF: SIL
DG: SIL UM
DF: OPPOSITE WALL
DG: SIL
DF: [pause]
DG: I WOULD [pause] SIL HOW MANY WALLS ARE THERE SIL [pause]
DF: SIL WELL IT'S A ROOM SO THERE ARE FOUR WALLS [pause]
DG: SIL WELL SIL [pause] SIL PUT IT ON THE OPPOSITE WALL SIL [pause]

Situated conversation #1 II

DF: SIL OK [pause] SIL CONTROL PICKS THE SIL [pause] SIL CONTROL'S SUPPOSED TO PICK THINGS UP AND [pause] SIL AM I SUPPOSED TO PICK UP THIS THING [pause]
DG: SIL I CAN SAY THAT SIL [pause] SIL I CAN SAY THAT NUMBER SIL [pause] SIL NO SIL [pause] SIL OH THAT'S WHERE I HAVE TO MOVE IT TO SIL [pause] SIL THAT'S WHERE I HAVE TO MOVE IT TO SIL [NOISE LAURA NO YOU CAN DESCRIBE THAT THIS BUTTON CONTROLS IT] SIL WELL THERE IS A BUTTON THAT CONTROLS IT BUT
DF: SIL
DG: OH
DF: CONTROLS WHAT
DG: SIL
DF: SIL [pause]
DG: NOW I UNDERSTAND

Expressing meaning with our body I

- Gestures and emotions
- Conversational resources
  - Non-verbal cues and information
  - ...but not any kind of movement and prosody.
- Help with coordination of conversation:
  - understanding and misunderstanding
  - turn-taking
  - topic progression
  - empathy
  - sarcasm
  - attitude
  - mood
  - ...

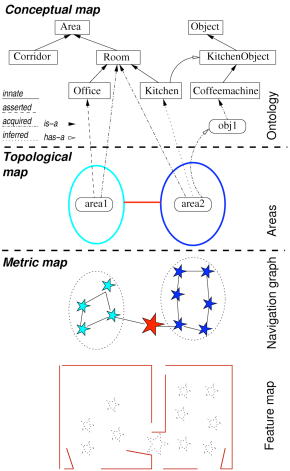
Expressing meaning with our body II

- Social referencing: film of Leonardo robot (Thomaz et al., 2005; Breazeal et al., 2006)

- Eye-gaze and multi-party dialogue: Furhat (Skantze, 2016)

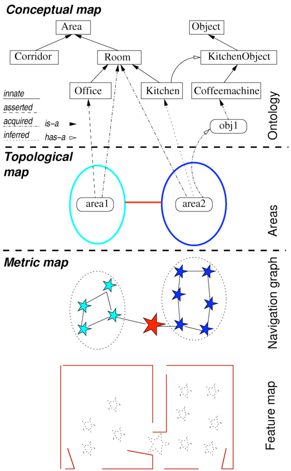
Building robotic systems

A layered approach (Kruijff et al., 2007; Zender et al., 2008)



Building robotic systems

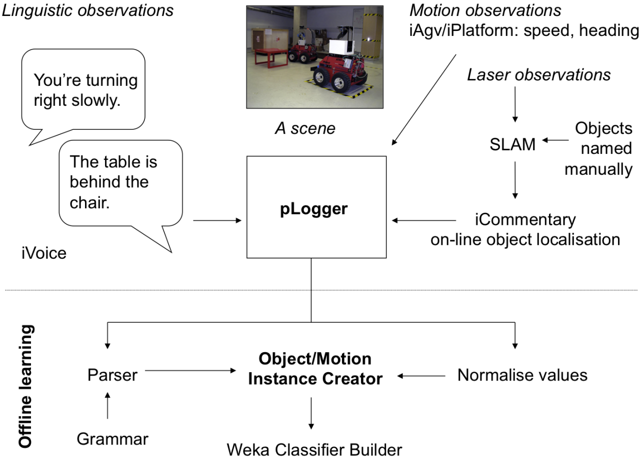
A layered approach (Kruijff et al., 2007; Zender et al., 2008)



- Integration of (independent) processes
- Information exchange and flow
- Temporal processing
- Information fusion
- Increased abstraction of representations

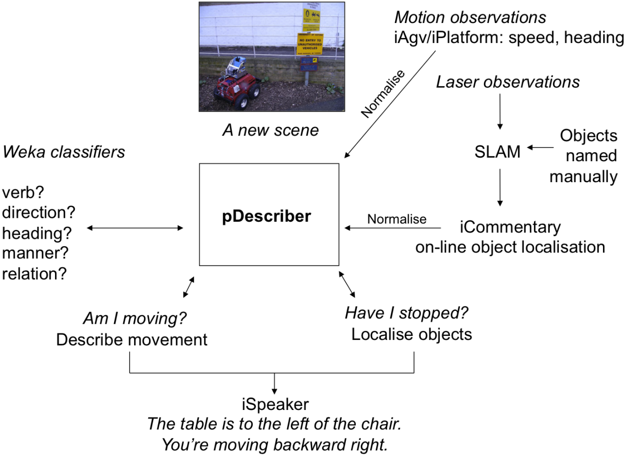
Building robotic systems, II

Data collection and offline learning (Dobnik, 2009)



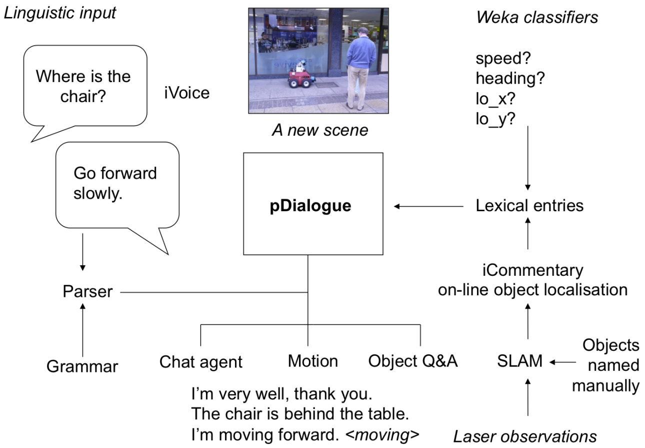
Building robotic systems, III

pDescriber



Building robotic systems, IV

pDialogue



Robot operating system (ROS)

- ROS: Robot Operating System
- A middle-ware that assists writing robotic applications on the top of OS
- Processes and information flow between them
- Portable: several robots supported, https://robots.ros.org
- Easy to program
- http://wiki.ros.org/ROS/Introduction

ROS basics

- Each task separate process (a ROS node)
- ROS nodes communicate over network directly with each other
- They either publish or subscribe to information (ROS topics/services)
  `/camera/rgb/image_color` of type `sensor_msgs/Image`
- ROS master (`roscore`) coordinates the communication between the nodes

(Quigley et al., 2009)

Tutorial #4: writing your ROS node

by Mehdi Ghanimifard

In this tutorial we will give you some quick guidance how to start programming with ROS.

Code on Github

KILLE: Kinect Is Learning LanguagE

- A framework for situated agents that learn grounded language incrementally and online from a human tutor
- Language technology:
  - A platform for building models of situated language learning
  - Apply existing LT tools in an interactive learning framework
- Linguistics and psychology:
  - a platform for testing theories, formalisms and applications on grounding language in interaction
- A proof-of-concept evaluation of tutor assisted learning

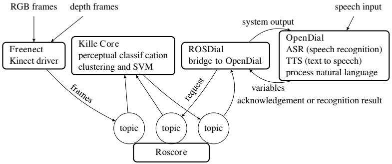
(de Graaf, 2016; Dobnik and de Graaf, 2017)

KILLE: Kinect Is Learning LanguagE

UNIVERSITY OF GOTHENBURG

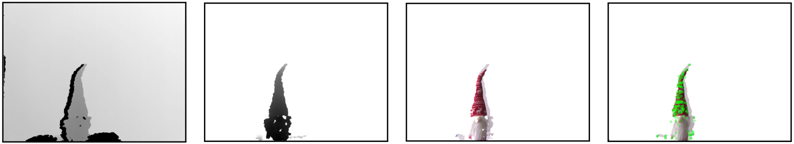
Hardware

Software



Recognising objects

"This is a gnome."



Dialogue games as learning strategies

Direct instruction: objects

U: This is a cup.

S: [Object is focused on and observed.]

S: OK, I learned a cup.

Dialogue games as learning strategies

UNIVERSITY OF GOTHENBURG

Direct instruction: objects

U: This is a cup.

S: [Object is focused on and observed.]

S: OK, I learned a cup.

Direct instruction: spatial relations

U: The book is to the right of the mug.

S: [Objects are focused on and observed.]

S: OK. This relation is called "to the right of".

S: [Objects are focused on and observed.]

U: Where are the objects?

S: The book is to the right of the mug.

CLASP centre for linguistic theory and studies in probability

Dialogue games as learning strategies

UNIVERSITY OF GOTHENBURG

Reinforcement from feedback

U: What is this?

S: [Object is focused on and observed.]

S: This is a cup.

U: That is correct.

S: [Observation is added to the models.]

CLASP centre for linguistic theory and studies in probability

Dialogue games as learning strategies

Reinforcement from feedback

U: What is this?

S: [Object is focused on and observed.]

S: This is a cup.

U: That is correct.

S: [Observation is added to the models.]

Self-initiative: spatial relations

S: [Objects are focused on and observed.]

S: One of the objects is book, the other one is box. Please tell me where things are.

U: The book is to the left of the box.

S: OK, this relation is called "to the left of".

Demo

UNIVERSITY OF GOTHENBURG

https://www.youtube.com/watch?v=AsEgaka6tH0

CLASP centre for linguistic theory and studies in probability

Summary

UNIVERSITY OF GOTHENBURG

- Challenges for working with robots in a dynamic world
- A brief overview of language in interaction
- Architectures of robotic systems and information fusion
- ROS (Robot Operating System) as a software platform for building interactive systems
- KILLe as a framework for learning grounded language incrementally and online from a human tutor

CLASP centre for linguistic theory and studies in probability

References I

UNIVERSITY OF GOTHENBURG

Simon Baron-Cohen, Alan M. Leslie, and Uta Frith. 1985. Does the autistic child have a “theory of mind” ? *Cognition*, 21(1):37–46.

Cynthia Breazeal, Matt Berlin, Andrew Brooks, Jesse Gray, and Andrea L. Thomaz. 2006. Using perspective taking to learn from ambiguous demonstrations. *Robotics and Autonomous Systems*, 54(5):385–393.

Joyce Chai. 2019. Collaboration in situated language communication. Invited talk, Combined Workshop on Spatial Language Understanding (SpLU) and Grounded Communication for Robotics (RoboNLP), The 2019 Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019).

Herbert H. Clark and Deanna Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition*, 22(1):1–39.

Simon Dobnik. 2009. *Teaching mobile robots to use spatial words*. Ph.D. thesis, University of Oxford: Faculty of Linguistics, Philology and Phonetics and The Queen’s College, Oxford, United Kingdom.

References II

Simon Dobnik and Erik de Graaf. 2017. KILLE: a framework for situated agents for learning language through interaction. In *Proceedings of the 21st Nordic Conference on Computational Linguistics (NoDaLiDa)*, pages 162–171, Gothenburg, Sweden. Northern European Association for Language Technology (NEALT), Association for Computational Linguistics.

Simon Dobnik, Christine Howes, Kim Demaret, and John D. Kelleher. 2016. Towards a computational model of frame of reference alignment in Swedish dialogue. In *Proceedings of the Sixth Swedish language technology conference (SLTC)*, pages 1–3, Umeå. Umeå University.

Simon Dobnik, Christine Howes, and John D. Kelleher. 2015. Changing perspective: Local alignment of reference frames in dialogue. In *Proceedings of goDIAL – Semdial 2015: The 19th Workshop on the Semantics and Pragmatics of Dialogue*, pages 24–32, Gothenburg, Sweden.

Erik de Graaf. 2016. Learning objects and spatial relations with Kinect. Master's thesis, Department of Philosophy, Linguistics and Theory of Science. University of Gothenburg, Gothenburg, Sweden, June, 8th. Supervisor: Simon Dobnik, examiner: Richard Johansson, opponent: Lorena Llozhi.

Geert-Jan M. Kruijff, Hendrik Zender, Patric Jensfelt, and Henrik I. Christensen. 2007. Situated dialogue and spatial organization: what, where... and why? *International Journal of Advanced Robotic Systems*, 4(1):125–138. Special issue on human and robot interactive communication.

Paul M. Newman and Hugh F. Durrant-Whyte. 2001. An efficient solution to the SLAM problem using geometric projections. In *Proceedings of the November 2001 SPIE conference*, Boston, USA.

Morgan Quigley, Ken Conley, Brian Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, and Andrew Y Ng. 2009. ROS: an open-source robot operating system. In *ICRA workshop on open source software*, volume 3, page 5.

References IV

UNIVERSITY OF GOTHENBURG

Gabriel Skantze. 2016. Real-time coordination in human-robot interaction using face and voice. *AI Magazine*, 37(4):19–31.

Laura Stoia, Darla Magdalena Shockley, Donna K. Byron, and Eric Fosler-Lussier. 2008. Scare: a situated corpus with annotated referring expressions. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, pages 650–653, Marrakech, Morocco. European Language Resources Association (ELRA).

Andrea Lockerd Thomaz, Matt Berlin, and Cynthia Breazeal. 2005. An embodied computational model of social referencing. In *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005.*, pages 591–598. IEEE.

Hendrik Zender, Óscar Martínez-Mozos, Patric Jensfelt, Geert-Jan M. Kruijff, and Wolfram Burgard. 2008. Conceptual spatial representations for indoor mobile robots. *Robotics and Autonomous Systems*, 56(6):493–502. Special issue "From sensors to human spatial concepts".