Emergent Language-Based Coordination In Deep Multi-Agent Systems

Marco Baroni
Universitat Pompeu Fabra
mbaroni@gmail.com

Roberto Dessì
Universitat Pompeu Fabra
Facebook AI Research
rdessi@fb.com

Angeliki Lazaridou
DeepMind
angeliki@deepmind.com

Abstract

Large pre-trained deep networks are the standard building blocks of modern AI applications. This raises fundamental questions about how to control their behaviour and how to make them efficiently interact with each other. Deep net emergent communication tackles these challenges by studying how to induce communication protocols between neural network agents, and how to include humans in the communication loop. Traditionally, this research had focussed on relatively small-scale experiments where two networks had to develop a discrete code from scratch for referential communication. However, with the rise of large pre-trained language models that can work well on many tasks, the emphasis is now shifting on how to let these models interact through a language-like channel to engage in more complex behaviors. By reviewing several representative papers, we will provide an introduction to deep net emergent communication, we will cover various central topics from the present and recent past, as well as discussing current shortcomings and suggest future directions. The presentation is complemented by a hands-on section where participants will implement and analyze two emergent communications setups from the literature. The tutorial should be of interest to researchers wanting to develop more flexible AI systems, but also to cognitive scientists and linguists interested in the evolution of communication systems.

Brief description and motivation

Just like interaction and communication are pivotal to humans engaging in complex problem solving and coordination, communication among artificial agents allow for effective coordination (both when they cooperate and when they compete). While multi-agent communication protocols can be prespecified and coded, emergent communication has emerged as a successful paradigm – agents are left free to create protocols whose semantics are not pre-determined by any form of supervision, but are rather shaped by the need to achieve their goals.

This utilitarian view of communication is familiar to linguistics (Wittgenstein, 1953). As such, initial work on multi-agent emergent communication studied the conditions under which artificial agents in constrained setups would evolve shared protocols and the latter's similarity to human language (Kirby and Hurford, 1997; Wagner et al., 2003; Steels, 1997). Recently, and after a break of some years, the topic of emergent communication has re-emerged, partially due to the successful and widespread use of deep learning in many fields. In addition to using these simulations to understand the underpinnings of natural language, much work in the field today focuses on how deep network agents could evolve robust protocols, on whether these protocols are interpretable and how it is possible to make them more natural-language-like, in order to enable human-machine communication. Given this recent turn, we started seeing papers on this topic appearing at the major NLP conferences and occasionally being recognized with best-paper awards (Kottur et al., 2017). We believe this is the right time to bring together researchers that wish to know more about the field by offering a structured tutorial on the theme.

Given the interdisciplinarity of the topic, a computational linguistics conference would allow us to reach researchers interested in it from diverse perspectives: AI and NLP researchers who want to develop flexible and robust agents able to coordinate in natural language, but also cognitive scientists/linguists wishing to use simulations to test theories about language evolution.

We will start with an introduction to the emerging field of emergent communication. We will discuss foundational work and we will introduce common experimental setups (i.e., data, training algorithms, analysis and protocol interpretability

Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: Tutorial Abstracts, pages 11 - 16
December 7-8, 2022

methods). We will also critically examine the standard practices in the field. Having established the basics, we will then move to discussing promising current directions (i.e., beyond simplistic simulations, linking emergent language to natural language and emerging protocols in situated environments). We will conclude with a hands-on session to deepen attendees' understanding of core concepts by grounding them in actual experiments, but also providing an entry point for researchers who wish to learn how to design such simulations.

Tutorial Structure

The tutorial is divided into 3 slots of around half hour, 1 and a half hour, and 1 hour, respectively. We will have 15 minutes break between each section.

Introduction Early work investigated the necessary conditions for emergence of a shared communication code among artificial agents. Experiments often employed hand-crafted models and/or very simplified environments, and the simulations focused on studying linguistic properties of the emergent protocols (Batali, 1998; Cangelosi and Parisi, 2002; Christiansen and Kirby, 2003).

Recent progress in deep (reinforcement) learning and its successful application in several fields has revamped interest in language emergence. Unlike earlier work, the use of powerful general-purpose neural network models enables experiments with agents that can interact and communicate in complex and dynamic environments. This has led to the introduction of new setups probing language-based coordination between deep agents (Sukhbaatar et al., 2016; Foerster et al., 2016; Mordatch and Abbeel, 2018). Examples of collaborative tasks in "deep emergent communication" include developing a shared code to solve riddles, crossing intersections or goal-oriented navigation.

Another line of research in deep emergent communication focuses on one of the most basic functions of human communication, namely that of referring to a specific object in the surrounding environment. The ability to denote specific items is the building block for more complex forms of collaboration, such as object use and manipulation. Work in this area tends to use a discrimination task called *referential game* (Lewis, 1969). In the game, a sender Agent generates a message that describes a target object. The message is transmitted to a Receiver agent that is tasked with recognizing the object of interest from a set of candidates. Initial work in this domain showed that agents evolve an effective communication policy to denote the content of realistic images (Lazaridou et al., 2017; Havrylov and Titov, 2017). However, later experimental findings suggested that the agents' "language" does not point to semantically meaningful concepts, relying instead on low-level visual features. Subsequent work showed that, unless explicitly constrained, emergent protocols do not develop core properties similar to natural languages, such as compositionality and efficient coding (Chaabouni et al., 2019; Rita et al., 2020). This highlights the importance of bridging the gap between emergent and natural languages, a topic that we will return to in the second part of the tutorial.

Communication between agents in typical setups happens through the exchange of either continuous or discrete messages. In this tutorial, we will focus on experiments with a discrete channel, a prerequisite for language-like human-machine communication. Channel discretization poses an important optimization challenge, given that it is not possible to back-propagate gradients through discrete nodes. We will cover the main approach to overcome this problem that is based on a widely policy gradients method, namely a varient of the REINFORCE algorithm (Williams, 1992).

Given the lack of supervision on the emergent protocol, it is not sufficient to evaluate agents' accuracy on the target task. Such performance-based analysis must be complemented by an analysis of the evolved protocol. This is a far-from-trivial task, somewhat akin to linguistic fieldwork. We will thus end the first part of the tutorial reviewing standard quantitative and qualitative protocol analysis methods currently used in the literature. (Brighton and Kirby, 2006; Lazaridou et al., 2018; Chaabouni et al., 2020; Lowe et al., 2019)

Current themes in emergent communication In the second part, we will introduce in more detail three currently "hot" topics in emergent communication research, presenting main findings along with possible research directions.

The first theme is whether deep nets can communicate about their visual input on a large scale. Lazaridou et al. (2017) showed that two interacting agents can develop a shared lexicon to describe natural images from standard computer vision datasets. The setup of Lazaridou and colleagues used single-symbols messages and sampled images from a limited set of image categories.

Later work by Havrylov and Titov (2017) and Dessi et al. (2021) scaled the visual referential game to variable-length messages and a richer pool of object categories, respectively. Another line of research tries to study the biases that emergent protocols have and whether they are similar to natural language features (Chaabouni et al., 2019, 2020). An example is the work of Rita et al. (2020), it studies which optimization constraints can lead to the emergence of languages that exhibit a human-like word-length distribution. We are still far, however, from robust and flexible visually-aware interactive agents. For instance, most simulations employ a single pair of agents in single-turn interactions, and there is currently no evidence that the emergent protocol will support successful communication with new partners. Additionally, contextual information is not modeled by the agents' protocol, whereas there is ample evidence that human language relies on contextual knowledge to discriminate objects (Glaser and Glaser, 1989; Munneke et al., 2013).

A second important theme is the ability to collaborate in more realistic, dynamic scenarios. Starting from the fully cooperative symbolic agents of Foerster et al. (2016), follow-up work looked at how to integrate different aspects of realistic coordination as they unfold between human agents. For instance, Evtimova et al. (2018) studied multi-turn interactions in a multimodal discrimination task. Das et al. (2019) experimented with embodied agents cooperating to solve a target-reaching navigation task in naturalistic 3D environments. Finally, all these experimental configurations are tied to a single task. On the other hand, natural language allows coordination to be carried out for an unlimited number of goals. However, scaling the an emergent communication setup does not come free of challenges (Chaabouni et al., 2022; Carroll et al., 2019). Future research directions should also investigate the ability of the emergent lexicon to adapt to new tasks, without forgetting those previously learnt.

The third research line studies how emergent protocols can be constrained to resemble natural language and how such languages can be used to interact with large pre-trained networks. Several approaches attempted to interleave game-playing with supervised tasks such as image labelling (Lazaridou et al., 2017; Gupta et al., 2019) and multimodal grounding (Lee et al., 2019), or tried to optimize the agents' communication based on statistics inferred from natural language corpora (Havrylov and Titov, 2017). However, later evidence found that this type of interlaced learning does not protect against forms of pragmatic drift where emergent and natural language interpretation diverges (Lazaridou et al., 2020). Yao et al. (2022) used emergent protocols as a pre-training corpus for image captioning and language modelling, showing performance benefits on downstream tasks. This shows how these protocols could be applied to improve standard NLP tasks, hinting at some structural similarities between emergent and natural languages. Language prompting have recently shown to be effective to extract information from large pre-trained models that are able to excel at many tasks. Such prompts, often manually designed, can be used to combine several powerful and diverse multimodal models (Zeng et al., 2022). Deng et al. (2022) shows how automatic prompt discovery, a method similar to language emergence in deep agents, can improve over several other prompting methods.

Future work should bridge the gap between the language evolved in interactive simulations, usually consisting of short denotational messages, and the syntactic and semantic knowledge acquired by deep networks pre-trained on static large-scale datasets. Additionally how these emergent languages can be used to interact with large and powerful pre-trained models remains an important open challenge.

Hands-on session The final part of the tutorial consists in an interactive hands-on session using EGG (Kharitonov et al., 2021), a Python toolkit designed to offer an easy entry point into emergent communication simulations. By providing implementations of common neural network architectures and simulation setups, it allows developers to quickly code and run a language emergence experiment on both CPU and GPU devices.

In this interactive coding session, we will guide the audience through two experimental setups. In a first configuration, we experiment with a realistic scenario involving natural data. We will provide pre-trained agents that, through a large-scale visual discrimination task, successfully converged on a shared communication policy. We will then probe the agents' communication skills by analysing the messages triggered by unseen input images. This exercise will give the audience a flavor for common challenges involved in interpreting agents' protocol.

In the second half of the session, we will show

how emergent protocols could be used to interact with (large) language models. We will show how automatic discovery of prompts can be used to extract information from pre-trained task-agnostic networks for downstream NLP tasks. This will show the connection between emergent communication and modern NLP.

Further information

Presenters **Marco Baroni** is ICREA research professor at Universitat Pompeu Fabra. **Angeliki Lazaridou** is staff research scientist at DeepMind. Marco and Angeliki co-authored one of the earliest and most influential papers on emergent communication among deep net agents (Lazaridou et al., 2017) as well as a recent survey of the area (Lazaridou and Baroni, 2020). Marco has extensive teaching experience, including interdisciplinary classes aimed at computer scientists, linguists and cognitive scientists, and lectures and tutorials in international venues such as ESSLLI, ACL and the CIFAR Deep Learning Summer School (where he presented an introduction to deep net emergent communication). He was recently awarded an ERC Advanced Grant to work on emergent communication. Angeliki's work in the area was recognized with a 2019 ICML best-paper mention (Jaques et al., 2019). She co-initiated the Emergent Communication Neurips Workshop series (which ran successfully for 6 years). **Roberto Dessì** is a 3rd-year PhD student at Facebook AI Research and Universitat Pompeu Fabra. His work focuses on scaling up emergent communication research, including a paper on the topic to appear at NeurIPS 2021. Roberto was a co-organizer of the last two Emergent Communication workshops and is currently the maintainer of the EGG toolkit for emergent communication simulations.

Tutorial type and breadth We propose a tutorial on an emerging area that has not been previously covered in ACL/EMNLP/NAACL/COLING tutorials. While we are active researchers in the field and we will review some of our own work, the tutorial attempts to survey the area as a whole, as shown by the fact that the majority of references in this proposal are to papers we did not author.

Audience: target, background and size We target two audience types: AI/NLP researchers who might look at emergent communication protocols as a tool to build more flexible multi-agent AI systems; and linguists/cognitive scientists interested in how emergent communication simulations might provide insights into the origins and nature of human and animal communication. The only strict prerequisite consists in basic programming skills in Python, in order to follow the hands-on part of the tutorial. We do not expect the audience to have a technical background in linguistics. While we will rely on standard notions from machine learning, such as cost functions and backpropagation, attendees can get a good high-level view of the area even without this background. This is the first time the tutorial has been offered, but several regular talks by Lazaridou and Baroni introducing the area have registered high attendance. On the one hand, the tutorial has broad interdisciplinary appeal and introduces a novel area to NLP.

Recommended reading While not strictly necessary, participants would benefit from a look at the survey of Lazaridou and Baroni (2020).

Diversity We are a diverse team of instructors, gender-wise and seniority-wise (one senior professor, one senior researcher, one advanced-stage PhD student). We are affiliated with one university and two different industry labs. We expect that the tutorial topic will attract a diverse audience, as it is of interest to both AI/NLP practitioners and linguists/cognitive scientists. While the focus is not on natural language *per se*, we observe that emergent communication research looks at typological research on language variety for inspiration, and it is not reliant on language-specific resources.

Ethics Autonomous agent communication raises ethical issues specifically in terms of *transparency* (see, e.g., `https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence`). Problems related to the development of opaque protocols (including bias control) and how to spur the emergence of interpretable inter-agent communication will be discussed in the tutorial.

Materials and technical requirements We will use slides and provide scripts for the hands-on part section, where we will use Google Colab and the EGG library (Kharitonov et al., 2021). Attendees should bring a laptop and all materials will be made publicly available.

References

John Batali. 1998. Computational simulations of the emergence of grammar. In James Hurford, Michael Studdert-Kennedy, and Chris Knight, editors, Approaches to the Evolution of Language: Social and Cognitive Bases, pages 405–426. Cambridge University Press, Cambridge, UK.

Henry Brighton and Simon Kirby. 2006. Understanding linguistic evolution by visualizing the emergence of topographic mappings. Artificial Life, 12:229–242.

Angelo Cangelosi and Domenico Parisi. 2002. Computer simulation: A new scientific approach to the study of language evolution. In Simulating the evolution of language, pages 3–28. Springer.

Micah Carroll, Rohin Shah, Mark K Ho, Tom Griffiths, Sanjit Seshia, Pieter Abbeel, and Anca Dragan. 2019. On the utility of learning about humans for human-ai coordination. In Advances in Neural Information Processing Systems, volume 32. Curran Associates, Inc.

Rahma Chaabouni, Eugene Kharitonov, Diane Bouchacourt, Emmanuel Dupoux, and Marco Baroni. 2020. Compositionality and generalization in emergent languages. In Proceedings of ACL, pages 4427–4442, virtual conference.

Rahma Chaabouni, Eugene Kharitonov, Emmanuel Dupoux, and Marco Baroni. 2019. Anti-efficient encoding in emergent communication. In Proceedings of NeurIPS, Vancouver, Canada. Published online: https://papers.nips.cc/paper/2019.

Rahma Chaabouni, Florian Strub, Florent Altché, Eugene Tarassov, Corentin Tallec, Elnaz Davoodi, Kory Wallace Mathewson, Olivier Tieleman, Angeliki Lazaridou, and Bilal Piot. 2022. Emergent communication at scale. In International Conference on Learning Representations.

Morten Christiansen and Simon Kirby, editors. 2003. Language Evolution. Oxford University Press, Oxford, UK.

Abhishek Das, Théophile Gervet, Joshua Romoff, Dhruv Batra, Devi Parikh, Mike Rabbat, and Joelle Pineau. 2019. Tarmac: Targeted multi-agent communication. In International Conference on Machine Learning, pages 1538–1546. PMLR.

Mingkai Deng, Jianyu Wang, Cheng-Ping Hsieh, Yihan Wang, Han Guo, Tianmin Shu, Meng Song, Eric P. Xing, and Zhiting Hu. 2022. Rlprompt: Optimizing discrete text prompts with reinforcement learning.

Roberto Dessi, Eugene Kharitonov, and Marco Baroni. 2021. Interpretable agent communication from scratch (with a generic visual processor emerging on the side). In Advances in Neural Information Processing Systems.

Katrina Evtimova, Andrew Drozdov, Douwe Kiela, and Kyunghyun Cho. 2018. Emergent communication in a multi-modal, multi-step referential game. In Proceedings of ICLR Conference Track, Vancouver, Canada. Published online: https://openreview.net/group?id=ICLR.cc/2018/Conference.

Jakob Foerster, Ioannis Alexandros Assael, Nando de Freitas, and Shimon Whiteson. 2016. Learning to communicate with deep multi-agent reinforcement learning. In Proceedings of NIPS, pages 2137–2145, Barcelona, Spain.

Wilhelm R Glaser and Margrit O Glaser. 1989. Context effects in stroop-like word and picture processing. Journal of Experimental Psychology: General, 118(1):13.

Abhinav Gupta, Ryan Lowe, Jakob Foerster, Douwe Kiela, and Joelle Pineau. 2019. Seeded self-play for language learning. In Proceedings of the Beyond Vision and LANguage: inTEgrating Real-world kNowledge (LANTERN), pages 62–66, Hong Kong, China. Association for Computational Linguistics.

Serhii Havrylov and Ivan Titov. 2017. Emergence of language with multi-agent games: Learning to communicate with sequences of symbols. In Proceedings of NIPS, pages 2149–2159, Long Beach, CA.

Natasha Jaques, Angeliki Lazaridou, Edward Hughes, Caglar Gulcehre, Pedro Ortega, Dj Strouse, Joel Leibo, and Nando De Freitas. 2019. Social influence as intrinsic motivation for multi-agent deep reinforcement learning. In Proceedings of ICML, pages 3040–3049, Long Beach, CA.

Eugene Kharitonov, Roberto Dessì, Rahma Chaabouni, Diane Bouchacourt, and Marco Baroni. 2021. EGG: a toolkit for research on Emergence of lanGuage in Games. https://github.com/facebookresearch/EGG.

Simon Kirby and James Hurford. 1997. Learning, culture and evolution in the origin of linguistic constraints. In Fourth European conference on artificial life, pages 493–502. Citeseer.

Satwik Kottur, José Moura, Stefan Lee, and Dhruv Batra. 2017. Natural language does not emerge 'naturally' in multi-agent dialog. In Proceedings of EMNLP, pages 2962–2967, Copenhagen, Denmark.

Angeliki Lazaridou and Marco Baroni. 2020. Emergent multi-agent communication in the deep learning era. https://arxiv.org/abs/2006.02419.

Angeliki Lazaridou, Karl Moritz Hermann, Karl Tuyls, and Stephen Clark. 2018. Emergence of linguistic communication from referential games with symbolic and pixel input. In Proceedings of ICLR Conference Track, Vancouver, Canada. Published online: https://openreview.net/group?id=ICLR.cc/2018/Conference.

Angeliki Lazaridou, Alexander Peysakhovich, and Marco Baroni. 2017. Multi-agent cooperation and the emergence of (natural) language. In Proceedings of ICLR Conference Track, Toulon, France. Published online: https://openreview.net/group?id=ICLR.cc/2017/conference.

Angeliki Lazaridou, Anna Potapenko, and Olivier Tieleman. 2020. Multi-agent communication meets natural language: Synergies between functional and structural language learning. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 7663–7674, Online. Association for Computational Linguistics.

Jason Lee, Kyunghyun Cho, and Douwe Kiela. 2019. Countering language drift via visual grounding. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 4385–4395, Hong Kong, China. Association for Computational Linguistics.

David Lewis. 1969. Convention. Harvard University Press, Cambridge, MA.

Ryan Lowe, Jakob Foerster, Y-Lan Boureau, Joelle Pineau, and Yann Dauphin. 2019. On the pitfalls of measuring emergent communication. In Proceedings of AAMAS, pages 693–701, Montreal, Canada.

Igor Mordatch and Pieter Abbeel. 2018. Emergence of grounded compositional language in multi-agent populations. In Proceedings of AAAI, pages 1495–1502, New Orleans, LA.

Jaap Munneke, Valentina Brentari, and Marius Peelen. 2013. The influence of scene context on object recognition is independent of attentional focus. Frontiers in Psychology, 4:552.

Mathieu Rita, Rahma Chaabouni, and Emmanuel Dupoux. 2020. "LazImpa": Lazy and impatient neural agents learn to communicate efficiently. In Proceedings of the 24th Conference on Computational Natural Language Learning, pages 335–343, Online. Association for Computational Linguistics.

Luc Steels. 1997. The synthetic modeling of language origins. Evolution of communication, 1(1):1–34.

Sainbayar Sukhbaatar, arthur szlam, and Rob Fergus. 2016. Learning multiagent communication with backpropagation. In Advances in Neural Information Processing Systems, volume 29. Curran Associates, Inc.

Kyle Wagner, James A Reggia, Juan Uriagereka, and Gerald S Wilkinson. 2003. Progress in the simulation of emergent communication and language. Adaptive Behavior, 11(1):37–69.

Ronald Williams. 1992. Simple statistical gradient-following algorithms for connectionist reinforcement learning. Machine learning, 8(3-4):229–256.

Ludwig Wittgenstein. 1953. Philosophical investigations. Blackwell, Oxford, UK.

Shunyu Yao, Mo Yu, Yang Zhang, Karthik R Narasimhan, Joshua B. Tenenbaum, and Chuang Gan. 2022. Linking emergent and natural languages via corpus transfer.

Andy Zeng, Maria Attarian, Brian Ichter, Krzysztof Choromanski, Adrian Wong, Stefan Welker, Federico Tombari, Aveek Purohit, Michael Ryoo, Vikas Sindhwani, Johnny Lee, Vincent Vanhoucke, and Pete Florence. 2022. Socratic models: Composing zero-shot multimodal reasoning with language. arXiv.