The role of population structure in language evolution

Yoosook Lee* Travis C. Collier Edward P. Stabler Charles E. Taylor
Dept. of EEB Dept. of EEB Dept. of Linguistics Dept. of EEB
University of California, Los Angeles
Los Angeles, CA 90095

Abstract

The question of language evolution is of interest to linguistics, biology and recently, engineering communicating networks. Previous work on these problems has focused mostly on a fully-connected population. We are extending this study to structured populations, which are generally more realistic and offer rich opportunities for linguistic diversification. Our work focuses on the convergence properties of a spatially structured population of learners acquiring a language from one another. We investigate several metrics, including mean language coherence and the critical learning fidelity threshold.

1 Introduction

The question of linguistic divergence is of interest to linguistics[1, 2, 3, 4, 5], biology[6, 7],and engineering communicating networks[8, 9, 10, 11]. For linguists the question is: "What causes languages to change[12, 13, 14], and why do humans have so many different languages?[15, 16, 17]". From an engineering point of view, how to achieve convergence to a single language in a distributed adaptive system[18, 19] is an important issue, as in adversarial conditions, where we would like to maintain high coherency among "friendlies" with minimal understanding from the adversary.

More generally, the dynamics of language evolution provides insight into convergence to a common understanding where distributed learning is a goal. At a theoretical level, these issues are fundamentally similar. The evolution of language takes on special importance for robotics and artificial life because it provides a superb platform for studying the emergence of united behavior from distributed, separate agents.

Previous work on these problems has focused mostly on a fully-connected population where all individuals have an equal probability of learning from each other and the fitness contribution of language is evaluated using the frequency among the entire population[20, 21]. We are extending this study to structured populations, which are generally more realistic and offer rich opportunities for diversification. Our work focuses on the convergence properties of a population of learners acquiring a language from one another under different connectivity connectivity conditions, called *topologies*. This approach is motivated in part by studies indicating that whom a person learns language from can heavily influence one's language [22, 23, 24, 25].

Breaking the symmetry that a fully-connected population provided makes finding an analytical solutions much more difficult, though perhaps not impossible. Therefore, we are using simulations to explore the convergence properties of variety of distinct topologies. We compare the topologies on several metrics, including mean language coherence and critical error-threshold. Our results show that topology has a large effect on overall convergence and can create stable multi-language solutions.

The multi-language solutions are a third distinct phase of local convergence between no-convergence ("Tower of Babel") on the one hand, where all languages are represented in roughly equal frequencies, and global convergence ("Lingua Franca"), where a single language and its close variants predominate. In a multi-language solution, the average individual belongs to a neighborhood predominated by a single language, but no single language dominates across the entire population.

In our paper these simulations are described and discussed. Among our conclusions is that local convergence has important implications for developing systems such as sensor networks where adaptive communication between agents in a heterogeneous environment is desirable.

*yoosook@ucla.edu

2 Methods

Our system is constructed with each individual possessing a parameterized grammar, in the principles and parameters tradition, which can be encoded as a sequence of symbols. However, for the baseline tests we report here, each grammar consists of a single symbol and all grammars have the same expressive power and equal distance from each other. This is a necessary simplification to make our results comparable to the analytic results from Komarova *et. al.*[20].

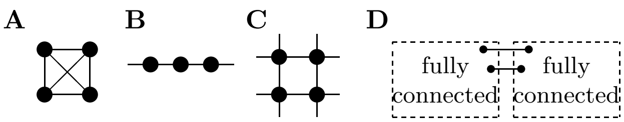

Figure 1: Topologies: (**A**) Fully-connected, denoted FC. The number of connection for each individual n_c is $N-1$. (**B**) Linear, $n_c = 2$. (**C**) A von Neumann lattice with $r = 1$, denoted VN, $n_c = 4$. (**D**) Bridge, which has multiple fully-connected subpopulations and a fixed number of connections between subpopulations.

Each individual exists within a topology defining a set of neighboring individuals. We explore four different topologies: fully-connected (FC), linear, von Neumann lattice (VN), and bridge, illustrated in Figure 1.

The fitness of an individual has two parts: the base fitness, denoted as f_0, and a *linguistic merit* proportional to the probability that the individual could successfully communicate with its neighbors. In the simplified system, linguistic merit is proportional to the number of neighbors which share the same grammar. In the fully-connected topology, each individual of a given grammar will have the same fitness, but this does not hold for other topologies.

Specifically, the fitness of individual i, f_i, is f_0 plus the sum over each neighbor j of the similarity between i's grammar and j's grammar.

$$f_i = f_0 + \frac{1}{2} \sum_{j=1}^{n_c} (a_{ij} + a_{ji}) \tag{1}$$

Each time step, an individual is chosen proportional to its fitness to reproduce. Reproduction can be thought of as the chosen individual producing an offspring which inherits the parent's grammar and replaces one of the parent's neighbors. The offspring learns the parent's grammar with a certain *learning fidelity*, q. This learning fidelity is properly a function of the specifics of the learning method the child uses and the complexity of the grammar, but in the simplified system the learning fidelity is reducible to a transition probability function between grammar G_i and grammar G_j equal to q for $i = j$, and $(1-q)/(n-1)$ for $i \neq j$.

The algorithm of our program is as follows:

```
for each individual i in a population P
    set a random language L_i of i
end for
for each individual i ∈ P
    compute fitness f_i of i
end for
do until number of updates is met
    select an individual k ∈ P
    select a random neighbor j of individual k
    replace the neighbor j with an offspring of individual k
        the offspring becomes an individual j
        if the offspring is mutant( mutation rate = μ)
          get a random language for L_j
        else
          Lj = L_k
        end if
        update fitness of the individual j
end do
```

One important metric is the *dominant* grammar frequency. We measure this directly each time step by counting the abundance of each grammar. Which grammar is the dominant one may change each time it is measured; in other words, the dominant grammar is whichever grammar happens to be at the highest frequency at the time.

The linguistic coherence, denoted as ϕ, is measured using the following equation:

$$\phi = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{2} \sum_{j=1}^{n_c} (a_{ij} + a_{ji}) \tag{2}$$

Various different "levels" of coherence exist as defined by the set of individuals in n_c the second summation occurs over. *Local coherence*, ϕ_0, only sums over the neighbors of each individual and is proportional to mean fitness (equal if $f_0 = 0$). ϕ_1 is the coherence measured over the set of neighbor's neighbors, and generally, ϕ_i is measured using the set of (neighbor's)^i neighbors. *Global coherence*, ϕ_∞, corresponds to summation is over the entire population. In the fully-connected topology, all of these convergence levels reduce to the same value.

For the experiments, we used a population size N of 500, except for the von Neumann lattice which was a 22×22 torus giving a population size of 484. The similarity of between languages a was set at .5, the base fitness f_0 was 0, and the number of different possible grammars n was 10. All relevant parameters are summarized in Table 1.

Parameter	Topologies FC	Linear	VN	Bridge
a		0.5		
f_0		0		
n (# grammars)		10		
N (pop size)	500	500	484	500
# subpops				2
subpopsize				250
# connections				10
# time steps	10^5	10^6	5×10^5	10^5

Table 1: Parameters. Note that 10 connections of bridge topology are randomly selected from each subpopulation.

The experiments, or runs, are done for a set number of time steps that varies with topology. The goal is to make each run long enough that the system will very probably reach an equilibrium. A set of 5 replica runs, varying only the random number generator seed, were done at each q value between 0.65 and 1 at 0.01 intervals.

3 Analytic Model

For the fully-connected topology given a uniform similarity a between n different grammars, and the learning fidelity of q, three equilibrium solutions for grammar frequency were derived by Komarova *et. al.*[20]:

$$X_0 = 1/n \tag{3}$$

$$X_{\pm} = \frac{(a-1)(1+(n-2)q) \mp \sqrt{D}}{2(a-1)(n-1)} \tag{4}$$

where

$$D = 4[1+a(n-2)+f_0(n-1)](1-q)(n-1)(a-1) + (1-a)^2[1+(n-2)q]^2$$

Below a specific learning fidelity q_1, D is negative and there is no real solution for $X_{\pm}$. Therefore, for $q < q_1$, only the symmetric solution X_0 exists and no grammar dominates. Solving for q when $D = 0$ determines the critical leaning fidelity threshold q_1, which corresponds to the *error threshold* in molecular evolution.

$$q_1 = \frac{4 - 2f_0(n-1)^2 - 3n - a(2n^2 - 7n + 6)}{(1-a)(n-2)^2} + \frac{2(n-1)^{\frac{3}{2}}\sqrt{1+f_0[1+a(n-2)+f_0(n-1)]}}{(1-a)(n-2)^2} \tag{5}$$

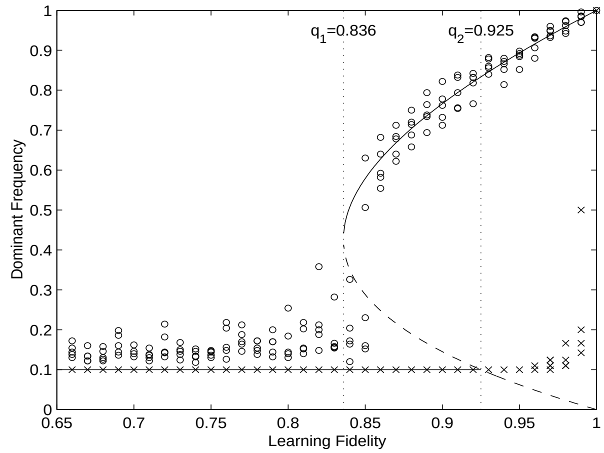

Figure 2: The dominant(○) and average(×) grammar frequency at the last time step of a set of fully-connected runs, overlaid with symmetric (horizontal line) and asymmetric (curved line) analytic solutions for $a = 0.5$, $n = 10$, $f_0 = 0$.

When $q_1 < q < q_2$ for a specific q_2, both the symmetric $X_{\pm}$ and asymmetric X_0 solutions exist and are stable. For $q > q_2$ however, only the asymmetric solution where one grammar dominates the population is stable. This q_2 value is the point where $X_0 = X_-$, giving:

$$q_2 = \frac{n^2(f_0+a) + (n+1)(1-a)}{n^2(f_0+a) + 2n(1-a)} \tag{6}$$

Komarova *et. al.* provide much more detail and proofs[20]. We plot these solutions and compare them to experimental results in Figure 2.

4 Results

The empirical results for the fully-connected topology well match the expectation from the analytic results arrived at by Komarova *et. al.*[20], as shown in Figure 2. In the region where only the symmetric solution is stable ($q < q_1$), the average grammar frequency is $1/n$. The dominant grammar frequency appears high because it is the upper end of a distribution of grammar frequencies which has a non-zero variance due to the finite population size.

In the bi-stability region ($q_1 < q < q_2$), a discrepancy between the analytic and empirical results presumably derives from a lack of runs settling at the symmetric solution. With a finite population, the basin of attraction of the symmetric solution in this region is very weak. Choosing which individual reproduces each

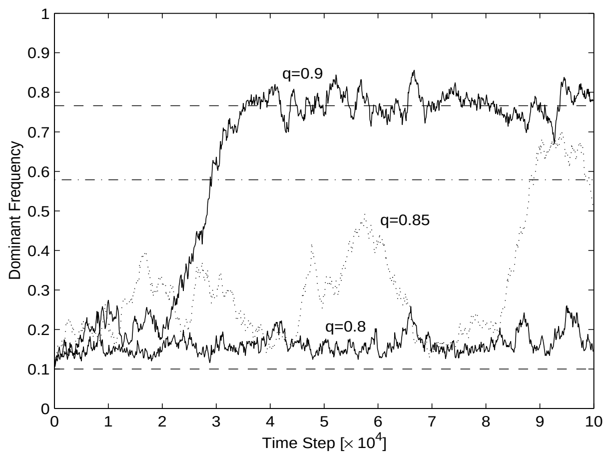

Figure 3: Time-series of fully-connected single runs. The dash line(--) is X_+ for $q = 0.9$, the upper dot-dash line(·-) is X_+ for $q = 0.85$, the lower dot-dash line is X_0. When $q = 0.8$, only the X_0 is stable.

time step is stochastic. This combined with stochastic learning errors appear to be sufficient perturbation to make the symmetric solution unstable empirically in this region.

The time series of single runs with three different learning fidelities in the fully-connected topology are shown in Figure 3. These learning fidelities correspond to the three different regions of stable solutions. The run in the region where symmetric and asymmetric solutions are possible shows the very weak attraction of the symmetric solution. Even starting with every individual in the population being initialized to the same grammar, a dominant frequency of 1, runs at this learning fidelity settle into a similar pattern (data not shown).

For topologies other than fully-connected, convergence provides a more clear picture of system dynamics than dominant frequency. Global coherence ϕ_∞ correlates very closely with dominant frequency, but local coherence ϕ_0 corresponds directly to the linguistic contribution to fitness and is directly operated on by the evolutionary process.

Figure 4 shows the coherence values by taking the average values at the end-points of the 5 replica runs at each q value. The learning fidelity threshold, or error-threshold, for the emergence of a dominant grammar is where indicated by the inflection point in the coherence curve. The emergence of a dominant universal grammar among the entire population is reflected in the global coherence curve.

The bridge topology with 2 subpopulations of size 250 and 10 random connections between subpopulations (Figure 4 B) is indistinguishable from two isolated fully connected populations. It demonstrates a very similar learning fidelity threshold, $q_1 \approx 0.84$, as the fully-connected run shown in panel A. However, global convergence is never achieved reliably in excess to the probability that both subpopulations individually converge to the same grammar by chance. The up-tick in the ϕ_∞ line at $q = 0$ is not statistically significant due to this effect. Additionally, ϕ_1 is extremely close to ϕ_0 while ϕ_∞ only rises to approximately 0.5, indicating that there is a large degree of separation between many individuals.

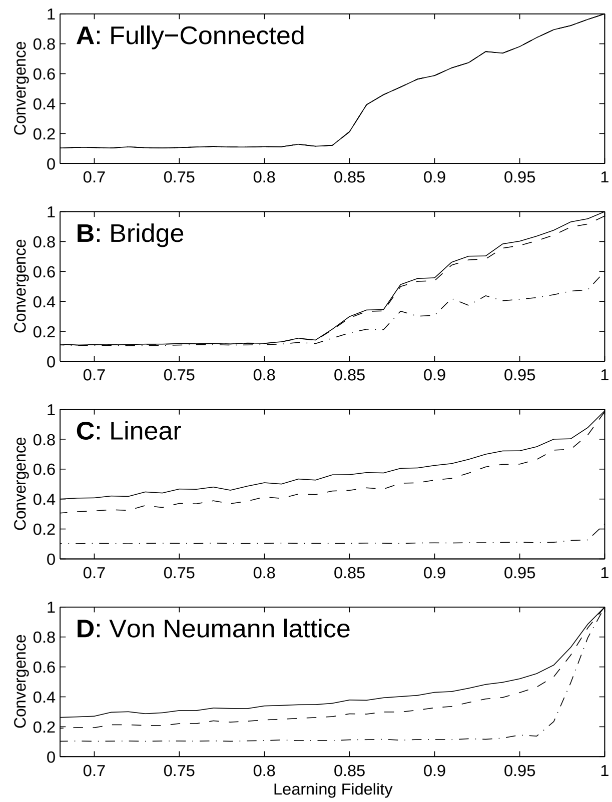

Figure 4: Linguistic coherence. The solid line(–) for ϕ_0, the dash line(--) for ϕ_1, and the dot-dashed line(·-) for global coherence ϕ_∞.

For the linear topology shown in Figure 4 C, ϕ_0 and ϕ_1 slowly rise over the entire range shown, and the trend extends all the way from $q \approx 0.2$ where learning error makes each offspring essentially random (data not shown). Since ϕ_1 trends upward along with ϕ_0, we can conclude that extended "patches" of individuals with the same grammar form. Near $q \approx 0.99$, ϕ_1 begins to approach ϕ_0, and ϕ_∞ shows a slight uptick. However, even with perfect learning fidelity, ϕ_∞ is only slightly different from the symmetric solution

$1/n$. There appears to be no possible global convergence learning fidelity threshold for this topology.

In contrast to the linear topology, the toroidal von Neumann lattice topology (VN) shows a clear learning fidelity threshold for all three coherence metrics at $q \approx 0.96$. Below this threshold, the VN topology behaves similarly to the linear topology with an expected decrease in ϕ_0 and ϕ_1 due to the doubling of neighbors.

5 Discussion

Empirical results using agent-based simulations closely match the analytic results produced by Komarova, Nowak, and others for the fully-connected topology. However, a relatively small population size combined with stochastic scheduling and learning errors lead to sufficient perturbations that empirical results show less stability than the pure math would suggest.

The instability of the fully-connected model at learning fidelities just above the critical threshold q_1 tempts the conclusion that human languages exist in this "edge of chaos" region. Humans exist in complex social connectivity networks, which are probably closer to the bridge topology. The instability of human language is more probably related to changing connectivity and topology than a specific learning fidelity.

Topologies other than fully-connected can behave quite differently. The bridge topology for the parameters we have tested quickly and stably converges to two independent grammars, one for each subpopulation, above a critical learning fidelity. A linear topology fails to converge to a single dominant grammar, but does converge to many "patches" that increase in size as q increases. Both of these cases correspond to stable multi-language solutions which do not exist in the fully-connected topology.

Additionally, the bridge topology has parameters that quite likely change its dynamics. At a higher number of connections between subpopulation and/or a higher number of smaller subpopulations, there is probably a global coherence threshold.

The lack of apparent learning fidelity threshold for the linear topology and similarity to the behavior of the VC topology below its threshold suggests that there is a critical connectivity value that a regular lattice must exceed before global convergence is possible. This result, while only hinted at here, would fit very well with percolation theory. Percolation theory may also provide insight into what parameterized random graph topologies a learning threshold exists for.

The fully-connected topology provides the scenario with the lowest critical learning fidelity, but it also requires the most intensive communication. A topology with much more limited connectivity such as a lattice or clustered graph may still globally converge with much more limited communication. For many engineering situations such as adaptive sensor networks, this is an important consideration.

Language in this study is sufficiently abstract that these results apply to many situations where agents adapt by learning from one another and convergence is desirable. In an adaptive sensor network setting, it may be beneficial for sensor nodes to adapt their communicative coding and recognition/detection systems based on the specific topology of deployment and the actual inputs to the network. An evolutionary strategy where nodes adopt successful schemes from their neighbors with a fitness bonus for agreement is a general option with great promise. Such a system maps directly onto the linguistic systems we present.

6 Summary

We demonstrated the role of topology is critical in determining the degree of linguistic coherence and the learning fidelity threshold through empirical studies informed by the theoretical results for an idealized population. The reality of complex population structure makes evident the importance of topology in studying the dynamics of language acquisition and language evolution. Further investigation on various topologies with different parameter settings may provide a more in depth understanding of language evolution and diversification.

Acknowledgements

This work was supported by the UCLA Center for Embedded Network Sensors, the Defense Advance Research Projects Agency (DARPA), administered by the Army Research Office under Emergent Surveillance Plexus MURI Award No. DAAD19-01-1-0504, and DARPA MURI award administered by the US Airforce No. F49620-01-1-0361. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the sponsoring agencies.

References

[1] S. Kirby. Fitness and the selective adaptation of language. In J. R. Hurford, M. Studdert-Kennedy, and C. Knight, editors, *Approaches to the Evolution of Language: Social and Cognitive Bases*, Cambridge, 1998. Cambridge University Press.

[2] K. Smith. Natural selection and cultural selection in the evolution of communication. *Adaptive Behavior*, 10(1):25–44, 2002.

[3] G. Jäger. Evolutionary game theory and linguistic typology: a case study. In *Proceedings of the 14th Amsterdam Colloquium*. ILLC, 2003.

[4] M. A. Nowak, D. C. Krakauer, and A. Dress. An error limit for the evolution of language. *Proc. Roy. Soc. Lond. B.*, pages 2131–2136, 1999.

[5] N. Grassly, A. von Haesler, and D. C. Krakauer. Error, population structure and the origin of diverse sign systems. *J. theor. Biol.*, 206:369–378, 2000.

[6] S. Nowicki, W. A. Searcy, M. Hughes, and J. Podos. The evolution of birdsong: male and female response to song innovation in swamp sparrows. *Animal Behaviour*, 62:1189–1195, 2001.

[7] W. J. Sutherland. Parallel extinction risk and global distribution of languages and species. *Nature*, 423:276–279, 2003.

[8] T. C. Collier and C. E. Taylor. Self-organization in sensor networks. *Journal of Parallel and Distributed Computing*, 64(7), 2004.

[9] D. Livingstone and C. Fyfe. Modelling the evolution of linguistic diversity. In J. N. D. Floreano and F. Mondada, editors, *ECAL99*, pages 704–708, Berlin, 1999. Springer-Verlag.

[10] L. Steels and F. Kaplan. Stochasticity as a source of innovation in language games. In C. Adami, R. Belew, H. Kitano, and C. Taylor, editors, *Proceedings of Artificial Life VI*, pages 368–376, Cambridge, MA, June 1998. The MIT Press.

[11] T. Arita and C. E. Taylor. A simple model for the evolution of communication. In L. Fogel, P. J. Angeline, and T. Bäck, editors, *The Fifth Annual Conference On Evolutionary Programming*, pages 405–410. MIT Press, 1996.

[12] A. Fox. *Linguistic Reconstruction*. Oxford University Press, Oxford, UK, 1995.

[13] H. H. Hock and B. D. Joseph. *Language History, Language Change, and Language Relationship: An Introduction to Historical and Comparative Linguistics*. Mouton de Gruyter, Berlin, 1996.

[14] A. M. S. McMahon. *Understanding Language Change*. Cambridge University Press, Cambridge, UK, 1 edition, 1994.

[15] M. Ruehlen. *The Origin of Language*. John Wiley & Sons Inc., New York, 1994.

[16] L. L. Cavalli-Sforza, P. Menozzi, and A. Piazza. *The History and Geography of Human Genes*. Princeton University Press, 1994.

[17] L. L. Cavalli-Sforza. gene, peoples and languages. *Proc. Natl. Acad. Sci. USA*, 94:7719–7724, 1997.

[18] L. Steels. Language games for autonomous robots. *IEEE Intelligent systems*, pages 17–22, October 2001.

[19] H. Yanco and L. A. Stein. An adaptive communication protocol for cooperating mobile robots. In J.-A. Meyer, H. Roitblat, and S. Wilson, editors, *From Animals to Animats 2: Proceedings of the Second International Conference on the Simulation of Adaptive Behavior*, pages 478–485. The MIT Press/Bradford Books, 1993.

[20] N. L. Komarova, P. Niyogi, and M. A. Nowak. Evolutionary dynamics of grammar acquisition. *Journal of Theoretical Biology*, 209(1):43–59, 2001.

[21] M. A. Nowak, N. L. Komarova, and P. Niyogi. Evolution of universal grammar. *Science*, 291:114–118, 2001.

[22] S. Pinker. *The Language Instinct*. HarperPerennial, New York, NY, 1995.

[23] A. Senghas and M. Coppola. Children creating language: How nicaraguan sign language acquired a spatial grammar. *Psychological Science*, 12:323–328, 2001.

[24] J. Kegl, A. Senghas, and M. Coppola. Creation through contact: Sign language emergence and sign language change in nicaragua. In M. DeGraff, editor, *Language Creation and Language Change:Creolization, Diachrony, and Development*, pages 179–237. MIT Press, Cambridge, MA, 1999.

[25] L. Steels. Social language learning. In M. Tokoro and L. Steels, editors, *The Future of Learning*, pages 133–162, Amsterdam, 2003. IOS Press.