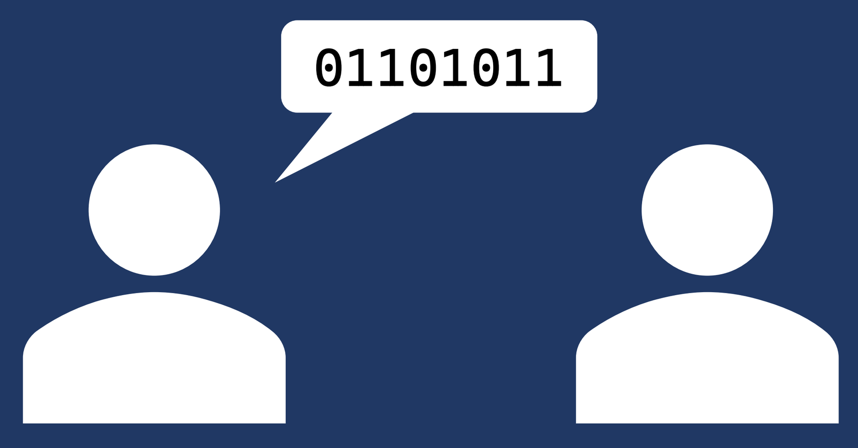
Agent-based Model of Language Change

Jack Bunyan
jackmbunyan.github.io

To understand how human language evolved, models have been created to replicate the process. One such model is the Iterative Learning Model. I explain the original model, then discover its flaws, and propose new models. I am attempting to decrease the computational complexity required for ILMs.

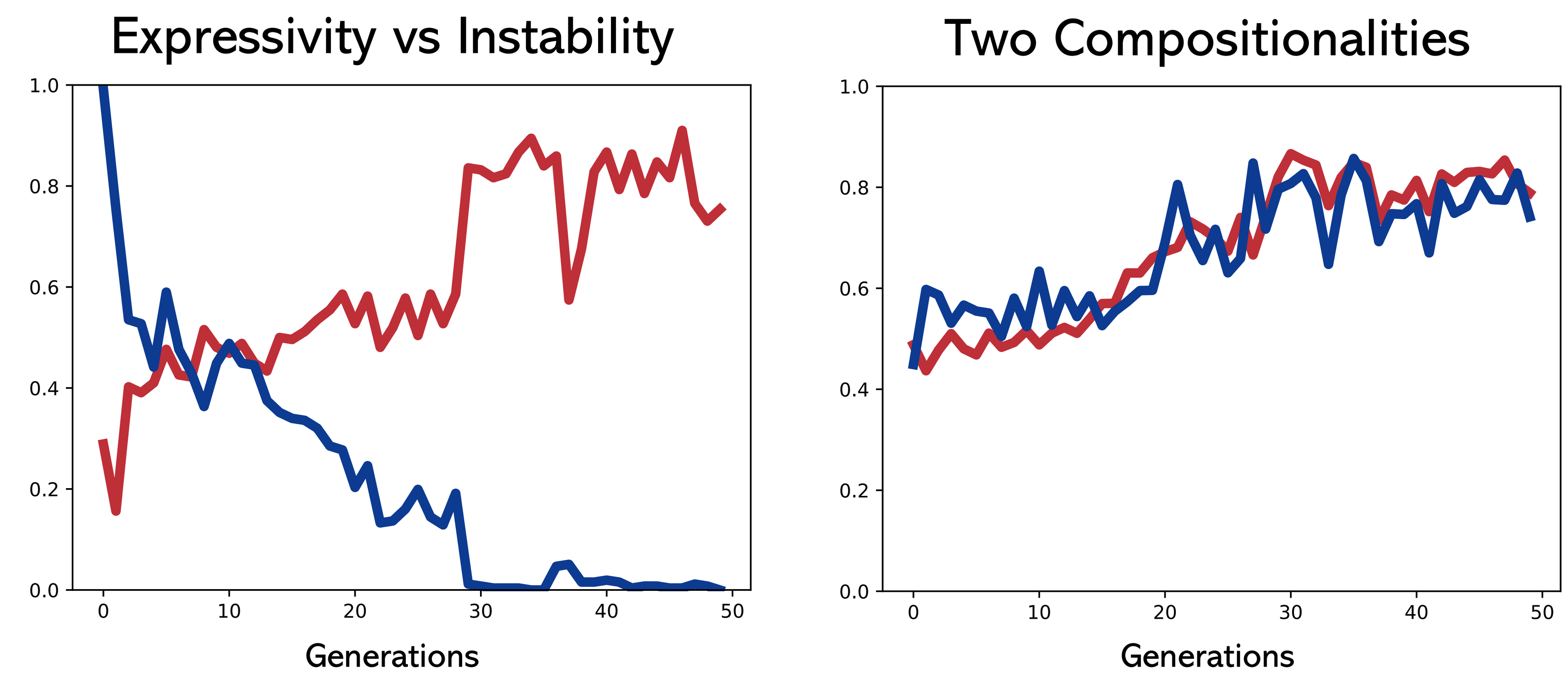


The Language Game:

Agents learn to communicate through meanings and signals represented by 8-bit binary strings. Initially, we provide a random mapping from the 256 signals to the 256 meanings and use this to train the first agent. Each "Teacher" has an internal mapping from signals to meanings which it attempts to pass on to the naïve "Learner". We investigate the different mechanisms that facilitate this training.

The KH ILM

The Iterative Learning Model we replicate is Kirby and Hurford's from 2002 [1]. Alongside the original expressivity and instability scores used, I implemented an information theoretic measure of compositionality alongside the Pearson correlation based measure.



Expressivity is a proportion of how much of the language space is being used.
Instability is a proportion of signal-meaning pairs that the agents disagree on.

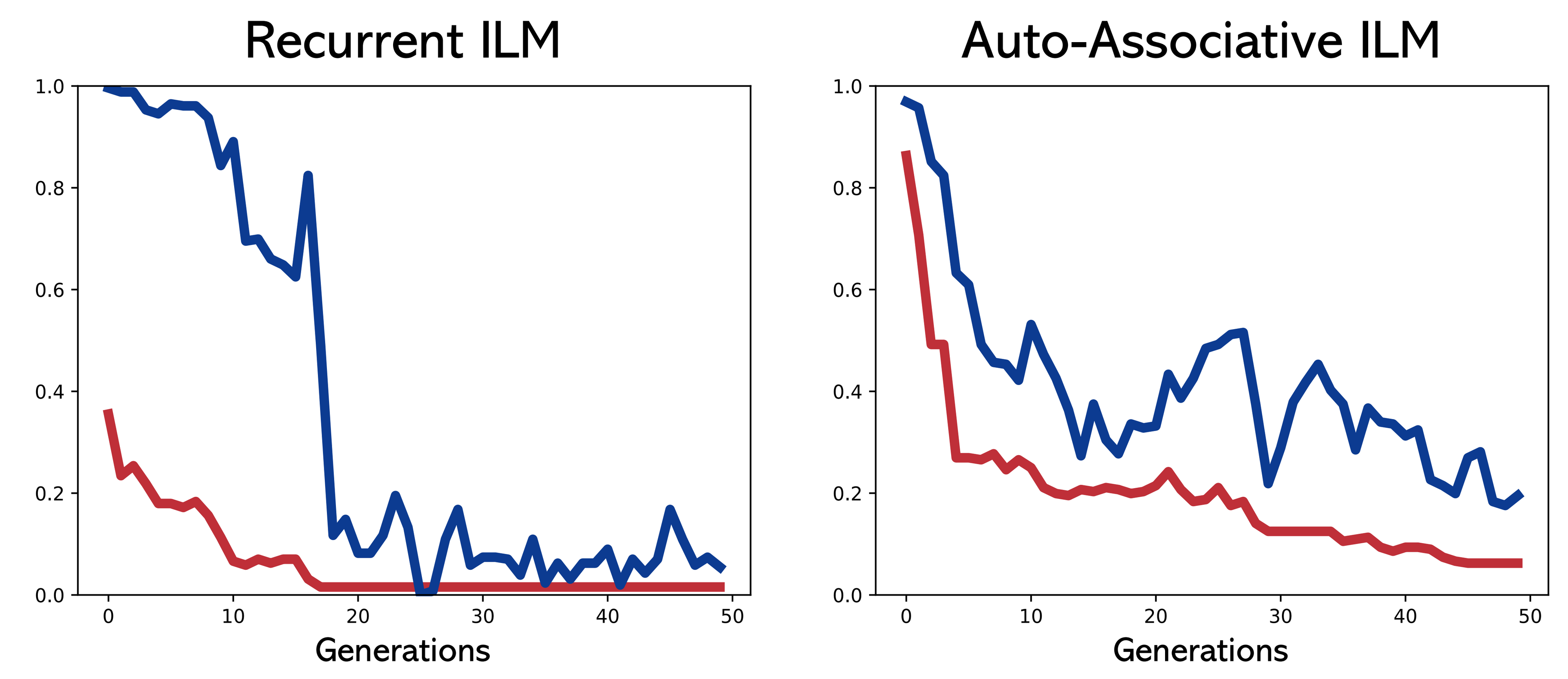
Compositionality is the correlation between bits in the signal and bits in the meaning. Two measures are used, the **original method** using correlation between bits and one using the **entropy** between each bit in the signal to the bits in the meaning.

The Obversion Problem

To facilitate the language learning in the KH model, an obverter learning strategy is used, similar to the one discussed by Oliphant and Batali [2]. It is used to create a meanings to signals mapping, based off of the neural network's mapping of signals to meanings. A confidence score is calculated for each signal as to how likely it is to have produced the given meaning; the highest confidence signal associated to the meaning is used for the mapping. However, this process is computationally expensive and thus prevents scaling without the use of a supercomputer.

Attempted Non-Obverter ILMs

My first attempt at avoiding obversion was to simply to create two neural nets to be trained, one from signals to meanings and one the other way, dubbed the Recurrent ILM. Then, I tried an auto-associative neural network for learning signal-meaning pairs; the benefit of this would be the ability to recover the signal with just one of the signal or meaning fed through the network. Both methods results are shown below.



Both models preserve the idea of agents' **instability** decreasing over time but both show **expressivity** rapidly decreasing; implying the language collapses into only a few utterances.

Results

So far, I have shown that the recurrent and auto-associative ILMs do not generate similar results to the KH ILM. However, from this I have learned a lot about what properties the obverter injects into the iterative learning process. The act of obversion seems to encourage the language being used by the agents to be expressive; to not collapse into a smaller subset of utterances. To build on this knowledge, one final ILM attempt will use a new learning function for the backpropagation to use: **a contrastive learning function** that takes into account the previous utterance to discourage using the same output for successive inputs. Hopefully, this will fulfil the role of obversion within the initial learning period and prove to be a viable alternative.

Supervised by Conor Houghton and Seth Bullock

References:
[1] Simon Kirby and James R. Hurford. The Emergence of Linguistic Structure: An Overview of the Iterated Learning Model, pages 121–147. Springer London, London, 2002.
[2] Michael Oliphant and John Batali. Learning and the emergence of coordinated communication. Center for Research on Language Newsletter, 11, 03 1997.

University of BRISTOL