Interactive Grounded Language Understanding

Jean Rouat[1], Simon Brodeur[1], Sean Wood[1], Luca Celotti[1], Florian Strub[2], Bilal Piot[2], Olivier Pietquin[2], Stéphane Dupont[3], Gueorgui Pironkov[3], Huseyin Cakmak[3], Giampiero Salvi[4], Niklas Vanhainen[4], Aaron Courville[5], Harm de Vries[5], Olivier Mastropietro[5], Ana C. Murillo[6], Javier Civera[6], Pablo A. Millan[6], Manuel Lopes[7], Florian Golemo[7], Yoan Mollard[7], Roger K. Moore[8]

[1] Université de Sherbrooke, Canada
[2] Université de Lille 1, France
[3] Université de Mons, Belgium
[4] Royal Institute of Technology, Sweden
[5] Université de Montréal, Canada
[6] Universidad de Zaragoza, Spain
[7] Inria Bordeaux Sud-Ouest, France
[8] University of Sheffield, United-Kingdom

chist-era
Projects Seminar
2016

IGLU

Consortium partners

Contributors
The IGLU consortium is composed of 8 research teams, across 6 different countries.
The project is a total effort of 325 person-months (PM).

Experts
- Deep learning – A. Courville
- Reinforcement learning – O. Pietquin, B. Piot
- Neurosciences and cognitive sciences – J. Rouat, R. K. Moore
- Robotics – M. Lopes, A. C. Murillo, J. Civera
- Signal processing (audition, vision) – J. Rouat, S. Dupont, G. Salvi
- Human-machine interaction – S. Dupont

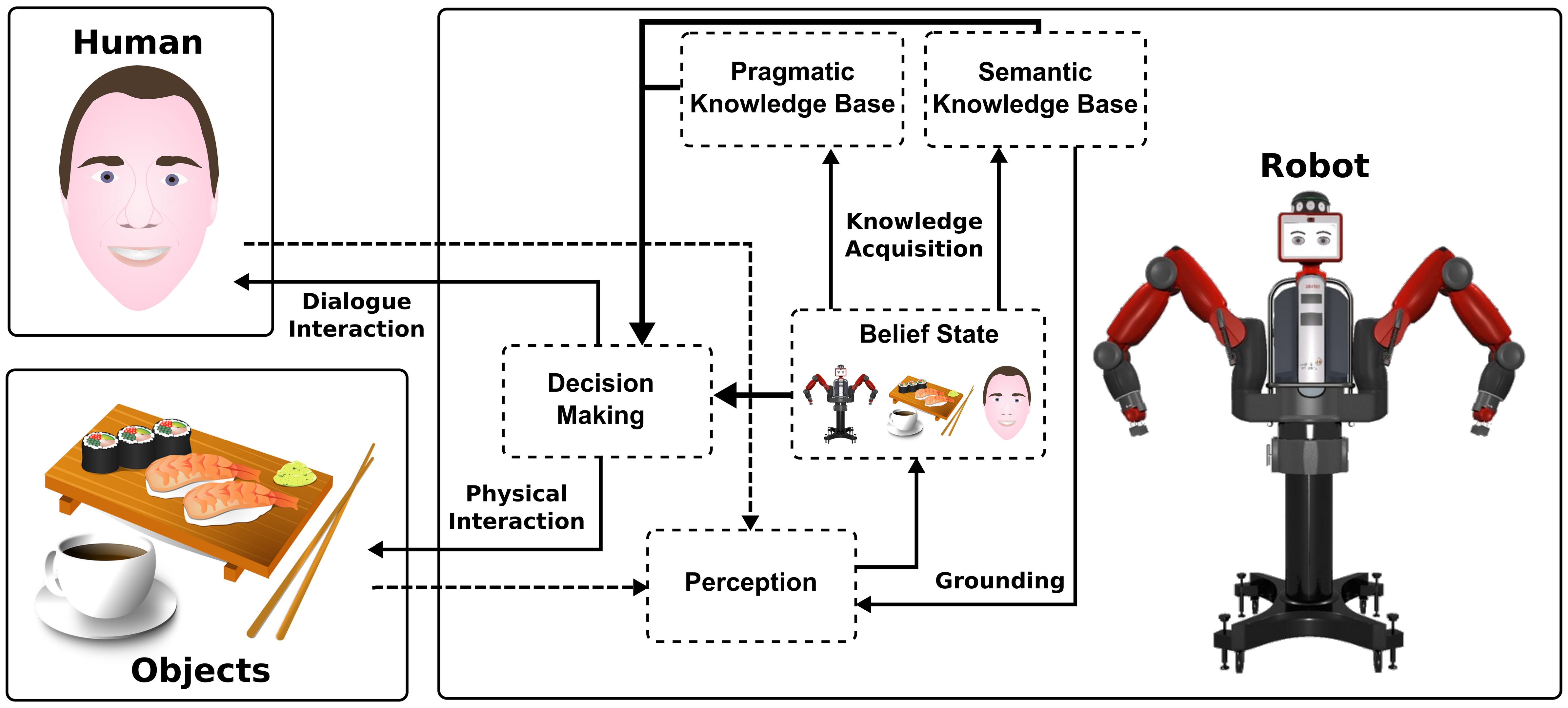
Objectives of the project

Contribute to HLU by conducting research in the following topics, so as to unify the respective contributions made to these topics into a single architecture:
- Semantic knowledge acquisition, modelling and grounding.
- Pragmatic knowledge acquisition through interaction and modelling.
- Decision-making based on reinforcement learning able to tackle semantic and pragmatic knowledge representations.

Evaluate the architecture in the context of human-agent interaction, aiming for human language acquisition and understanding:
- Reproducible experiments will be made possible by the creation of database suitable for this task.
- Research results will be published in relevant publications, conferences and videos.
- Dissemination of the tools and data generated as part of the project will be open to the scientific community.

Application

An unified architecture with application in cooking: a situated/embodied agent interacting with a human physically and through verbal dialogue



Grounded knowledge acquisition:
- Acquisition of such perceptually-grounded knowledge in virtual (avatar) and embodied (robotic) agents.
- Driven by multimodal experience and language interaction with a human.
- Handle large-scale multimodal inputs.

Interaction-driven learning:
- Pragmatic knowledge (identifying or conveying intention) must be present to complement semantic knowledge.
- Developmental approach where knowledge grows in complexity while driven by multimodal experience and language interaction.

Decision-making:
- Models of dialogues, human emotions and intentions as part of the decision-making process.
- Anticipation and reaction not only based on its internal state (own goal and intention, perception of the environment), but also on the perceived state and intention of the human interactant.

Research method

Semantic Knowledge Modelling, Acquisition and Grounding:
- Online learning of the semantics of new multimodal objects, and update of the semantic knowledge base for new vocabulary items.
- Linking objects with their perceptual properties, modelling relations between objects (e.g. IS-A, CAN relations), and modelling action and affordance at the conceptual level.
- Multimodal feature extraction through deep learning methods for the sensory pathway.
- Visual and auditory scenes analysis.
- Automatic speech recognition and natural language processing for the dialogue pathway, to provide symbolic information.
- Regenerate the perceptual representations related to stored concepts.
- Offline learning to build prior knowledge from available databases and ontology (e.g. WordNet, ImageNet).

Pragmatic Knowledge Acquisition through Interaction and Modelling:
- Scene description to provide hints on the topic of the dialogue and interaction.
- Human action and affordance from scene understanding and dialogue interactions.
- Human emotion from visual and acoustic cues.
- Human semantic simulation as a prediction of human intentions and behaviors.

Decision-making based on Reinforcement Learning:
- Central executive to integrate many high-level functions in managing the dialogue and actions.
- Perform inference (action selection), accessing the knowledge bases as needed.
- Reinforcement learning to select the optimal actions based on the belief state and what must be inferred through the semantic and pragmatic knowledge bases.
- Physical interactions to use motion planning and execution to achieve complex manipulation tasks.
- Verbal interactions to use natural language generation and text-to-speech synthesis tools.

Scientific impacts

Scientific impacts on machine learning and knowledge representation:
- Move toward interaction and cooperation with situated agent, where temporal aspect is important.
- Deep learning on spatio-temporal multimodal data could be the next leap leading to even more success in complex problems.
- Data and semantic knowledge collected from the project could be contributed and merged with RoboBrain.

Scientific impacts on neuroscience and cognitive science:
- Emerging trend in neurosciences to consider the brain as a hierarchical generative model of the world.
- Insight about the significance and impact of using semantic simulation in the context of human-agent interaction.
- New highlights and contribution to the parallel scientific debate of connectionism versus symbolism for cognition.

Ongoing work

We study metrics to evaluate
- the impacts of grounded/not grounded and interaction/no interaction.
- the impact of HLU on the human-agent interaction.
- the adaptation of the agent to new situations.
- the anticipation abilities of the agent.
- the assimilation of new knowledge by the agent.

Implementation
We study scenarios, will soon proceed to data acquisition and are completing the roadmap while research is beginning with the different teams attached to WPs.