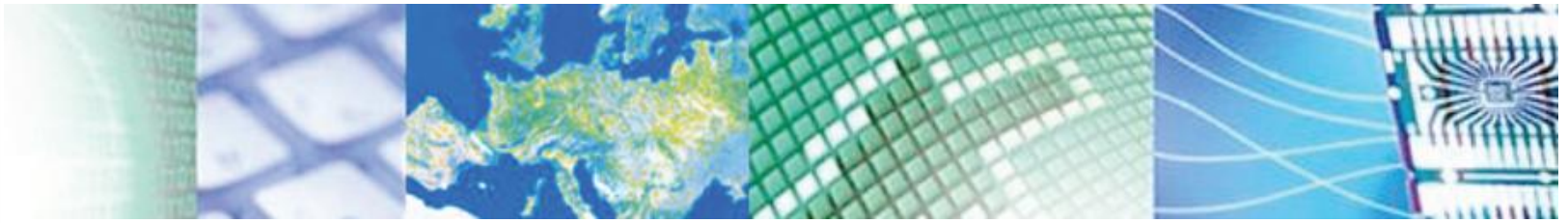


CHIST-ERA Projects Seminar
Topic Human Language Understanding – Grounded Language Learning

Patrick Gallinari

Brussels, March 22-23, 2017

FUNDING OPPORTUNITIES from the
FUTURE & EMERGING TECHNOLOGIES scheme

Introduction: Topic description

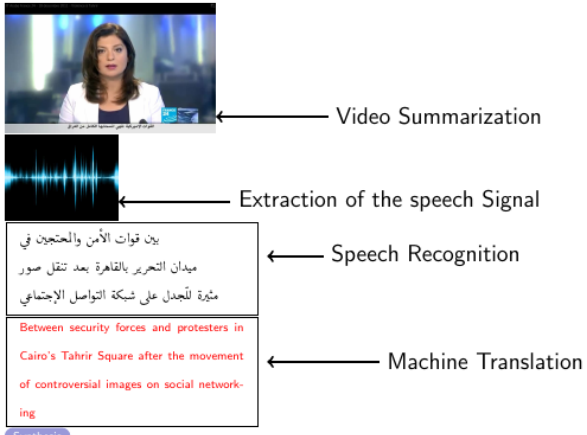
- **The goal:**
  - ✓ Ground language learning in the perceptual, emotional and sensorimotor experience of the system
- **Why:**
  - ✓ To model high-level, semantic & pragmatic knowledge in a robust way, from varied data, considering situational context
- **How:**
  - ✓ Multidisciplinary approach: combine human language processing with related fields such as developmental robotics or cognitive science.
  - ✓ Evaluation
    - **Well defined metrics and protocols to measure progress**

Introduction: Projects of the topic

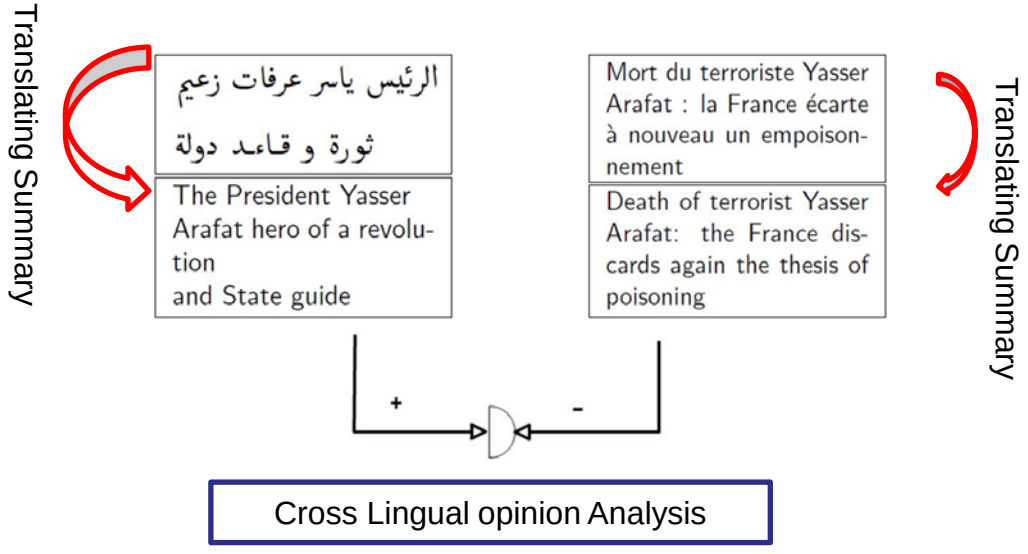
HLU	AMIS	Carlos-Emiliano	Gonzalez-Gallardo	University of Avignon	FR - France	carlos.cicak@gmail.com
HLU	AMIS	Mikołaj	Leszczuk	AGH University of Science and Technology	PL - Poland	leszczuk@agh.edu.pl
HLU	AMIS	Kamel	Smaili	University of Lorraine LORIA	FR - France	smaili@loria.fr
HLU	ATLANTIS	Brigitte	Krenn	Austrian Research Institute for Artificial Intelligence	AT - Austria	brigitte.krenn@ofai.at
HLU	ATLANTIS	Thierry	Poibeau	LATTICE-CNRS	FR - France	thierry.poibeau@ens.fr
HLU	IGLU	Jean	Rouat	Univ. Sherbrooke	CA - Canada	jean.rouat@usherbrooke.ca
HLU	IGLU	Simon	Brodeur	University of Sherbrooke	CA - Canada	Simon.Brodeur@USherbrooke.ca
HLU	IGLU	Jean-Benoit	Delbrouck	UMONS	BE - Belgium	Jean-Benoit.DELBROUCK@umons.ac.be
HLU	IGLU	Stéphane	Dupont	Université de Mons	BE - Belgium	stephane.dupont@umons.ac.be
HLU	MUSTER	Dengxin	Dai	ETH Zurich	CH - Switzerland	dai@vision.ee.ethz.ch
HLU	MUSTER	Oier	Lopez de Lacalle	University of the Basque Country	ES - Spain	oier.lopezdelacalle@ehu.eus
HLU	MUSTER	patrick	Gallinari	University Pierre et Marie Curie	FR - France	patrick.gallinari@lip6.fr
HLU	MUSTER	Benjamin	Piwowarski	Université Pierre et Marie Curie	FR - France	benjamin.piwowarski@lip6.fr
HLU	MUSTER	Aitor	Soroa	University of Basque Country	ES - Spain	a.soroa@ehu.eus
HLU	MUSTER	Marie-Francine	Moens	KU Leuven	BE - Belgium	sien.moens@cs.kuleuven.be
HLU	MUSTER	Guillem	Collell Talleda	KU Leuven	BE - Belgium	guillem.collelltalleda@kuleuven.be
HLU	M2CR	Loïc	Barrault	LIUM, University of Le Mans	FR - France	loic.barrault@univ-lemans.fr
HLU	ReGROUND	Luc	De Raedt	KU Leuven	BE - Belgium	luc.deraedt@cs.kuleuven.be
HLU	ReGROUND	Andreas	Persson	Örebro University	SE - Sweden	andreas.persson@oru.se
HLU	ReGROUND	Pedro	Zuidberg Dos Martires	KU Leuven	BE - Belgium	pedro.zuidbergdosmartires@cs.kuleuven.be
HLU	ReGROUND	Jesse	Davis	KU Leuven	BE - Belgium	jesse.davis@cs.kuleuven.be
D2K	CAMOMILE	Claude	Barras	LIMSI, CNRS, Univ. Paris-Sud	FR - France	claude.barras@limsi.fr

Introduction: Projects of the topic

- **AMIS: Access Multilingual Information opinionS**
  - ✓ Goals
    - **(1) Understanding / summarize video in a foreign language;**
      - How a user can access information in a foreign language
      - Develop a multilingual system of understanding without human intervention
    - **(2) Cross-lingual opinion analysis**
- **Distinctive features**
  - ✓ Collaborative architecture for understanding and comparing videos



Understanding by summarizing

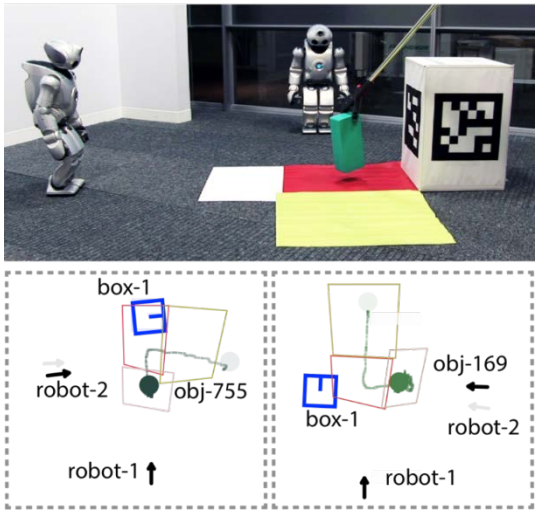


Cross Lingual opinion Analysis

- **Partners:** LORIA (France), AGH (Poland), DEUSTO (Spain), UA (France)

Introduction: Projects of the topic

- **ATLANTIS:** Artificial Language Understanding in Robots
  - ✓ Goal:
    - **Understand and model the first stages of language learning**
    - Synthesize the major transitions in the **emergence of languages using agent-based computational models**
- **Distinctive features**
  - ✓ task-based grounded learning
    - **object reference: draw attention to objects**
      - How pointing and gestures emerge from action
    - **Navigation: describe a path of movements**
  - ✓ reversible language processing
    - **parsing and production**
    - **How words and grammar emerge from language games**
  - ✓ Learning with a tutor or by interaction
  - ✓ empirical corpora as a blueprint for computational models



- **Partners:** Vrije Universiteit Brussel (AI Lab), Austrian Research Institute for Artificial Intelligence, Universitat Pompeu Fabra (IBE), Lattice-CNRS, Sony Computer Science Lab

IGLU

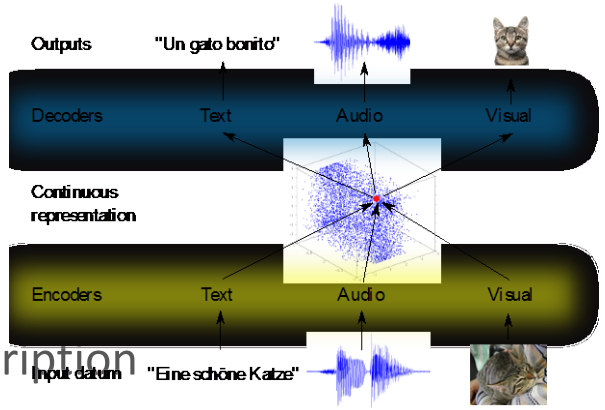
- **IGLU:** Interactive Grounded Language Understanding
  - ✓ Goal: Synthesize the major transitions in the **emergence of languages using agent-based Cognitive - Reinforcement Learning framework**
- **Distinctive features & realizations**
  - ✓ Motor, physical & verbal interactions
    - **Language interactions with humans: incorporate models of dialog**
    - **Learn physical interaction with objects**
  - ✓ 3 open access databases – 3 learning scenarios: GuessWhat?! game for supervised learning on large data (https://guesswhat.ai); Multi-modal Human Robot Interaction with incremental learning (http://robots.unizar.es/IGLUdataset/); Create for multimodal learning (https://github.com/sbrodeur/ros-icreate-bbb)
  - ✓ **International Workshop on Grounding Language Understanding, Satellite of Interspeech 2017 (**http://www.speech.kth.se/glu2017/**).** Submission deadline: 24 May 2017
  - ✓ Website for publications list and databases: (https://iglu-chistera.github.io)
  - ✓ Others: Initial studies on novelty detection, a sound source separation system, a neural network architecture for speech recognition, a proposal for incremental learning of multimodal objects and an interactive dense segmentation of images to be used for databases labelling.

Partners: Québec (Sherbrooke, Montréal), Sweden (KTH), France (Lille 1, INRIA), Spain (Zaragoza), Belgium (Mons), UK (Sheffield – not funded)

Introduction: Projects of the topic

- **M2CR:** **Multimodal Multilingual Continuous Representations for HLU**
  - ✓ Goal
    - **Define an common architecture for continuous representation of multimodal and multilingual data**
    - **Address major tasks in HLU via a unified deep learning architecture**
    - **Joint learning from multiple modalities**

- **Achievements:**
  - ✓ Pure neural MT and ASR systems
  - ✓ Image encoder: invertible conditional GAN
  - ✓ CRF-based and Neural SLU system
  - ✓ **Multimodal** Machine Translation and Image description



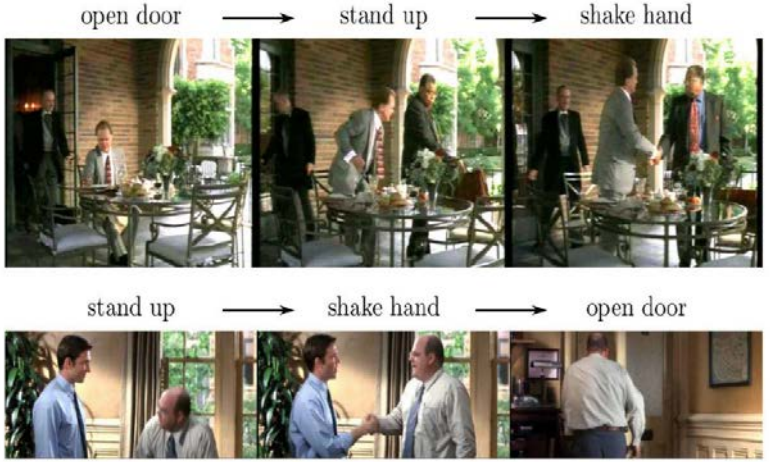
- **Next:**
  - ✓ Multi-task systems: multi-purpose representations

- **Partners:** CVC (Barcelona, Spain), LIUM (Le Mans, France), MILA (Montreal, Québec)

MUSTER

KU Leuven (Be), ETH Zurich (Ch), UPMC – Paris (Fr), U. Basque Contry (Spain)

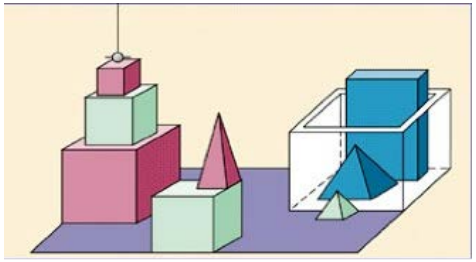
- **MUSTER – Mu**ltimodal processing of **S**patial and **T**emporal Exp**R**essions
  - ✓ Video, images and text to ground semantic representations of words and sentences, and express their spatial and temporal relations
  - ✓ Use the novel improved semantic representations to improve HLU
- **Distinctive features**
  - ✓ Visual modality to help understand language
  - ✓ New paradigm: processing linguistic phenomena related to objects, actions, space and time in language
  - ✓ Evaluation on a series of semantic tasks
- **Main results**
  - ✓ Datasets: Semantic Text similarity and Video
  - ✓ Annotation tool for videos
  - ✓ Image-grounded embeddings



- **Partners**: KU Leuven (B), ETH Zurich (Ch), UPMC (Fr), Univ. Basque country (Sp)

Introduction: Projects of the topic

- **ReGround: Relational Symbol Grounding through Affordance learning**
  - ✓ Associate symbols in language with referents in an environment
  - ✓ Goal: From Winograd's SHRDLU to the real world, here kitchen environment
- **Distinctive features**
  - ✓ multi-modal input (perception and language)
  - ✓ take into account the context & environment; multiple objects and their relationships
  - ✓ build on a notion of affordance from robotics
    - potential actions in objects in the environmental context
- **Results so far**
  - ✓ Anchoring + relational affordances
    - **Link perception of physical world to object properties**
  - ✓ Identification of position in NLP (symbol grounding)



Put the blue pyramid on the block in the box

Bring me the tea pot and the sugar

- **Partners:** KU Leuven (Belgium), Koç University (Turkey), Örebro University (Sweden)

Introduction: Projects of the topic

	Multi-modal	Multi-lingual	Physical Robots	Actions	Relations	Open Data	Systematic Evaluation
AMIS	✓	✓				✓	✓
Atlantis	✓		✓	✓	✓	✓	✓
IGLU	✓		✓	✓	✓	✓	✓
M2CR	✓	✓				✓	✓
Muster	✓			✓	✓	✓	✓
ReGround	✓		✓	✓	✓	✓	✓

Upcoming challenges and needs (1/3)

- **Scientific challenges & needs**
  - ✓ Heterogeneous data (multi-modal, multi-lingual)
  - ✓ Many current techniques need large amounts of data: how to address this challenge?
    - **investigate techniques that only need few data ...**
      - e.g: unsupervised / weakly supervised learning
    - **... or can use data from different sources independently**
      - e.g: lifelong learning / autonomous learning
      - e.g. additional sources, unlabeled data
  - ✓ How to transfer learned knowledge across different contexts?

Upcoming challenges and needs (2/3)

- **Scientific challenges & needs**
  - ✓ How to connect data to actions?
    - **use data for action**
      - e.g: anticipation, planning
    - **use action for data acquisition**
      - e.g: sensor planning, perception focus
  - ✓ How to evaluate an intelligent interactive system in real situations?
    - **what are the performance metrics?**
    - **how can we define benchmarks?**

Upcoming challenges and needs (3/3)

- **Organizational challenges & needs – provide resources for:**
  - ✓ Data
    - **create bodies of annotated data**
  - ✓ Evaluation
    - **shared test facilities, standard challenges, evaluation campaigns**
  - ✓ Common platform (hardware/software)
    - **affordable, maintenable**
  - ✓ Exchange data, software components, knowledge
    - **Make public the outputs of the inititiative**
      - Wshops, summer school, …
        - **April 2018 ! Jean Rouat**
      - HLU Website listing the facilities developed by the projects
        - **e.g. data sharing, platforms produced by the partners**

Remaining challenges and needs - SWOT

- **Strengths**
  - ✓ Large multimodal datasets are starting to be available usually developed by communities or large groups
  - ✓ Deep Learning as common framework for different modalities - works well
- **Weaknesses**
  - ✓ Deep Learning is not enough
  - ✓ Relevant multi-modal data not yet always available
  - ✓ Copyright and distribution of corpora
- **Opportunities**
  - ✓ Needs from industry – hopefully
  - ✓ Common practices, component and data reuse, availability of data, software, computation
  - ✓ Common representations for multi-modal phenomena
  - ✓ Better perception and actuation hardware
- **Threats**
  - ✓ Ethics – use of technology by companies
  - ✓ Research driven by some companies
  - ✓ Research diversity

Remaining challenges and needs

- **Challenges**
  - ✓ AI, reasoning
  - ✓ Transfer knowledge from different contexts
  - ✓ Integrating different approaches
  - ✓ Data Sharing
- **In progress**
  - ✓ Data production
  - ✓ Formulation of problem settings

Questions

Questions ?