Emergence of Linguistic Conventions in Multi-Agent Systems Through Situated Communicative Interactions

Doctoral Consortium

Jérôme Botoko Ekila
Vrije Universiteit Brussel
Brussels, Belgium
jerome@ai.vub.ac.be

ABSTRACT

The field of emergent communication investigates the emergence of shared linguistic conventions among autonomous agents engaged in cooperative tasks that require communication. Conventions that arise through self-organisation are known to be more robust, flexible, and adaptive, and it removes the need for hand-crafting communication protocols. In my PhD research, I investigate how artificial agents can co-construct such conventions of linguistic structures in reference-based tasks. This problem is tackled using the language game experimental paradigm which aims to model the processes underlying the emergence and evolution of human languages. My primary contribution thus far introduces a novel methodology for the language game paradigm in the emergent setting. Using the methodology, agents can establish through self-organisation an emergent language that enables them to refer to arbitrary entities in their environment using single-word utterances. For the first time, the methodology is directly applicable to any dataset that describes entities in terms of continuously-valued features. The next phase in my research is to move from single-word utterances to multi-word utterances through the emergence of grammatical structures.

KEYWORDS

language emergence; multi-agent systems; autonomous agents; emergent communication; self-organisation; language evolution

ACM Reference Format:
Jérôme Botoko Ekila. 2024. Emergence of Linguistic Conventions in Multi-Agent Systems Through Situated Communicative Interactions: Doctoral Consortium. In *Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), Auckland, New Zealand, May 6 – 10, 2024,* IFAAMAS, 3 pages.

1 INTRODUCTION

In recent years, the field of emergent communication has witnessed an unprecedented surge of interest propelled by developments in deep multi-agent reinforcement learning. These developments have enabled agents to learn to communicate in complex environments and deal with high-dimensional input data such as images [17]. As a result, a number of impressive experiments have been conducted, addressing a range of tasks including, but not limited to, reference [6, 18, 19], navigation [14, 21, 37], and robotics [22]. In such settings, the shared task requires establishing an emergent communication protocol which enables coordination among the agents.

While these results are incredibly promising, the conditions under which the languages in these experiments emerge, differ significantly from the way human languages do [42]. For example, as discussed in [42], populations sometimes consist of only two agents [5, 26], learning is not decentralised [10, 15], or agents can either speak or listen, but not both [6, 16, 21]. To ensure that effective and coherent conventions can emerge in heterogeneous populations and can adapt to changes in tasks and environments, it is thus important to model the circumstances under which artificial languages emerges (as much as possible) to those under which human languages emerge [42].

Human languages, similar to biological evolution, are evolutionary systems where linguistic structures are shaped through the processes of variation and selection [7, 20, 29]. These processes operate at the level of concepts, words, and grammatical structures [42]. Conventions constantly emerge and evolve through local communicative interactions between members of a linguistic community. Indeed, language emerges through self-organisation, rather than through central control [8, 34, 35]. For that purpose, the language game experimental paradigm was conceived as a framework to model the emergence and evolution of languages in populations of artificial agents by simulating these aforementioned evolutionary mechanisms [24, 34, 36]. It has been demonstrated that such self-organising systems, driven by the processes of variation and selection, exhibit desirable properties such as robustness and adaptability [1, 12, 28].

In my PhD research, I investigate how artificial agents can co-construct conventions of linguistics structures in reference-based tasks using the language game experimental paradigm. The goal is to develop generally mechanisms for the emergence of such conventions in large-scale experiments for agents embedded in continuous environments. This project is divided into two concrete stages. In the first, detailed in Section 2, I developed a general methodology for the emergence of conceptual structures in a fully decentralised manner. In the second stage, detailed in Section 3, I plan to explore the emergence of grammatical structures which provide agents with the ability to compositionally assemble concepts into meaningful combinations.

Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), N. Alechina, V. Dignum, M. Dastani, J.S. Sichman (eds.), May 6 – 10, 2024, Auckland, New Zealand. © 2024 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

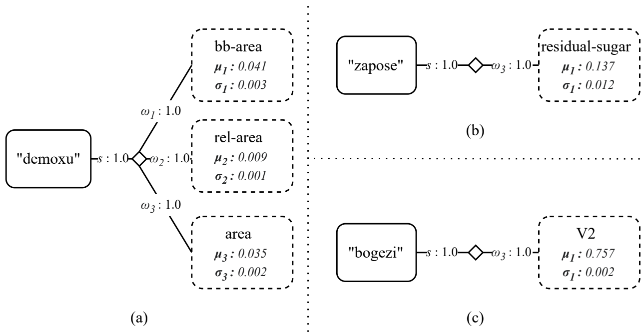

Figure 1: Examples of emerged concepts for the CLEVR (a), WINE (b) and CREDIT (c) datasets. Reprinted from [4].

2 EMERGENCE OF CONCEPTUAL STRUCTURES

Inspired by a discrimination-based approach in the tutor-learner setting [25], I introduced a methodology through which a population can establish a linguistic convention in a decentralised manner. This convention enables agents to refer to arbitrary entities within their environment [4]. In our experimental set-up, agents participate in pairwise local communicative interactions. In each interaction, two agents are sampled from the population and are randomly assigned the role of either speaker or listener. The agents are placed in a simulated environment and perceive a set of 3 to 10 entities through their own sensors. The goal of the speaker is to produce a linguistic utterance that draws attention to a particular entity. An interaction is successful if the listener, based on the utterance, is able to distinguish the entity from all others. Agents cannot transmit their raw sensor values as each agent perceives the world possibly differently [25]. For example, agents might be endowed with different sensors (heterogeneous agents), or sensor readings might differ due to noise or calibration differences.

Therefore, the agents must come up with a convention that allows them to abstract away from these raw sensor values. Specifically, agents build from the bottom up an inventory of concepts, which associate discrete symbolic labels (word forms) to concept representations (word meanings). While all concept representations are individually constructed and grounded in an agent's own sensory-motor endowment and experiences, the emergent linguistic systems of the agents are compatible on a communicative level. In other words, only the linguistic forms are shared and observed, while the meaning remains tied to the individual agent.

To illustrate, Figure 1(a) visualises a word "*demoxu*" that emerged in the inventory of an agent in one of the experiments. In this experiment, agents were situated in a simulated scene consisting of geometric objects from the CLEVR dataset [13]. The other two concepts in Figure 1 (b) and (c) emerged in other experiments detailed in the paper. The concept representation corresponds to a set of weighted Gaussian distributions, one distribution for each sensor that the agent is endowed with. The weights of the distributions are learned and reflect the importance of the feature channel to the concept. For visualisation purposes, Figure 1 only shows the channels with positive weights. A qualitative analysis (detailed in the original paper) demonstrated that the word "demoxu" emerged and was conventionalised to primarily refer to small objects. This representation can be used in both production and comprehension processes. As a speaker, the agent finds a concept that matches with the topic entity and utters the word for it. Conversely, as a listener, the same representation can be used to map an utterance to an entity in the world.

The novelty in our approach lies in the way in which agents represent, invent, adopt and align concepts. In contrast to prior work, the methodology combines three properties that have never been achieved together. It proposes an approach for decentralised, communication-based concept learning in (i) continuous feature spaces (as opposed to the discrete setting as in [43]), (ii) in a multi-agent setting (as opposed to a tutor-learner setting as in [25]), and (iii) without labels [38]. As a result, the methodology is now directly applicable to any dataset that describes entities in terms of continuously-valued features.

The methodology is experimentally validated on three tabular datasets, each describing hundreds of thousands of entities in terms of continuous features. Through a range of experiments, it is demonstrated that the methodology enables a population to converge to an effective and coherent linguistic convention, that it can effectively deal with noisy observations, uncalibrated sensors and heterogeneous populations, that the method is adequate for continual learning, and lastly that the convention can self-adapt to changes in the environment and to the communicative needs of the agents.

3 WORK PLAN

The next step of my research is to look beyond the emergence of single-word concepts and investigate how grammatical structures can emerge within populations of autonomous agents. Grammatical structures provide agents with the ability to compositionally assemble concepts into meaningful combinations. For example, consider a scene where neither "*cube*" or "*green*" is discriminative, but the combination is. Instead of inventing distinct concepts, the agent should learn to combine these concepts compositionally. The ability to create new combinations using existing building blocks improves thus the expressiveness of the emergent language. This capability enables agents to communicate effectively about objects that they were previously unable to.

Prior work on the emergence of such grammatical structures in the language game paradigm [2, 3, 11, 27, 30–33, 39–41] is limited in terms of broad applicability. The focus of these experiments was limited to the emergence of specific linguistic phenomena and was constrained to rather small-scale environments. At the moment, the language experimental paradigm lacks a general computational mechanism that facilitates the emergence of grammatical structures in large-scale experiments. Recent advancements in the domain of language acquisition through repeated situated communicative interactions are certainly promising [9, 23], but research into general operators for the emergent setting remains largely unexplored.

ACKNOWLEDGMENTS

The research reported on in this paper was financed by the European Union's Horizon 2020 research and innovation programme under grant agreement no. 951846 (MUHAI).

REFERENCES

[1] Clay Beckner, Richard Blythe, Joan Bybee, Morten H. Christiansen, William Croft, Nick C. Ellis, John Holland, Jinyun Ke, Diane Larsen-Freeman, and Tom Schoenemann. 2009. Language is a complex adaptive system: Position paper. *Language learning* 59 (2009), 1–26. https://doi.org/10.1111/j.1467-9922.2009.00533.x

[2] Katrien Beuls, Kateryna Gerasymova, and Remi van Trijp. 2010. Situated learning through the use of language games. In *Proceedings of the 19th Annual Machine Learning Conference of Belgium and The Netherlands (BeNeLearn)*. 1–6.

[3] Joris Bleys and Luc Steels. 2009. Linguistic selection of language strategies, a case study for color. In *Proceedings of the 10th European Conference on Artifical Life*. 150–157.

[4] Jérôme Botoko Ekila, Jens Nevens, Lara Verheyen, Katrien Beuls, and Paul Van Eecke. 2024. Decentralised Emergence of Robust and Adaptive Linguistic Conventions in Populations of Autonomous Agents Grounded in Continuous Worlds. In *Proceedings of the 23rd International Conference on Autonomous Agents and Multi-Agent Systems*.

[5] Diane Bouchacourt and Marco Baroni. 2018. How agents see things: On visual representations in an emergent language game. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, Ellen Riloff, David Chiang, Julia Hockenmaier, and Jun'ichi Tsujii (Eds.). Association for Computational Linguistics, 981–985. https://doi.org/10.18653/v1/D18-1119

[6] 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 *10th International Conference on Learning Representations (ICLR 2022)*. 1–30.

[7] Charles R. Darwin. 1871. *The descent of man, and selection in relation to sex* (1st ed.). Vol. 1. John Murray, London, United Kingdom.

[8] Bart de Boer. 2001. *Self-organisation in language*. Cambridge University Press, Cambridge, United Kingdom, 123–139. https://doi.org/10.1017/cbo9780511542275.009

[9] Jonas Doumen, Katrien Beuls, and Paul Van Eecke. Forthcoming. A Mechanistic Model of Constructivist Language Acquisition through Syntactico-Semantic Pattern Finding. *Artificial Intelligence* (Forthcoming). Forthcoming.

[10] Jakob Foerster, Ioannis Alexandros Assael, Nando de Freitas, and Shimon Whiteson. 2016. Learning to communicate with deep multi-agent reinforcement learning. In *Advances in Neural Information Processing Systems 29 (NIPS 2016)*, Daniel Lee, Masashi Sugiyama, Ulrike Von Luxburg, Isabelle Guyon, and Roman Garnett (Eds.). Curran Associates Inc., Red Hook, NY, USA, 2137–2145.

[11] Kateryna Gerasymova and Michael Spranger. 2010. Acquisition of Grammar in Autonomous Artificial Systems. In *Proceedings of the 19th European Conference on Artificial Intelligence (ECAI-2010)*, Helder Coelho, Rudi Studer, and Michael Woolridge (Eds.). 923–928.

[12] Francis Heylighen. 2001. The science of self-organization and adaptivity. In *Knowledge management, organizational intelligence and learning, and complexity. The encyclopedia of life support systems*, Lowell Douglas Kiel (Ed.). EOLSS Publishers, Oxford, United Kingdom, 253–280.

[13] Justin Johnson, Bharath Hariharan, Laurens van der Maaten, Li Fei-Fei, C. Lawrence Zitnick, and Ross Girshick. 2017. CLEVR: A Diagnostic Dataset for Compositional Language and Elementary Visual Reasoning. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, Washington, D.C., USA, 2901–2910.

[14] Ivana Kajić, Eser Aygün, and Doina Precup. 2020. Learning to cooperate: Emergent communication in multi-agent navigation. In *Proceedings of the 42nd Annual Conference of the Cognitive Science Society*. Cognitive Science Society, Toronto, Canada, 1993–1999.

[15] Jooyeon Kim and Alice Oh. 2021. Emergent Communication under Varying Sizes and Connectivities. In *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)*, Marc'Aurelio Ranzato, Alina Beygelzimer, Yann Dauphin, Percy S. Liang, and Jennifer W. Vaughan (Eds.). Curran Associates Inc., Red Hook, NY, USA, 17579–17591.

[16] Satwik Kottur, José Moura, Stefan Lee, and Dhruv Batra. 2017. Natural Language Does Not Emerge 'Naturally' in Multi-Agent Dialog. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, Martha Palmer, Rebecca Hwa, and Sebastian Riedel (Eds.). Association for Computational Linguistics, 2962–2967. https://doi.org/10.18653/v1/D17-1321

[17] Angeliki Lazaridou and Marco Baroni. 2020. Emergent multi-agent communication in the deep learning era. *arXiv preprint arXiv:2006.02419* (2020).

[18] Angeliki Lazaridou, Alexander Peysakhovich, and Marco Baroni. 2016. Multi-agent cooperation and the emergence of (natural) language. *arXiv preprint arXiv:1612.07182* (2016).

[19] Angeliki Lazaridou, Alexander Peysakhovich, and Marco Baroni. 2017. Multi-agent cooperation and the emergence of (natural) language. In *5th International Conference on Learning Representations (ICLR 2017)*. 1–11.

[20] John Maynard Smith and Eörs Szathmáry. 1999. *The origins of life: From the birth of life to the origin of language*. Oxford University Press, Oxford, United Kingdom.

[21] Igor Mordatch and Pieter Abbeel. 2018. Emergence of grounded compositional language in multi-agent populations. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, Sheila McIlraith and Kilian Q. Weinberger (Eds.). AAAI Press, Washington, D.C., USA, 1495–1502.

[22] Yao Mu, Shunyu Yao, Mingyu Ding, Ping Luo, and Chuang Gan. 2023. EC2: Emergent Communication for Embodied Control. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, 6704–6714.

[23] Jens Nevens, Jonas Doumen, Paul Van Eecke, and Katrien Beuls. 2022. Language acquisition through intention reading and pattern finding. In *Proceedings of the 29th International Conference on Computational Linguistics*, Nicoletta Calzolari and Chu-Ren Huang (Eds.). International Committee on Computational Linguistics, 15–25.

[24] Jens Nevens, Paul Van Eecke, and Katrien Beuls. 2019. A Practical Guide to Studying Emergent Communication through Grounded Language Games. In *AISB 2019 Symposium on Language Learning for Artificial Agents*. AISB, 1–8.

[25] Jens Nevens, Paul Van Eecke, and Katrien Beuls. 2020. From continuous observations to symbolic concepts: A discrimination-based strategy for grounded concept learning. *Frontiers in Robotics and AI* 7, 84 (2020). https://doi.org/10.3389/frobt.2020.00084

[26] Michael Noukhovitch, Travis LaCroix, Angeliki Lazaridou, and Aaron Courville. 2021. Emergent Communication under Competition. In *Proceedings of the 20th International Conference on Autonomous Agents and Multi-Agent Systems*. 974–982.

[27] Simon Pauw and Joseph Hilferty. 2012. The emergence of quantifiers. In *Experiments in Cultural Language Evolution*, Luc Steels (Ed.). Vol. 3. John Benjamins, Amsterdam, Netherlands, 277–304. https://doi.org/10.1075/ais.3.14pau

[28] Rolf Pfeifer, Max Lungarella, and Fumiya Iida. 2007. Self-organization, Embodiment, and Biologically Inspired Robotics. *Science* 318, 5853 (2007), 1088–1093. https://doi.org/10.1126/science.1145803

[29] August Schleicher. 1869. *Darwinism tested by the science of language. English translation of Schleicher 1863, translated by Alex V. W. Bikkers*. John Camden Hotten, London, United Kingdom.

[30] Michael Spranger. 2015. Incremental grounded language learning in robot-robot interactions: Examples from spatial language. In *2015 Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)*. 196–201.

[31] Michael Spranger. 2016. *The evolution of grounded spatial language*. Language Science Press, Berlin, Germany. https://doi.org/10.17169/langsci.b53.183

[32] Michael Spranger. 2017. Usage-based grounded construction learning: A computational model. In *The 2017 AAAI Spring Symposium Series*. AAAI Press, Washington, D.C., USA, 245–250.

[33] Michael Spranger and Luc Steels. 2015. Co-acquisition of syntax and semantics: an investigation in spatial language. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*, Qiang Yang and Michael Wooldridge (Eds.). AAAI Press, Washington, D.C., USA, 1909–1915.

[34] Luc Steels. 1995. A self-organizing spatial vocabulary. *Artificial Life* 2, 3 (1995), 319–332. https://doi.org/10.1162/artl.1995.2.3.319

[35] Luc Steels. 1997. Synthesising the origins of language and meaning using co-evolution, self-organisation and level formation. In *Evolution of Human Language*, James R. Hurford (Ed.). Edinburgh University Press, Edinburgh, United Kingdom.

[36] Luc Steels. 2003. The evolution of communication systems by adaptive agents. In *Symposium on Adaptive Agents and Multi-Agent Systems*, Eduardo Alonso, Daniel Kudenko, and Dimitar Kazakov (Eds.). 125–140. https://doi.org/10.1007/3-540-44826-8_8

[37] Sainbayar Sukhbaatar, Arthur Szlam, and Rob Fergus. 2016. Learning multiagent communication with backpropagation. In *Advances in Neural Information Processing Systems 29 (NIPS 2016)*, Daniel Lee, Masashi Sugiyama, Ulrike Von Luxburg, Isabelle Guyon, and Roman Garnett (Eds.). Curran Associates Inc., Red Hook, NY, USA, 2244–2252.

[38] Samarth Swarup and Les Gasser. 2010. The classification game: combining supervised learning and language evolution. *Connection Science* 22, 1 (2010), 1–24. https://doi.org/10.1080/09540090802638766

[39] Paul Van Eecke. 2018. *Generalisation and specialisation operators for computational construction grammar and their application in evolutionary linguistics Research*. Ph.D. Dissertation. Vrije Universiteit Brussel, Brussels: VUB Press.

[40] Paul Van Eecke and Katrien Beuls. 2017. Meta-layer problem solving for computational construction grammar. In *The 2017 AAAI Spring Symposium Series*. AAAI Press, Washington, D.C, USA, 258–265.

[41] Paul Van Eecke and Katrien Beuls. 2018. Exploring the creative potential of computational construction grammar. *Zeitschrift für Anglistik und Amerikanistik* 66, 3 (2018), 341–355. https://doi.org/10.1515/zaa-2018-0029

[42] Paul Van Eecke, Katrien Beuls, Jérôme Botoko Ekila, and Roxana Rădulescu. 2022. Language games meet multi-agent reinforcement learning: A case study for the naming game. *Journal of Language Evolution* 7, 2 (2022), 213–223. https://doi.org/10.1093/jole/lzad001

[43] Pieter Wellens. 2012. *Adaptive Strategies in the Emergence of Lexical Systems*. Ph.D. Dissertation. Vrije Universiteit Brussel, Brussels: VUB Press.