Frontiers of Language and Robotics:
Learning, Understanding, Usage and Symbol Emergence
in the Real-World Environment

Tadahiro Taniguchi
1) Professor, College of Information Science & Engineering,
   Ritsumeikan University
2) Visiting General Chief Scientist, AI solution center,
   Panasonic

Invited talk at Workshop on Language and Robotics,
IEEE/RSJ International Conference on Intelligent Robots and Systems
(IEEE IROS 2018),
30th September 2018
Tadahiro Taniguchi  @tanichu

• Professor, Emergent System Laboratory, College of Information Science and Engineering, Ritsumeikan University, Japan
  – 2003-2006: PhD student, Kyoto University
  – 2005-2008: JSPS research fellow, Kyoto University
  – 2008: Assistant professor, Ritsumeikan University
  – 2010: Associate professor, Ritsumeikan University
  – 2015-2016: Visiting Associate Professor, Imperial College London
  – 2017: Professor, Ritsumeikan University
  – 2017: Visiting General Chief Scientist, Panasonic Corporation
    AI solution center (20% C.A.)

• Research Topics
  – Machine learning, Intelligent Robotics,
    Symbol emergence in robotics, Language acquisition
Success of deep learning: Encoder-decoder architecture has solved many problems

Labeled data/
Perfect simulation environment

Image captioning

Labeled data matters!

Machine translation

Input
Encoder NN

Decoder NN
Output

Visual recognition

Playing video games

Labeled data/
Perfect simulation environment

Input
Encoder NN
Decoder NN
Output

Visual recognition

Playing video games

Labeled data matters!
How can we really achieve linguistic/semiotic communication between human users and autonomous robots in the real-world environment in our future society?
Symbol emergence in robotics [Taniguchi 2016]

Tadahiro Taniguchi, Takayuki Nagai, Tomoaki Nakamura, Naoto Iwahashi, Tetsuya Ogata, and Hideki Asoh,
Language understanding of unconstrained sentences in the real-world picking up task
[Hattori+ 2018]

Fig. 1: An illustration of object picking via human-robot interaction.

Human: hey move that brownish fluffy stuff to the lower right bin.

Robot: which one?
(two objects highlighted in [ ])

Human: the one next to the green and blue box.

Robot: I got it.
(one object highlighted in [ ])

Fig. 5: Robot setup for experiments

Reinforcement learning-based approach towards language acquisition [Hermann+ 2017]

SpCoA++ simultaneously learn spatial concepts and a language model, i.e., word dictionary in an unsupervised manner.

Cross-Situational Learning with Bayesian Generative Models for Multimodal Category and Word Learning in Robots [Taniguchi+ 2017]

Natural language processing (NLP):
How can we go further in language learning, understanding and usage

- Progresses in NLP tasks
  - Machine translation
  - Image & video captioning
  - Dialogue systems
  - Speech recognition

- Natural language processing in the real-world environment
  - Text data alone cannot give an AI to understand the full meaning of natural language
  - Symbol grounding/emergence
  - Language learning, understanding and usage in the situated embodied systems

Let’s leap out from “text-only” NLP
We humans can use language differently from other animals. The linguistic capability enables us to collaborate with other agents, i.e., multi-agent coordination, and to form social norm and structure.

Such magnificent capabilities gave human species the highest position of species on the earth, this capability related to language can be regarded as a fruit of adaptation to the real-world environment from the viewpoint of evolution.

Humans can not only use language but also learn a language. The input data received by language learners during the learning process is NOT written text data, but multimodal sensorimotor information including speech signal, haptic information, and visual information. The language learning needs to be performed in the real-world environment which is full of uncertainty.

In language understanding, the real-world multi-modal information is essential as well. When I say "please take it" pointing at an object, you will use visual information to reduce uncertainty about the interpretation of "it," i.e., exophora. In various situations, the existence of real-world, i.e., embodied, information is essential.
Language & Robotics

- **Viewpoints of NLP and Linguistics**
  - Include real-world, i.e., multimodal and physical, information into language understanding.
  - Put more importance on embodiment.
  - We need robot having a body with sensorimotor system in a real-world, i.e., embodied cognitive system.

- **Viewpoints of Robotic and AI**
  - Create robots that can linguistically/symbolically communicate with people.
  - To understand sentences given by people in the real-world environment, the robot need to interpret their meanings on the basis of their real-world experiences.
  - Reconstruct the bottom-up organization of human cognitive system from physical interaction to linguistic knowledge.

Workshop on Language and Robotics

A human child acquires many physical skills, concepts, and knowledge, including language, through physical and social interaction with his/her environment.

How do we become able to communicate via symbols?

We’d like to obtain an understanding of the computational process of mental development and language acquisition.

Constructive approach
Develop robotic and computational models to better understand the original

Symbol Emergence in Robotics
Development through a self-organizational learning process based on real-world sensorimotor information.
Development though a self-organizational learning process based on real-world sensorimotor information.
Development through a self-organizational learning process based on real-world sensorimotor information

- **Estimation of intention**
- **Planning capability**
- **Perceptual categories**
- **Motor skills**
- **Symbol systems (Language)**

**Sensorimotor information (observations)**

- **Bottom-up organization** of a variety of mutually dependent cognitive functions based on sensorimotor information
Symbol emergence in robotics: a survey

Tadahiro Taniguchi, Takayuki Nagai, Tomoaki Nakamura, Naoto Iwahashi, Tetsuya Ogata, and Hideki Asoh
Online spatial concept acquisition method

**SpCoSLAM** [Taniguchi+ 2017]
(including word discovery task)

Akira Taniguchi, Yoshinobu Hagiwara, Tadahiro Taniguchi and Tetsunari Inamura, Online Spatial Concept and Lexical Acquisition with Simultaneous Localization and Mapping, IEEE IROS 2017
Spatial concept is multimodal

Position

Word

Visual information

Sound, smell

“Where is the space?”

“How do they call the space?”

“How does the space look like?”

“This is the third table”
“A meeting space”
“Under the air conditioner”
Graphical model of SpCoSLAM

Map

Self location

Cluster of positions

Where is the space?

Online learning of Multimodal spatial concept

Map

How does the space look like?

Visual information

How do they call the space?

“Third table” “Meeting space”

Words

Online learning of Multimodal spatial concept

Map

How does the space look like?

Visual information

How do they call the space?

“Third table” “Meeting space”

Words
Online spatial concept acquisition method

**SpCoSLAM** [Taniguchi+ 2017]  
(including word discovery task)

Akira Taniguchi, Yoshinobu Hagiwara, Tadahiro Taniguchi and Tetsunari Inamura, Online Spatial Concept and Lexical Acquisition with Simultaneous Localization and Mapping, IEEE IROS 2017
Visualizing Robot’s knowledge and memories with HoloLens (Mixed reality!)

- Occupancy grid
- Rendering the robot’s map on the real world
- Obtained occupancy grid
- Permeabilization
- Position distribution of a spatial concept
- SpCoSLAM forms position distribution with Gaussian distributions
- Generating hemisphere instead of Gaussian
- Sending learned parameters to HoloLens
- Place concept
Visualizing robot’s memory and perception

Human-linked cyber technology??
SERKET: An Architecture for Connecting Stochastic Models to Realize a Large-Scale Cognitive Model

- Connecting cognitive modules developed as probabilistic generative models and letting them work together as a single unsupervised learning system.
- Having inter-module communication of probabilistic information and guaranteeing theoretical consistency.
- Providing a cognitive module as a ROS model. This kind of idea allow us distributed development of probabilistic model-based cognitive systems.

SERKET: An Architecture for Connecting Stochastic Models to Realize a Large-Scale Cognitive Model

Tomoaki Nakamura¹*, Takayuki Nagai¹ and Tadahiro Taniguchi²

¹ Department of Mechanical Engineering and Intelligent Systems, University of Electro-Communications, Tokyo, Japan,
² Department of Information Science and Engineering, Ritsumeikan University, Shiga, Japan

To realize human-like robot intelligence, a large-scale cognitive architecture is required for robots to understand their environment through a variety of sensors with which they are equipped. In this paper, we propose a novel framework named Serket that enables the construction of a large-scale generative model and its inferences easily by connecting sub-modules to allow the robots to acquire various capabilities through interaction with their environment and others. We consider that large-scale cognitive models can be constructed by connecting smaller fundamental models hierarchically while maintaining

Frontiers: open challenges

✓ **Unsupervised lexical acquisition**
  ✓ How can a robot discover words and phonemes?
  ✓ Grounding words on sensorimotor information.

✓ **Unsupervised syntactic parsing**
  ✓ How can a robot learn dependency analysis and phase structure analysis?

✓ **Grounding phrases**
  ✓ Find chunks like “a bottle of milk” and “from the kitchen,” and its relation to the real-world phenomena.

✓ **Learning and inference of speech act**
  ✓ How can a robot understand that the speech is an imperative sentence and an interrogative sentence? The robot needs to understand speech acts.

✓ **Unsupervised learning of verbs (and predicates).**
  ✓ Learning motion/action primitives is more difficult problem for robots than perceptual categories, e.g., objects, color and places.
  ✓ “Good bye trajectories and welcome object-oriented/goal-directed actions” e.g., through, grasp and roll. Words acting like functions taking input variables.

✓ **NLP over probabilistic distribution over sentences**
  ✓ Recognition results are usually (/naturally) obtained as a posterior distribution over sentences or other structural internal representations like \( P(\text{Sentence}|\text{Speech}) \) and \( P(\text{Tree}|\text{Sentence}) \). Marginalization should be considered.

✓ Other linguistic phenomena
  ✓ Metaphor, euphoria, and
Symbol Emergence in Cognitive Developmental Systems: a Survey

[Taniguchi+ 2018]

(Accepted to IEEE Transactions on Cognitive Developmental Systems)
Symbol Emergence in Robotics for Future Human-Machine Collaboration
記号創発ロボティクスによる人間機械コラボレーション基盤創成 (2015-2021)
JST-CREST 5.5 years 300 million JPY (2 million £ approx)

Leader
T. Nagai (UEC)

- Home Robotics
- Human-Robot communication
- Embodied language acquisition
- Modeling human-robot collaboration
- Multimodal Information Processing
- Bayesian modeling of knowledge
- Deep learning
- Symbol Emergence
- Cloud robotics

...Embodied semantic communication and collaboration with Robots
Ritsumeikan Global Innovation Research Organization
International and Interdisciplinary Research Center of
Next-generation Artificial Intelligence and Semiotics

PL: Tadahiro Taniguchi 谷口忠大 (情報理工学部・教授)

Group 1
Self-driving cars

Group 2
Service robots

Group 3
Semiotics

Group 4
Future AI society

Group 5
Sound technology

GL1: 和田隆広（情理・教授）
TL1-2 深尾隆則（理工・教授）

GL2: 島田伸敬（情理・教授）

GL3: 吉田寛（先端研・教授）
TL3-1 北野圭介（映像・教授）

GL4: 青山敦（MOT・教授）

GL5: 西浦敬信（情理・教授）
理工・情理（BKC）
Information

Funded by

Email: taniguchi@ci.ristumei.ac.jp
Twitter: @tanichu
Facebook: Tadahiro Taniguchi
HP: http://www.tanichu.com