

Complexity can be maintained in small populations: a model of lexical variability in emerging sign languages / Bill Thompson, Limor Raviv & Simon Kirby

Research on emerging sign languages suggests that social structure affects the process of language convergence ([Meir, Israel, Sandler, Padden & Aronoff, 2012](#)). Specifically, sign languages that emerge in small, highly-connected communities are less conventionalized, showing greater lexical variability between speakers. On the other hand, languages that emerge in larger and sparser communities tend to be more uniform. This finding is somewhat surprising in light of theoretical results suggesting that convergence rate scales with population size ([Baronchelli, Felici, Loreto, Caglioti & Steels, 2006](#)).

In this paper, we argue that the evidence from emerging sign language can be explained by an interaction between population size and memory limitations. Put simply, in the early stages of language formation, community members may employ two potential strategies in order to successfully interact with each other: memorize each other's unique lexical variants, or try to align on a shared language. Importantly, the efficacy of these strategies and the ease with which they can be employed will vary in different population contexts. Members of small communities are better able to keep track of each other's variants, allowing them to successfully communicate with each other without the need to converge on a single variant at all. In contrast, such a strategy is much harder to maintain in larger groups with many more individuals' variants to keep track of. Members of larger communities are therefore under a stronger pressure to reduce variability and converge on a shared lexical form. We hypothesize that when memory constraints are taken into account, rather than lexical convergence proceeding more rapidly in small populations, there will be situations where small populations preserve high levels of variability for longer - explaining the data we see in emerging sign languages.

We tested this hypothesis by simulating interaction in populations of language learners. In our model, each individual remembers the lexical variants used by specific individuals they have encountered, but also represents lexical variation in the population as a whole. We analysed the process of conventionalisation on a shared lexical form under the assumption that individuals combine these sources of information using hierarchical Bayesian inference. Under this model, learners draw on individual-specific representations when interacting with somebody familiar, but draw on a population-level generalisation when interacting with a stranger. We made the simplifying assumption that lexical variants can be represented in a one-dimensional continuous space, and that the distributions maintained by individuals can be approximated by Gaussian distributions. We measured conventionalisation (i.e. loss of lexical variation) as the variance in the lexical forms in the population.

Our analysis shows that, in important parts of the parameter space, memory limitations lead to an inverse relationship between population size and lexical variance. Small populations end up with languages that are highly variable, while larger populations converge on a uniform language. This result is modulated by memory size: increasing agents' memory capacity leads to more variance and less convergence overall in both small and large populations. This result supports the idea that simpler (i.e., more compressible) communication systems evolve in the presence of learning bottlenecks ([Kirby, Tamariz, Cornish & Smith, 2015](#)), and supports the hypothesis that convergence in the early stages of language formation is driven by group size ([Meir et al, 2012](#)). Our findings also resonate with the idea that interacting with more strangers is an important factor in driving languages to be systematic and predictable ([Wray & Grace, 2007](#)).