SIMULATION OF EMERGENT COMMUNICATION WITH LARGE SCALE MACHINE LEARNING

Rahma Chaabouni[*1], Florian Strub[*1], Florent Altché[1], Corentin Tallec[1], Eugene Trassov[1], Elnaz Davoodi[1], Kory Mathewson[1], Olivier Tieleman[1], Angeliki Lazaridou[1], and Bilal Piot[1*]

[*]Corresponding Authors: {rahmac,fstrub,piot}@deepmind.com
[1]DeepMind

The origins of human languages are unarguably a complex research problem (Christiansen & Kirby, 2003). Over the last fifty years, this inherent complexity has led to various research approaches from many fields ranging from biology, linguistics, anthropology, computer science etc. (Harnad et al., 1976; Bickerton, 2007; Nölle et al., 2020). Here, we walk in the footprint of the computational community (Steels, 1997; Kirby, 2002; Myers-Scotton, 2002). Roughly, the computational approach aims to simulate the prerequisite or processes that could trigger the emergence of a structured language within a controlled environment. In practice, the classic modeling paradigm simulates two agents that must develop a communication protocol to solve the Lewis game; the first agent describes an object such as the second agent can locate it among a set of candidate objects. Therefore, the two agents must settle on a communication protocol to solve the game. In this paper, we discuss how new computational resources and recent machine learning algorithms, namely deep learning (LeCun et al., 2015) and reinforcement learning (Sutton & Barto, 2018), allow scaling language simulation emergence. We then discuss how this more realistic model can foster new research directions in language evolution.

Despite the recent astonishing successes of deep neural networks in solving complex tasks, computational emergent language has made little profit from these advances. Most works still consider small disentangled input spaces (Lazaridou & Baroni, 2020), where the expected language can often be reduced to a basic identity operator. In fact, modeling scenarios barely evolved over twenty years, e.g. Kirby (2002) and Ren et al. (2019) both use the same binary input vector of size eight, and only a few papers went beyond artificial input spaces (Havrylov & Titov, 2017; Lu et al., 2020). In this paper, we endorse Bickerton (2015)'s view about the necessity of complex tasks and stimuli to model human language communication. Furthermore, such challenging settings have been proven to be paramount to emulate complex distributions in the machine learning literature (Krizhevsky et al., 2012; Brown et al., 2020). We argue for enhancing the

Lewis Game to more realistic settings to gain novel and more conclusive insights.

We cast the language emergence modeling into a deep reinforcement learning framework similar to Lazaridou, Peysakhovich, and Baroni (2017). Then, we summarize the modeling tools and algorithms, and provide guidance towards good practices from the machine learning field[1]. As a core contribution, we focus on three independent aspects to challenge language emergence modeling: increasing task complexity, using complex visual inputs, generating large populations of agents. We then assess the different properties of emergent languages through; (1) Generalization by computing the agents' communication success at test time, (2) Topographic similarity as a proxy for compositionality (Brighton & Kirby, 2006), (3) Ease of learning by using the emergent protocol for transfer tasks.

First, we consider more complex tasks by increasing the number of candidates among which the listener must retrieve the target input. Specifically, agents typically discriminate between less than 20 candidates in the emergent language field (Mu & Goodman, 2021); we experiment with up to 1024 candidates in this study. We found that scaling up the task complexity entails unstable optimization. We propose to smooth language learning by using classic mathematical regularization. Furthermore, we observe that complexifying the task has two positive aspects: it better discriminates the different models and improves the generalization of the learned communication protocol.

Second, most emergent language studies situated agents in a simple one-hot vector environment (Ren et al., 2019; Rita et al., 2020), we here scale-up the input space with continuous and ambiguous visual cues by using pretrained representations of natural images (Grill et al., 2020). We note no correlation between generalization and the widely used topographic similarity metric in this set of experiments. We hence question if this metric is adequate to assess compositionality in complex setups. Inspired by the computer vision community (Grill et al., 2020), we then discuss how ease of learning may be surrogate for protocol evaluation.

Finally, we investigate the impact of population size. In particular, we scale up the Lewis game from 2 to up to 100 agents. Here, unlike what was observed in human communication (Gary Lupyan, 2010; Raviv et al., 2019), we find little to no systematic benefit on emergent languages' properties when increasing the population size. We propose alternative methods to leverage populations, namely voting and imitation among speakers (Hester et al., 2018). Our results show that such population dynamics lead to more robust, productive, and in some cases easy-to-learn languages, opening up new research opportunities.

In the end, we expect that these observations, baselines, and good practices would allow the language emergence community to benefit further from deep RL advances. We believe that such a more realistic and challenging framework is a prerequisite to moving the field closer to its goals of modeling language evolution.

[1] Code: `https://github.com/deepmind/emergent_communication_at_scale`

References

Bickerton, D. (2007). Language evolution: A brief guide for linguists. *Lingua*, *117*(3), 510–526.

Bickerton, D. (2015). *Roots of language.* Language Science Press.

Brighton, H., & Kirby, S. (2006). Understanding linguistic evolution by visualizing the emergence of topographic mappings. *Artificial life*, *12*(2), 229–242.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., et al.. (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.

Christiansen, M. H., & Kirby, S. E. (2003). *Language evolution.* Oxford University Press.

Gary Lupyan, R. D. (2010). Language structure is partly determined by social structure. *PLoS ONE 5*, *1*.

Grill, J.-B., Strub, F., Altché, F., Tallec, C., Richemond, P., et al.. (2020). Bootstrap your own latent-a new approach to self-supervised learning. *Proc. of NeurIPS*.

Harnad, S., Steklis, H. D., & Lancaster, J. B. (1976). Origins and evolution of language and speech.

Havrylov, S., & Titov, I. (2017). Emergence of language with multi-agent games: Learning to communicate with sequences of symbols. In *Proc. of NeurIPS.*

Hester, T., Vecerik, M., Pietquin, O., Lanctot, M., Schaul, T., Piot, B., Horgan, D., et al.. (2018). Deep q-learning from demonstrations. In *Proc. of AAAI.*

Kirby, S. (2002). Natural language from artificial life. *Artificial life*, *8*(2), 185–215.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Proc. of NeurIPS.*

Lazaridou, A., & Baroni, M. (2020). Emergent multi-agent communication in the deep learning era. *arXiv preprint arXiv:2006.02419*.

Lazaridou, A., Peysakhovich, A., & Baroni, M. (2017). Multi-agent cooperation and the emergence of (natural) language. In *Proc. of ICLR.*

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *nature*, *521*(7553), 436–444.

Lu, Y., Singhal, S., Strub, F., Courville, A., & Pietquin, O. (2020). Countering language drift with seeded iterated learning. In *Procof ICML.*

Mu, J., & Goodman, N. (2021). Emergent communication of generalizations. In *Proc. of NeurIPS.*

Myers-Scotton, C. (2002). *Contact linguistics: Bilingual encounters and grammatical outcomes.* Oxford, UK: Oxford University Press.

Nölle, J., Hartmann, S., & Tinits, P. (2020). Language evolution research in the year 2020: A survey of new directions. *Language Dynamics and Change*.

Raviv, L., Meyer, A., & Lev-Ari, S. (2019). Larger communities create more systematic languages. *Proceedings of the Royal Society B*, *286*(1907), 20191262.

Ren, Y., Guo, S., Havrylov, S., Cohen, S., & Kirby, S. (2019). Enhance the compositionality of emergent language by iterated learning. In *Proc. of EmeCom - Neurips).*

Rita, M., Chaabouni, R., & Dupoux, E. (2020). "LazImpa": Lazy and impatient neural agents learn to communicate efficiently. In *Proc. of ACL.*

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

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction.* MIT press.