# Humanlike Emergent Language in Multi-Agent Systems

Doctoral Consortium

Jannik Peters
University of Wuppertal
Wuppertal, Germany
jpeters@uni-wuppertal.de

## ABSTRACT

The study of emergent language examines how artificial agents autonomously develop communication strategies to achieve shared goals. These emergent languages hold promise for creating robust, adaptive, and context-aware communication systems. However, there are significant limitations in the field, including a lack of evaluation and benchmarking frameworks and limited experimental setups that mimic human language use. My PhD research seeks to address these limitations by developing a comprehensive framework that integrates existing metrics and introduces novel measures to capture critical linguistic features. Additionally, I will analyze how experimental parameters influence the characteristics of emergent languages. My focus will be on the exploration of complex and humanlike communication settings, such as population-based, semi-cooperative, targeted, and symmetric linguistic exchange, drawing inspiration from multi-task and continual learning. Through these advancements, my research aims to enhance our understanding of the interplay between language, meaning, and environment in artificial emergent communication systems.

## KEYWORDS

emergent language; emergent communication; multi-agent systems; autonomous agents; linguistic characteristics

**ACM Reference Format:**
Jannik Peters. 2025. Humanlike Emergent Language in Multi-Agent Systems: Doctoral Consortium. In *Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Detroit, Michigan, USA, May 19 – 23, 2025*, IFAAMAS, 3 pages.

## 1 INTRODUCTION

Emergent communication (EC) among entities is based on conventions that arise from the need or benefit of coordination. Based on these conventions, Lewis [17] introduced signaling games to study the emergence of codes between autonomous agents, a concept that has profoundly influenced emergent language (EL) research. As a subfield of EC, EL research investigates the development of linguistic codes between autonomous agents [3, 16]. Early studies mainly used simple hand-crafted simulations and supervised learning, limiting insights into complex linguistic characteristics [30]. However, since 2016, there has been a significant growth in EL research driven by multi-agent reinforcement learning (MARL) approaches, enabling studies of more advanced language features [11, 12, 22, 23, 29].

One goal of EL research is to enable agents to develop communication skills that support agent-to-agent and agent-to-human interactions in a natural language (NL)-like fashion [3, 16]. In EL environments, agents develop communication by experiencing its utility in goal-oriented tasks [17]. MARL is a suitable approach to designing such an interactive learning environment that enables this goal-oriented language development [15]. Reinforcement learning (RL) settings offer agents the autonomy to learn languages that require and convey deeper understanding, distinct from the data-driven nature of large language models (LLMs) [4, 22]. ELs exhibit unique characteristics, emphasizing the discrete nature of linguistic structures and their vast potential due to adaptability and combinatorial richness. EL in MARL environments inherently poses a multi-task learning challenge, as agents must emerge a language from scratch to effectively address their environment tasks.

Despite progress towards more sophisticated EL in the last years, EL research still faces strong limitations [5, 16, 24]. First, the lack of standardized evaluation strategies, due to the absence of common metrics and compatible methodologies, leads to challenges in analyzing and foremost comparing results. Second, the paucity of research on the relationships between individual EL characteristics and their relationships with experimental settings. Third, the focus on simple and artificial problems, environments, and settings that generally do not require learning a humanlike language. Accordingly, the objective of my PhD research is twofold. Firstly, it aims to establish a comprehensive framework for evaluating humanlike ELs. Secondly, it seeks to systematically explore their emergence in more realistic communication scenarios.

## 2 EMERGENT LANGUAGE IN ARTIFICIAL INTELLIGENCE

### 2.1 Language Characteristics and Metrics

A defining goal of EL research is to develop communication systems that exhibit characteristics of NL, such as morphology [6, 20], semantics [7, 25], and pragmatics [19]. Unlike LLMs, which derive them from large-scale textual data and lack a grounded understanding of their meaning [21, 27], EL systems aim to develop these characteristics autonomously through interactions within an environment, enabling a deeper connection between language and context. Among these linguistic characteristics, a particularly interesting feature to achieve more humanlike communication systems is *positive listening* [8, 24]. Positive listening involves an agent's ability to critically analyze, interpret, and integrate information



*Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Y. Vorobeychik, S. Das, A. Nowé (eds.), May 19 – 23, 2025, Detroit, Michigan, USA.* 

from incoming messages to make informed decisions. This skill mirrors human communication, where understanding often involves careful evaluation and active engagement with shared information. However, existing metrics, such as Causal Influence of Communication (CIC) [14, 19], only measure the direct correlation between messages received and the resulting actions, thus highlighting order-following behavior [2, 14, 19]. They overlook the agent's ability to assess, reject, or reinterpret incoming information based on context or relevance.

### 2.2 Experimental Settings

The MARL problem definition significantly influences the quality and possible characteristics of EL. Four key features of a MARL problem stand out as essential for advancing EL research [24]:

**Population-based:** Moving beyond pairwise agent interactions, population-based environments offer a more realistic, robust, and scalable approach. These environments allow for diverse communication strategies and implement the need for an emergence of complex linguistic conventions, similar to those observed in human groups [1, 13, 28].

**Semi-Cooperative:** Fully cooperative settings dominate current research, often leading to the emergence of oversimplified communication strategies tailored to narrow tasks. Semi-cooperative settings, in which agents share partially aligned goals, provide a richer framework for studying the development of more versatile and robust languages [18, 19].

**Symmetric:** Symmetry ensures that all agents have an equal opportunity to assume both speaker and listener roles. This dynamic encourages the development of shared linguistic conventions, in contrast to unidirectional signaling systems that are less applicable to real-world scenarios [2, 10] but are often observable in research settings [24].

**Targeted:** Effective language use involves not just message generation but also the ability to address specific agents or groups. Targeted communication introduces complexity by requiring agents to discern whom to inform, fostering a deeper understanding of the utility and context of language use [9, 26].

### 2.3 Open Questions

The field of EL research lacks a unified framework, that provides methods and tools to examine EL, hindering progress and comparability. In our survey [24], we proposed a taxonomy aligned with NL concepts to guide research and facilitate the development of humanlike communication systems. Current metrics often focus on basic features and fail to capture complex aspects. For example, *positive listening* is oversimplified as a direct link between message and action, ignoring nuanced decision-making processes [19, 24]. To achieve humanlike EL major questions are: How to measure the properties and features of an EL or NL? Which features and metrics are most important and how do they relate to each other and the environment settings? What metrics are the most informative when comparing studies? How does EL emerge in more realistic and humanlike semi-cooperative settings and what are its characteristics?

## 3 RESEARCH OBJECTIVES

The limitations and questions highlighted earlier serve as a basis for defining my research objectives. To this end, I propose the following objectives:

**O1 - Develop a Consolidated Framework:** A key goal is to design a consolidated framework to evaluate ELs, encompassing existing metrics and introducing new ones to comprehensively measure linguistic characteristics, as discussed in Section 2.1. We have already started this by systematically reviewing existing EL metrics and identifying gaps where new measures can improve understanding [24].

**O2 - Investigate Experimental Parameters:** By systematically varying parameters such as reward structures, agent abilities, and environmental complexity, my research aims to elucidate the relationships and correlations between these parameters and specific linguistic features. This approach will provide more robust design principles for EL experiments to achieve desired NL characteristics.

**O3 - Explore Realistic Communication Settings:** This research involves adapting MARL setups to better reflect real-world communication. I will develop environments with population-based, semi-cooperative, symmetric, and goal-directed interactions. These settings will be evaluated for their impact on the emergence of humanlike language, enabling the study of nuanced behaviors such as selective information exchange and shared linguistic conventions.

## 4 WORK PLAN

**Phase 1: Emergent Language Metrics Library** The first priority is to establish a comprehensive library of metrics to evaluate ELs, guided by the taxonomy introduced in our previous survey [24] and addressing objectives *O1* and *O2*. Because the field currently lacks standardized metrics, this effort aims to move beyond superficial observations. The focus will be on reviewing existing metrics, identifying gaps, and implementing new or refined measures. While not definitive, this phase will be considered sufficiently advanced when a well-documented set of baseline metrics emerges that allows for consistent evaluation and meaningful comparisons across studies.

**Phase 2: Quantifying Positive Listening** A key gap in EL research is the definition of *positive listening*. Current agents primarily follow commands without making decisions based on the information they receive. To address objective *O1*, I will develop models that enable agents to evaluate and respond beyond mere compliance, fostering deeper understanding and humanlike behaviors. This research will contribute to the definition and measurement of more complex language features and introduce a new metric for *positive listening* as a key measure of actively processed communication.

**Phase 3: Learning Strategy Design for Language Emergence** Drawing on the principles of multitask learning, I will investigate strategies such as hierarchical and continual learning, which have shown potential for improving language acquisition and generalization across tasks. These strategies will be tested within the EL framework with the aim of promoting the emergence of more robust and naturalistic languages, thus addressing objective *O3*. The evaluation of this emergence will be guided by previously established metrics and strategies, allowing the assessment of more complex linguistic features.

## REFERENCES


[1] Akshat Agarwal, Swaminathan Gurumurthy, Vasu Sharma, Mike Lewis, and Katia Sycara. 2019. Community Regularization of Visually-Grounded Dialog. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems (ACM Digital Library)*, International Foundation for Autonomous Agents and Multiagent Systems (Ed.). International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1042–1050. https://www.ifaamas.org/Proceedings/aamas2019/pdfs/p1042.pdf

[2] Diane Bouchacourt and Marco Baroni. 2019. Miss Tools and Mr Fruit: Emergent communication in agents learning about object affordances. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, Association for Computational Linguistics (Ed.). Association for Computational Linguistics, Kerrville, TX, USA, 3909–3918. https://doi.org/10.18653/v1/P19-1380

[3] Nicolo' Brandizzi and Luca Iocchi. 2022. Emergent communication in human-machine games. In *5th Workshop on Emergent Communication at ICLR 2022*. OpenReview.net, Online, 1–11. https://openreview.net/forum?id=rqLgeQWCXZ9

[4] Jacob Browning and Yann Lecun. 2022. AI And The Limits Of Language: An artificial intelligence system trained on words and sentences alone will never approximate human understanding. https://www.noemamag.com/ai-and-the-limits-of-language/

[5] 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*, OpenReview.net (Ed.). OpenReview.net, Online, 1–30. https://openreview.net/forum?id=AUGBfDIV9rL

[6] Michael Cogswell, Jiasen Lu, Stefan Lee, Devi Parikh, and Dhruv Batra. 2019. Emergence of Compositional Language with Deep Generational Transmission. arXiv:1904.09067v2 http://arxiv.org/pdf/1904.09067v2

[7] Cédric Colas, Tristan Karch, Nicolas Lair, Jean-Michel Dussoux, Clément Moulin-Frier, Peter Ford Dominey, and Pierre-Yves Oudeyer. 2020. Language as a Cognitive Tool to Imagine Goals in Curiosity-Driven Exploration. In *Advances in Neural Information Processing Systems 33 (Advances in neural information processing systems)*, Neural Information Processing Systems Foundation (Ed.). Curran Associates Inc, Red Hook, NY, USA, 3761–3774. https://papers.neurips.cc/paper_files/paper/2020/file/274e6fcf4a583de4a81c6376f17673e7-Paper.pdf

[8] Alexander I. Cowen-Rivers and Jason Naradowsky. 2019. Emergent Communication with World Models. In *3rd Workshop on Emergent Communication at NeurIPS 2019*. NeurIPS, Online, 1–11. arXiv:2002.09604v1 http://arxiv.org/pdf/2002.09604v1

[9] Abhishek Das, Théophile Gervet, Joshua Romoff, Dhruv Batra, Devi Parikh, Michael Rabbat, and Joelle Pineau. 2019. TarMAC: Targeted Multi-Agent Communication. In *Proceedings of the 36th International Conference on Machine Learning (ICML 2019)*, PMLR (Ed.). PMLR, Online, 1538–1546. https://proceedings.mlr.press/v97/das19a/das19a.pdf

[10] Marina Dubova, Arseny Moskvichev, and Robert Goldstone. 2020. Reinforcement Communication Learning in Different Social Network Structures. In *1st Workshop on Language in Reinforcement Learning at ICML 2020*. ICML, Online, 1–8. arXiv:2007.09820v1 http://arxiv.org/pdf/2007.09820v1

[11] Jakob N. Foerster, Yannis M. Assael, Nando de Freitas, and Shimon Whiteson. 2017. Learning to Communicate with Deep Multi-Agent Reinforcement Learning. In *Advances in neural information processing systems 29*, Daniel D. Lee, Ulrike von Luxburg, Roman Garnett, Masashi Sugiyama, and Isabelle Guyon (Eds.). Curran Associates Inc, Red Hook, NY, USA, 2145–2153. arXiv:1605.06676v2 https://proceedings.neurips.cc/paper/2016/file/c7635bfd99248a2cdef8249ef7bfbef4-Paper.pdf

[12] Abhinav Gupta, Marc Lanctot, and Angeliki Lazaridou. 2021. Dynamic population-based meta-learning for multi-agent communication with natural language. In *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (Eds.). Curran Associates, Inc, Red Hook, NY, USA, 16899–16912. arXiv:2110.14241v1 https://papers.nips.cc/paper/2021/file/8caa38721906c1a0bb95c80fab33a893-Paper.pdf

[13] Laura Harding Graesser, Kyunghyun Cho, and Douwe Kiela. 2019. Emergent Linguistic Phenomena in Multi-Agent Communication Games. 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)*, Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan (Eds.). Association for Computational Linguistics, Stroudsburg, PA, USA, 3698–3708. https://doi.org/10.18653/v1/D19-1384

[14] Natasha Jaques, Angeliki Lazaridou, Edward Hughes, Caglar Gulcehre, Pedro A. Ortega, D. J. Strouse, Joel Z. Leibo, and Nando de Freitas. 2019. Social Influence as Intrinsic Motivation for Multi-Agent Deep Reinforcement Learning. In *Proceedings of the 36th International Conference on Machine Learning (ICML 2019)*, PMLR (Ed.). PMLR, Online, 3040–3049. https://proceedings.mlr.press/v97/jaques19a/jaques19a.pdf

[15] Seth Karten, Siva Kailas, Huao Li, and Katia Sycara. 2023. On the Role of Emergent Communication for Social Learning in Multi-Agent Reinforcement Learning. arXiv:2302.14276v1 http://arxiv.org/pdf/2302.14276v1

[16] Angeliki Lazaridou and Marco Baroni. 2020. Emergent Multi-Agent Communication in the Deep Learning Era. arXiv:2006.02419v2 http://arxiv.org/pdf/2006.02419v2

[17] David K. Lewis. 1969. *Convention: A philosophical study* (1 ed.). Harvard Univ. Press, Cambridge, Mass.

[18] Paul Pu Liang, Jeffrey Chen, Ruslan Salakhutdinov, Louis-Philippe Morency, and Satwik Kottur. 2020. On Emergent Communication in Competitive Multi-Agent Teams. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS '20)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 735–743. https://www.ifaamas.org/Proceedings/aamas2020/pdfs/p735.pdf

[19] Ryan Lowe, Jakob Foerster, Y-Lan Boureau, Joelle Pineau, and Yann Dauphin. 2019. On the Pitfalls of Measuring Emergent Communication. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems (ACM Digital Library)*, International Foundation for Autonomous Agents and Multiagent Systems (Ed.). International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 693–701. arXiv:1903.05168v1 https://www.ifaamas.org/Proceedings/aamas2019/pdfs/p693.pdf

[20] Diana Rodríguez Luna, Edoardo Maria Ponti, Dieuwke Hupkes, and Elia Bruni. 2020. Internal and external pressures on language emergence: least effort, object constancy and frequency. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu (Eds.). Association for Computational Linguistics, Stroudsburg, PA, USA, 4428–4437. https://doi.org/10.18653/v1/2020.findings-emnlp.397

[21] Melanie Mitchell and David C. Krakauer. 2023. The debate over understanding in AI's large language models. *Proceedings of the National Academy of Sciences* 120, 13 (2023), e2215907120. https://doi.org/10.1073/pnas.2215907120

[22] 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 and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence*, Association for the Advancement of Artificial Intelligence (Ed.). AAAI Press, Palo Alto, California, USA, 1495–1502. https://cdn.aaai.org/ojs/11492/11492-13-15020-1-2-20201228.pdf

[23] 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 MultiAgent Systems (AAMAS '21)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 974–982. arXiv:2101.10276v1 https://www.ifaamas.org/Proceedings/aamas2021/pdfs/p974.pdf

[24] Jannik Peters, Constantin Waubert de Puiseau, Hasan Tercan, Arya Gopikrishnan, Gustavo Adolpho Lucas de Carvalho, Christian Bitter, and Tobias Meisen. 2024. A Survey on Emergent Language. https://doi.org/10.48550/ARXIV.2409.02645 arXiv:2409.02645

[25] Shuwen Qiu, Sirui Xie, Lifeng Fan, Tao Gao, Song-Chun Zhu, and Yixin Zhu. 2022. Emergent Graphical Conventions in a Visual Communication Game. In *Advances in Neural Information Processing Systems 35 (Advances in neural information processing systems)*, Neural Information Processing Systems Foundation (Ed.). Curran Associates Inc, Red Hook, NY, USA, 13119–13131. https://proceedings.neurips.cc/paper_files/paper/2022/file/550ff553efc2c58410f277c667d12786-Paper-Conference.pdf

[26] David Simoes, Nuno Lau, and Luis Paulo Reis. 2020. Multi-agent actor centralized-critic with communication. *Neurocomputing* 390 (2020), 40–56. https://doi.org/10.1016/j.neucom.2020.01.079

[27] Settaluri Sravanthi, Meet Doshi, Pavan Tankala, Rudra Murthy, Raj Dabre, and Pushpak Bhattacharyya. 2024. PUB: A Pragmatics Understanding Benchmark for Assessing LLMs' Pragmatics Capabilities. In *Findings of the Association for Computational Linguistics ACL 2024*, Lun-Wei Ku, Andre Martins, and Vivek Srikumar (Eds.). Association for Computational Linguistics, Stroudsburg, PA, USA, 12075–12097. https://doi.org/10.18653/v1/2024.findings-acl.719

[28] Paul van Eecke and Katrien Beuls. 2021. Re-conceptualising the Language Game Paradigm in the Framework of Multi-Agent Reinforcement Learning. In *COMARL AAAI 2020-2021 - Challenges and Opportunities for Multi-Agent Reinforcement Learning, AAAI Spring Symposium Series*, Association for the Advancement of Artificial (Ed.). COMARL, Online, 1–8. arXiv:2004.04722v1 https://arxiv.org/pdf/2004.04722

[29] Shresth Verma. 2021. Towards Sample Efficient Learners in Population based Referential Games through Action Advising: Extended Abstract. In *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS '21)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1689–1691. https://www.ifaamas.org/Proceedings/aamas2021/pdfs/p1689.pdf

[30] 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 (2003), 37–69. https://doi.org/10.1177/10597123030111003