

# Language Emergence in Multi-Agent Communication Games

PhD Candidate: Xenia Ohmer<sup>1</sup>

Supervisors: Michael Franke<sup>1</sup>, Peter König<sup>1</sup>

<sup>1</sup>*University of Osnabrück*

Recently, there has been renewed interest in language emergence simulations involving agents interacting in communication games. In contrast to earlier work, current studies usually involve realistic visual input and rely on neural networks as agents that are trained with (deep) reinforcement learning algorithms (see for example [3], [1], [2]). One motivation for this research is the development of artificial agents which use a grounded language that generalizes well to new situations and new communication partners - including humans. Another motivation is the extension of existing language evolution research by working with more realistic input and more complex internal representations.

In my PhD project I would like to use such models to study the interdependencies between context, internal representations and language. So far, I have formulated possible research questions and implemented a simple communication system which can be extended to investigate these questions.

In my implementation I am working with two neural network agents, a *speaker* and a *listener*, playing a referential game (similar to the setup in [2]). The speaker is presented with an object belonging to the target class. It then selects one symbol from a set of symbols to describe that object. The listener in turn is presented with an object belonging to the same class plus additional distractor objects from other classes in random order. Based on the message from the speaker, it selects one of the objects. If it selects the object belonging to the target class, both speaker and listener receive a positive reward.

Both agents consist of an encoder and a decoder, each of which is implemented as a simple multilayer perceptron. The encoder maps the input objects onto an internal representation and the decoder generates a message (in case of the speaker) or a selection (in case of the listener).

The questions I would like to explore within this framework are:

1. How do the agents communicate conceptual hierarchies? What levels of abstraction emerge depending on the level of the hierarchy at which target and distractors can be discriminated?
2. Which internal representations emerge when the agents have explicit pragmatic reasoning abilities? Can we learn internal representations with the structure of a lexicon encoding the truth-conditions of messages?
3. How does category learning compare between humans and artificial agents as well as between single agent supervised learning and multi-agent reinforcement learning in communication games? In how far are the learning time-scales similar or different?

## References

- [1] Abhishek Das, Satwik Kottur, Jose M. F. Moura, Stefan Lee, and Dhruv Batra. Learning Cooperative Visual Dialog Agents with Deep Reinforcement Learning. *Proceedings of the IEEE International Conference on Computer Vision*, pages 2970–2979, 2017.
- [2] Angeliki Lazaridou, Karl Moritz Hermann, Karl Tuyls, and Stephen Clark. Emergence of Linguistic Communication from Referential Games with Symbolic and Pixel Input. *International Conference on Learning Representations*, pages 1–13, 2018.
- [3] Angeliki Lazaridou, Alexander Peysakhovich, and Marco Baroni. Multi-Agent Cooperation and the Emergence of (Natural) Language. *arxiv:1612.07182*, pages 1–11, 2016.