Emergence of Cohesion in the Evolutionary Dynamics of Grammar Networks

Yoosook Lee\textsuperscript{a}, Travis C. Collier\textsuperscript{b}, Charles E. Taylor\textsuperscript{c}, Reza Olfati-Saber\textsuperscript{d}

Received \ldots; accepted \ldots

This manuscript includes 5 figures and 0 tables.

\textsuperscript{a} yoosook@ucla.edu
TEL: 1-310-206-4079
FAX: 1-310-206-3987
Dept. Ecology and Evolutionary Biology, UCLA
621 Charles Young Dr. South (LS 3109)
Los Angeles, CA 90095

\textsuperscript{b} travc@taylor0.biology.ucla.edu
\textsuperscript{c} taylor@biology.ucla.edu
\textsuperscript{d} olfati@dartmouth.edu
Thayer School of Engineering
8000 Cummings Hall, Room 106
Hanover, NH 03755
ABSTRACT

Language provides a powerful instance of cultural evolution where individuals learn from each other and convergence to functionally similar states is advantageous. Recent studies exploring the evolutionary dynamics of populations of language learners have shown promising results regarding linguistic convergence and diversity. These studies often assume a set of equally related languages. In this paper, we deviate from this assumption and assume that some languages are more closely related than others; in particular we explore the population dynamics of grammar acquisition where the similarity between grammars is described by a ring lattice network. We specify the conditions for the emergence of linguistic coherence in terms of learning fidelity and network structure. In agreement with previous studies, our model exhibits a bifurcation from the symmetric state, where all grammars are in equal frequency, to an asymmetric state under dense grammar networks. However, different patterns of convergence, including a weak cohesion phase where the distribution of grammar frequencies is roughly Gaussian and centered around the most frequent grammar $g$, emerge in more sparse grammar networks. Weak cohesion is of particular interest from an adaptation standpoint since there is both linguistic coherence and standing heritable variation. In technological applications where agents learn from each other and it is desirable for the overall system to converge, our results may provide a guide to designing properties of the language or state representation depending on the desired degree of convergence. Weak cohesion provides a regime where overall functional convergence is relatively high, yet further adaptation is not hindered by low heritable variation.

Keywords: evolutionary dynamics, grammar acquisition, language evolution, grammar
network, adaptive system
1. Introduction

The human ability to think and communicate in expressive symbolic languages distinguishes us from any other species. The emergence of language is one of the truly major events in evolutionary history. Language enables a new mode of evolution by allowing much broader and more flexible transmission of heritable information (Smith & Szathmary 1999; Steels 2004; Jablonka & Lamb 2005).

The evolutionary dynamics of language takes on special importance for robotics and artificial life because it provides a superb platform for studying the emergence of united collective behavior from distributed, separate interacting agents. For the last decade, great progress has been made in the field of robotics to investigate the social and/or cognitive factors contributing the emergence of coherent communication among autonomous agents (Arita & Taylor 1996; Steels 2001; Marocco et al. 2003; Lee et al. 2004; Collier & Taylor 2004; Marocco & Nolfi 2006; Steels 2006).

How to achieve and maintain mutual intelligibility in a distributed adaptive system is an important issue, as in adversarial conditions, where it is advantageous to maintain high coherency among “friendlies” with minimal understanding from the adversary (Yanco & Stein 1993; Lee et al. 2004; Ficici & Pollack 1998). This is especially important in the context of heterogeneous sensors or robots collectively learning about their environment where different sensing modalities makes adaptive symbolic communication vital. 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.

In this study, we address the dynamics of a population of language learners as a less abstract instantiation of the general case of agents learning from one another. Grammars may be viewed as any state or knowledge base learned or derived from other agents where
Recent studies applying a biological perspective to the evolution of linguistic convergence and diversity have shown promising results (Hashimoto & Ikegami 1995, 1996; Steels 1996; Komarova et al. 2001; Nowak et al. 2001; Kirby 2001). Particularly, Komarova, Nowak and Niyogi have adopted a model based on Eigen and Schuster’s molecular evolution work (Eigen & Schuster 1979; Fontana & Schuster 1987; Eigen et al. 1989) for describing the dynamics of language evolution (Nowak et al. 2001; Komarova & Nowak 2001; Komarova et al. 2001).

Parameters such as the maximum error rate a communication channel can withstand before the agents lose their ability to communicate in a coherent language have been determined for some regimes (Nowak et al. 2001; Komarova et al. 2001). However, most strong results have been based on the convergence dynamics of a population learning languages from a set where each language is equally related to every other language. In contrast, natural languages are organized hierarchically, with language families, subgroups, and even dialects. Naturally some languages are more closely related than others.

In this paper, we define a grammar network to be a graph describing the similarity between grammars. We assume that the grammar network determines the distance of grammars from each other in mutational space as well as their contribution to an individual’s fitness. Such network, therefore, represents a space where only transitions between grammars that are connected by edges are allowed in a single reproduction/learning step. Preliminary results have shown that different grammar networks alone can qualitatively change the dynamics of convergence (Olfati-Saber 2005; Olfati-Saber & Murray 2004; Lee et al. 2005). Since our setting does not have limit on population size or on the number of grammars, our result may not be directly comparable with the independant study by Matsen and Nowak which recently have explored language convergence conditions on a “nearly-regular”
language network (Matsen & Nowak 2004).

Here we explore the population dynamics of grammar acquisition using a grammar network structured as a ring lattice. Although we have also explored other network structure such as small-world and random network, we are compelled to focus on ring lattice because it illustrates how grammar network alone can change the dynamics of language convergence. Two equilibrium states have previously been characterized in terms of language convergence: one is the symmetric state where all grammars exist in equal frequencies, and the other is an asymmetric state where the symmetry breaks and one grammar predominates in the population (Nowak et al. 2001; Komarova et al. 2001). Our results identify another state we label weak cohesion, where a group of structurally similar grammars predominates in the population. The high degree of standing variation characteristic of the weak cohesion state is more suggestive of real-world observations than the other two equilibria previously described. The weak cohesion regime appears desirable for many adaptive system since it achieves a high level of cohesion without eliminating too much variation and losing further adaptability.

1.1. Evolutionary Dynamics of a Population of Language Learners

Consider a network of grammars with nodes $U = \{G_1, G_2, \ldots, G_n\}$. Let $s_{ij}$ denote the probability that a learner acquires a grammar $G_j$ from a teacher who speaks grammar $G_i$. Then, we define the adjacency weights of the grammar network as $a_{ij} = (s_{ij} + s_{ji})/2$. The matrix $A = [a_{ij}]$ is determined by the set of edges $E$ of a particular graph $\Gamma = (U, E)$. In our setting, $A$ is defined as:

$$a_{ij} = \begin{cases} 
1 & \text{if } i = j \\
a & \text{if } (i, j) \in E \\
0 & \text{if } (i, j) \notin E
\end{cases}$$
Each grammar represents a language hypothesis, consisting of a set of rules that an individual deduced from its input. It is possible that two grammars can have a completely different set of rules and yet generate sentences that are somewhat mutually understandable. In order to prevent potential confusion, let us distinguish between two types of similarities, structural similarity and expressive similarity. Structural similarity is how many grammatical rules or lexical items two grammars share. Expressive similarity relates to the probability that a sentence generated from one grammar is similar to a sentence generated from another grammar. Structural similarity is analogous to genotype similarity, and expressive similarity is analogous to phenotype similarity.

In our setting, all grammars in a network are positioned on a polygon where their positions are indicative of structural similarity. If two grammars are positioned side by side, they share many common rules for generating sentences.

The relationship between grammars is represented as edges in a grammar network specifying which learner grammars each teacher grammar may produce. In a molecular evolution framework, this is the graph defining possible mutational transitions. For our model, the grammar network is completely specified by the matrix of mutual similarities $A$. Note that a mutual similarity $a_{ij}$ is equivalent to the expressive similarity of two language hypotheses. As long as two grammars are connected $(i, j) \in E$, they have some degree of mutual intelligibility ($a_{ij} > 0$).

Within this setting, each mutation or incorrect language learning step doesn’t necessarily yield a grammar that is structurally similar to its teacher’s grammar. Sometimes the learner can deduce a set of rules that are completely different from its parent’s and yet generate sentences that are very close to its input. Thus, the grammar network defines the space which an individual explores while learning language.

Let $x_i$ denote the proportion of an infinitely large population speaking grammar $G_i$
with \( n \) possible grammars existing. We assume that each individual uses only one grammar, thus we have \( \sum_{j=1}^{n} x_j = 1 \).

The fitness of individuals with grammar \( G_i \) is \( f_i = f_0 + \sum_j a_{ij} x_j \) where \( f_0 \) is the base fitness which does not depend on the language, and \( \sum_j a_{ij} x_j \) is the language contribution to fitness. Note that this fitness equation is frequency dependent.

The evolutionary dynamics of this population is of the form

\[
\dot{x}_i = \sum_j x_j f_j q_{ji} - \phi x_i, \quad 1 \leq j \leq n
\]  

(1)

where \( \phi = \sum_i x_i f_i \) is the average fitness, and \( Q = [q_{ij}] \) is the learning fidelity matrix. The term \(-\phi x_i\) maintains a constant population size. Equation (1) is known as the replicator-mutator dynamics in the generally accepted framework inspired by Eigen and Schuster (Eigen & Schuster 1979). \( Q \) is basically a Markov chain transition matrix, which is used as a learning fidelity matrix here and also by Nowak, Komarova (Komarova et al. 2001) and others. This dynamic system can be thought of either as having individuals that produce offspring to replace a randomly chosen individual, or as having individuals that change their grammars by learning a teacher’s language. We will use the latter terminology.

The learning model is a relationship between the matrix of mutual similarities \( A \) and the matrix \( Q \), which is defined by

\[
q_{ii} = q, \quad q_{ij} = (1 - q) \frac{a_{ij}}{\sum_{j \neq i} a_{ij}} \text{ for all } i \neq j.
\]  

(2)

The learning fidelity \( q \) is the probability that a learner acquires the same grammar as its teacher. \( Q \) satisfies the condition \( \sum_j q_{ij} = 1 \) for all \( i \). The special case of this transition matrix where \( a_{ij} = a \) for all \( i \neq j \) was analyzed by Komarova et.al. (Komarova et al. 2001; Komarova & Nowak 2001).
1.2. Grammar Networks

Graph theory provides a large number of metrics for describing the topology of grammar networks (Olfati-Saber & Murray 2004; Olfati-Saber 2005). The two we will focus on are density and the mean path length. Density $D$ is the measure of the number of links in the graph divided by the number of possible links (Albert & Barabasi 2002). Mean path length is the mean number of links in the graph that must be traversed to connect any pair of nodes.

For graphs of differing density and mean path length, we expect the dynamics to be different since a change in $A$ also changes $Q$ as well as fitness. Less than complete graphs make sense from a biological or sociological standpoint since learning a closely related language is more common than learning a drastically different language.

We systematically analyze the effects of learning fidelity $q$ on the evolutionary dynamics defined in equation (1) for two canonical types of graphs with substantially different structural properties. These are the complete graph and regular ring lattice graphs, depicted in Fig.1.

![Fig. 1.— Two types of network structure tested in this paper. Examples are drawn with 12 nodes. A the complete graph, B a ring lattice with $k = 2$.](image-url)
1.2.1. The Complete Graph

The complete graph with all-to-all links is considered in Komarova et al. (Komarova et al. 2001; Nowak et al. 2001) because the graph provides a simplified setting where finding an analytic solution is relatively tractable (example shown in Fig.1A). The adjacency matrix $A$ with similarity parameter $a > 0$ is defined by:

$$a_{ij} = \begin{cases} 
1 & \text{if } i = j \\
a & \text{if } i \neq j
\end{cases}$$

The total number of links for complete graph is $n(n - 1)/2$, and the density is 1. In terms of grammar networks, the complete graph is one in which every grammar has equal mutual intelligibility with every other grammar and an individual can change its grammar to any other grammar with a single step – evident from mean path length equal to 1.

1.2.2. Regular Ring Lattice

A regular ring lattice graph is a cyclic graph with $n$ nodes evenly spaced on a ring and each node linked to its $2k$ nearest neighbors on the ring (example shown in Fig.1B). This type of graph is also known as a one-dimensional lattice. The adjacency matrix of this graph is denoted by

$$A = I_n + aC(n,k)$$

where $I_n$ denotes an $n \times n$ identity matrix. $C(n,k)$ is a $n \times n$ matrix representing a regular ring lattice, where

$$c_{ij} = \begin{cases} 
1 & \text{if node } i \neq \text{ node } j \text{ and } \\
i \text{ and } j \text{ are separated by less } k \text{ vertices on the polygon defining the ring} \\
0 & \text{otherwise}
\end{cases}$$
Fig. 2.— The dominant grammar frequency, $x_{\text{max}}$ vs learning fidelity $q$ for a ring lattice with $n = 51$. **A**: $k = 25$ (a compete graph) **B**: $k = 15$ **C**: $k = 10$. The $q$ interval is $10^{-4}$.

A ring lattice has a total of $nk$ links. The density of this class of graphs is $2nk/n(n - 1) = 2k/(n - 1)$, which for $2k \ll (n - 1)$ is much less than 1. For the special choice of $n = 2k + 1$, one obtains a complete graph with the maximum density of 1.

In a ring lattice network, the mutual similarity can infer some structural similarity in the set of rules. In order for a learner to acquire a grammar that is very different from its teacher’s, the learner has to undergo many mutational steps. This trend is reflected by the large mean path length, which is proportional to $n/k$ on a ring lattice (Albert & Barabasi 2002).

### 2. Results

In Fig. 2 we show the equilibrium frequency of the dominant grammar over a range of $q$ values for a regular lattice graph with fixed $n = 51$ and $\alpha = 0.5$. Each point is the result from a single run, but the $q$ interval ($= 10^{-4}$) is small enough that the points appear as a line in places. The symmetric solution where all grammars exist in equal frequencies of $1/n$ is a stable attractor for $q \leq q_s$. The asymmetric state where single predominant grammar emerges become a stable attractor above a critical $q$ value of $q_a$. Fig. 2**A** shows the bi-stability region, where both symmetric and asymmetric solutions are stable, in agreement
Fig. 3.— The effective number of grammars, $n_e$ vs $q$: A: $k = 25$ (complete graph), B: $k = 15$, C: $k = 10$.

with the results analyzed by Komarova et al. (Komarova et al. 2001; Nowak et al. 2001). In a complete graph, $q_s$ is greater than $q_a$ allowing a bi-stability region where both symmetric and asymmetric solutions can be stable depending on the initial frequency distribution.

The ring lattice graph with $k = 15$ also has symmetric and asymmetric solutions as shown in Fig.2B. For the ring lattice network, the symmetric state breaks at $q_s$ well before asymmetric solutions can be stable. The symmetric state is not observed for the ring lattice network with $k = 10$ as shown in Fig.2C.

The learning fidelity threshold for asymmetric solutions, $q_a$, is the highest for the complete graph and it is lower in the networks with smaller $k$ values ($q_{a,k=25} = 0.9720 > q_{a,k=15} = 0.8280 > q_{a,k=10} = 0.7956$).

Fig. 3 shows the effective number of grammars for precisely the same runs as Fig.2. In agreement with the results shown in Fig.2A, Fig.3A shows two classes of equilibrium states as well as bi-stability region over a range of learning fidelities $q_a \leq q \leq q_s$ for the complete graph. We calculated the effective number of grammars $n_e$ (See Methods section for its definition) as a measure of grammar diversity. The $n_e$ is the maximum ($= n$) when a population is in symmetric state, while $n_e$ is the minimum ($\simeq 1$) when a single dominant grammar emerges. The value of $n_e$ is indicative of the degree of linguistic coherence.
Fig. 4.— Frequency of all grammars, sorted by index number, of a single run at the steady
state for a ring lattice with $n = 51$, $k = 15$, and $a = 0.5$. A: $q = 0$ B: $q = 0.82$ C: $q = .83$

Fig.3B suggests a third class of solutions which occurs at $q$ values between $q_s$ and $q_a$
for ring lattice networks. This class of stable attractors is characterized by a nearly linear
decrease in $n_e$ when $q_s \leq q \leq q_a$, as shown in Fig.3B.

For the regular ring lattice with $k = 10$, the symmetric state does not exist for $q \geq 0$
as shown in Fig.3C. The symmetric solution can still be obtained using negative $q$ values,
but the interpretation of such values is not obvious. Fig.3C also shows that some level of
coherence can be achieved even with the learning fidelity of 0 when $k = 10$.

In the regular ring lattice with $k = 15$, the frequencies of each grammar $x_i$ at
approximate equilibrium state for different learning fidelities are distributed as shown in
Fig.4. We cut the ring at the point opposite to the dominant grammar and spread it
along the x axis, so that the dominant grammar is always at the center. If the grammar is
positioned close to the dominant, the grammar index is close to the index of the dominant,
indicating that they are structurally similar.

Fig.4A shows the grammar frequency distribution when $q = 0$ as an example of a
symmetric state. If the learning fidelity $q$ is greater than $q_a$, only one grammar dominates
as shown in Fig.4C. We call this phase strong cohesion. When the learning fidelity is
between $q_s$ and $q_a$, the grammar frequencies form a smooth curve reminiscent of a Gaussian
Fig. 5.— Non-linear behavior of language convergence in response to network density. Various measures such as A grammar diversity $n_e$, B dominant frequency $x_{max}$, and C minimum frequency $x_{min}$ are plotted for varying $k = [1, 100]$ given a ring lattice with $n = 200$, $a = 0.5$, and a learning fidelity $q = 0.8$.

as shown in Fig.4B. We call this phase weak cohesion. In this phase learning fidelity is too low for a single grammar to dominate by faithfully reproducing itself, however structure in grammar space allows for a collection of closely structurally similar grammars to rise in frequency. Since learning errors produce similar grammars to the teacher’s grammar, the effective learning fidelity for the group of grammars is higher. This is analogous to the formation of a quasi-species in molecular evolution (Eigen & Schuster 1979).

Fig.5 shows the effect of graph density on the level of coherence. We plotted (a) grammar diversity, (b) dominant grammar frequency, and (c) minimum grammar frequency for a ring lattice with $n = 200$ and $a = 0.5$ given a fixed learning fidelity. Notice that the learning fidelity we used ($q = 0.8$) is smaller than $q_a = 0.829$ for a complete graph. The grammar diversity and dominant frequency changes in non-linear fashion in response to network density.

When a grammar network forms a complete graph ($k = 100$), the population is in the symmetric state ($n_e = 200$) as expected. When the density of a ring lattice is sufficiently high ($k \geq 83$), the population is also in the symmetric state, and both $x_{max}$ and $x_{min}$ are
equal to $1/n = 5 \times 10^{-3}$ as expected.

As the density of the graph decreases, grammar diversity $n_e$ decreases reflecting an increased level of linguistic coherence, indicating a weak cohesion state. When the network density is sufficiently small ($k \leq 10$), the grammar diversity is close to 1 and the dominant frequency is over 0.5 which is indicative of strong cohesion.

3. Discussion

The dynamics of the evolutionary system defined by equation (1) are characterized by three possible equilibrium states: (1) the symmetric state ($q \leq q_s$) where $x_i = 1/n$ for all $i$, (2) a weak cohesion state where the symmetry in grammar frequencies breaks and the distribution of grammar frequencies forms roughly Gaussian shape centered around the most frequent grammar, and (3) a strong cohesion state ($q \geq q_a$) where a single predominant grammar emerges.

If a grammar space is structured as a complete graph, high fidelity learner-driven change, such as the sort exhibited by human languages, can only occur just above the critical error threshold $q_a$ where the bifurcation of asymmetric and symmetric states begins. Additional empirical results from various grammar networks, data not shown, indicates that this threshold is highest for the complete grammar network.

For a grammar space defined by a mid to low density regular ring lattice, a weak cohesion phase of equilibria can be identified at learning fidelities between $q_s$ and $q_a$. This region is below the error threshold for a complete graph where no cohesion or evolution can take place. The existence of a weak cohesion phase is dependent on structure in grammar space allowing the formation of a quasi-species of related grammars. While the learning fidelity in this region is too low to allow a single grammar to faithfully reproduce itself well
enough to maintain a higher frequency than other grammars, the effective learning fidelity of the quasi-species as a whole is sufficiently high for the collection of grammars within it to dominate.

The existence of the weak cohesion phase suggests that a group of related grammars can emerge with lower learning fidelities than is required to establish a single consensus. Weak cohesion is also characterized by a large amount of standing heritable variation within the population which is particularly intriguing from evolutionary perspective.

Our results varying $k$ indicate that for regular ring lattice graphs, as the density of the grammar network decreases the level of equilibrium linguistic convergence increases for a fixed learning fidelity. We expect that this result holds generally for structured graphs, although the mean path length or clustering coefficient may be more indicative metrics of structure in grammar space.

Recognizing the importance of social networks in language learning, individual-based models have attracted attention as an experimental tool for language evolution. In the course of this investigation, we found that linguistic relatedness alone can affect the evolutionary dynamics.

The regular ring lattice is only an example of many different network structures. However, the qualitatively different dynamics with respect to the network density suggests that convergence to a set of closely related languages is significantly easier for a structured grammar space, in contrast to conclusions based only on the fully-connected model.

The linguistic system described here is studied not only as an aspect of biological organisms undergoing evolution but also as an evolutionary system in itself. Situations such as gene flow among subpopulations required for speciation or cohesion into a single population are similar to the problems explored here.
In technological applications where agents learn from each other and it is desirable for the overall system to converge, these results may provide a guide to designing properties of the language or state representation depending on the degree of convergence desired. If it is sufficient that agents of the system just mostly agree, i.e. converge to close variants of a dominant grammar, then a structured state space may provide a way to achieve faster convergence at higher mutation values. However, if absolute convergence is required, the state space must be designed such that minor variants are strongly selected against, producing a sharp fitness peak. This constraint also implies that a critical mutation/learning fidelity threshold exists.

4. Methods

Results are obtained from numerical evaluation of the system described by equation (1), referred to as runs, using the fourth-order Runge-Kutta method (Abramowitz & Stegun 1972). We assume the population is approximately in equilibrium state when the first and second derivatives for all grammar frequencies are below a predefined threshold.

We obtain the approximate equilibrium frequencies of grammars for a given adjacency matrix $A$ associated with a graph $G$, and a random choice of initial state $x(0)$, which satisfies $\sum_i x_i(0) = 1$. From the grammar frequencies we calculate the effective number of grammars $n_e$ as a measure of grammar diversity, which is defined as (Crow & Kimura 1970):

$$n_e = \frac{1}{\sum_i x_i^2}$$

Acknowledgments

We thank Gregory M. Kobele for insightful linguistic interpretation of our model. Special thanks to Prof. Emilio Frazzoli, Prof. Jeff S. Shamma, Prof. Edward P. Stabler. This work was
supported by MURI award administered by the US Airforce No. F49620-01-1-0361, NSF (CNS
0513058, EF-0410438), and the UCLA Center for Embedded Network Sensors. 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


This manuscript was prepared with the AAS \LaTeX\ macros v5.2.