How to Scale up Models of Language Emergence and Evolution

Remi van Trijp
Sony Computer Science Laboratories Paris
6 Rue Amyot
75005 Paris
remi.vantrijp@sony.com

Computational models of language evolution play an important role in exploring the emergence and evolution of human languages. However, a recent complexity analysis study suggests that models of language evolution, at least in their current formulation, are computationally intractable so they cannot be scaled up to more ecological scales of tens of thousands of words and are doomed to the study of toy languages (Woensdregt et al. 2021). The goal of this presentation is to demonstrate that this criticism is misleading because it reduces all models of language evolution to one specific kind of iterated learning based on Bayesian inference (e.g. Griffiths and Kalish 2007). Woensdregt et al. (2021) even discard other kinds of models as viable solutions because "it is not clear that [such models] would not run into the same wall of intractability" (p. 6). However, with this claim, the study ignores decades worth of research in the Language Game methodology (Steels 1995, 2012) that has already successfully addressed issues of scale. The goal of this presentation is therefore to show why and how this method scales so well by illustrating how it incorporates insights from constructivist usage-based learning (Tomasello 2003, Bybee 2006), heuristic decision making Gigerenzer et al. 2011), and Relevance theory (Sperber and Wilson 1986). Participants will be able to download the code supporting this presentation's claims as open-source software.

References

Andrea Baronchelli, Maddalena Felici, Vittorio Loreto, Emanuele Caglioti, and Luc Steels. "Sharp transition towards shared vocabularies in multi-agent systems." *Journal of Statistical Mechanics*, P06014, 2006.

Joan Bybee. "From usage to grammar: The mind's response to repetition." *Language*, 82(4):711–733, 2006. URL www.jstor.org/stable/4490266.

Bart De Vylder and Karl Tuyls. "How to reach linguistic consensus: A proof of convergence for the naming game." *Journal of Theoretical Biology*, 242 (4):818–831, October 2006.

Gerd Gigerenzer, Ralph Hertwig, and Thorsten Pachur, editors. *Heuristics: The Foundations of Adaptive Behavior*. Oxford University Press, Oxford, 2011.

Thomas L. Griffiths and Michael L. Kalish. Language evolution by iterated learning with Bayesian agents. *Cognitive Science*, 31:441–480, 2007.

Dan Sperber and Deirdre Wilson. *Relevance: Communication and Cognition.* Harvard University Press, Cambridge, MA, 1986.

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

Luc Steels, editor. *Experiments in Cultural Language Evolution*. John Benjamins, Amsterdam, 2012.

Michael Tomasello. *Constructing a Language. A Usage Based Theory of Language Acquisition*. Harvard University Press, Harvard, 2003.

Marieke Woensdregt, Matthew Spike, Ronald de Haan, Todd Wareham, Iris van Rooij, and Mark Blokpoel. "Why is scaling up models of language evolution hard?" *PsyArXiv Preprints*. Last edit on May 13, 2021. URL: https://doi.org/10.31234/osf.io/d2h5c. Last accessed on January 31, 2022.

Supplementary Materials

The following two graphs show how a population of 10 agents succeed at rapidly developing a shared vocabulary from scratch for referring to 25 objects, despite being faced with a hypothesis space that is too large to handle according to Woensdregt et al. (2021). More specifically, the average hypothesis space is $2^{125\times25}$, or about 5.2×10^{939}, possible languages to converge on. The language game method, however, allows the agents to already reach communicative success after less than 3000 *time steps* (600 interactions on average per agent) and full consensus after about 4000 time steps (800 interactions on average per agent). The bottom graph shows how many lexical constructions (or words) agents invent/learn before settling on an ideal lexicon size of 25. In other words, such simulations can be run in a matter of seconds on a present-day laptop or personal computer.