Grammar Structure and the Dynamics of Language Evolution

Yoosook Lee[1], Travis C. Collier[1], Gregory M. Kobele[1], Edward P. Stabler[1],
and Charles E. Taylor[1]

University of California, Los Angeles, CA 90095, USA
yoosook@ucla.edu,
webpage: http://taylor0.biology.ucla.edu/

Abstract. The complexity, variation, and change of human languages makes evident the importance of representation and learning in the acquisition and evolution of language. For example, analytic studies of simple languages in unstructured populations have shown complex dynamics, depending on the fidelity of language transmission. In this study we extend these analysis of evolutionary dynamics to include complex grammars inspired by the principles and parameters paradigm. In particular the space of languages is structured so that some pairs of languages are more similar than others, and mutations tend to change languages to nearby variants. We found that coherence emerges even with low learning fidelity.

1   Introduction

The evolutionary dynamics of language provides insight into the factors allowing subpopulations to converge on common or similar languages. The problem has a more general significance for robotics and artificial life as a clear and empirically supported platform for the study of how coherent behavior can emerge in a population of separate, distributed agents.

Of particular interest from the perspective of evolutionary dynamics are insights into the means and value of conserving linguistic diversity. The practical importance of linguistic diversity has attracted some attention [1, 2], though is perhaps not as so much as for biological diversity. Recent studies that have applied a biological perspective to the evolution of linguistic convergence and diversity have shown promising results [3–9]. Most such studies that apply a biological perspective to language evolution have been based on very simple languages arbitrarily related to each one another. We believe these studies may be enriched by a more realistic description of language.

Language models based on the Chomskian paradigm [10, 11] view language as an aspect of individual psychology. There has been some debate about the extent to which the underlying representation of languages are inherited or learned and of the relation to these to fitness. Pinker and Bloom, for example, suggested that a language instinct constrained by the universal grammar sets the stage for language acquisition which then contributes to individual fitness [12, 13].

Parameters		
N	population size	500
f_0	base fitness	10^{-3}
l	Number of language parameters	1, 2, 3, or 6
d	Number of values per each parameter	64, 8, 4, or 2
n	Number of possible grammars	$d^l = 64$
	Number of time steps	100,000

Table 1. Parameters used in the simulations.

Hauser, Chomsky and Fitch argue more recently that while certain perceptual and articulatory abilities may have been selected, it remains unclear how the most fundamental aspects of human language emerged [14, 15]. All parties agree that language-specific properties are to some extent learned, to some extent culturally transmitted, and in any event change through time. How this might occur has been the subject of many analytic and simulation studies [9, 16, 8, 7, ?].

In this paper, we study how certain properties of the space of possible languages and learning mechanisms could have an impact on language change. Because they both evolve, biological and linguistic systems have many features in common. As an organism is determined in part by its genome, language is determined in part by a lexicon of generators that restrict the phonology, semantics, morphology and syntax of a language; these properties may evolve [17, 18]. As the genome is subject to various mutations during reproduction, language is subject to change during learning and communication.

These similarities lead some researchers like Nowak and Komarova to adopt a quasi-species model [19, 20] for describing the dynamics of language evolution [6, 21]. Extensions of these studies have indicated connections between structured populations and language diversification [22], and suggests that the topology of a population may have a greater role than expected in shaping patterns of language evolution *(Reference removed)*.

In this paper we introduce a simple regularity in grammar space and explore its impact on the dynamics of language evolution. In this setting, the similarity between two languages is measured rather than assigned arbitrarily. These simulations go beyond previous analytic studies of simple models, and we find that structure has a significant impact on stability results.

2 Methods

Consider a fully-connected finite population of N individuals, each of whom possesses a language which is encoded as a sequence of l linguistic 'components' or 'parameters'. Each parameter can take only a limited number d of values. For example, a language L with 10 parameters each taking 3 values (A,B,C) can be represented by a linear sequence like this:

AABABCBABA.

It is important to recognize that this string is not an example of a statement from the language, but rather a represents the language itself. Such a representation is in the spirit of Chomsky's "principles and parameters" theory of language. See Niyogi[9] for a discussion with examples. Representing a language as a sequence, we define the language similarity between individual i and j, denoted a_{ij}, as the proportion of parameters on which the two individuals agree. For example, the language similarity($a_{ij} = a_{ji}$) between an individual i whose language is represented as AAA and an individual j whose language is represented as ABA is 2/3.

The fitness of an individual, f_i, has two parts: the base fitness, denoted f_0, and a *linguistic merit* proportional to the probability that the individual could successfully communicate. A linguistic merit of an individual is proportional to the sum of language similarity between the individual and others it is in linguistic contact with (which is the entire population for this model). In the fully-connected topology, each individual of a given language will have the same fitness. The overall fitness of an individual, f_i, is described as the following, as in [16]:

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

because $a_{ij} = a_{ji}$ according to our definition.

At each time step, an individual is chosen to reproduce, where this choice is made randomly and independently with a probability according to relative fitness. Reproduction can be thought of either as the individual producing an offspring which inherits the parent's language and replaces another in the population, or another individual changing its language to match the "teacher's" language. We will use the former terminology.

The offspring learns the parent's language 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 language, often modeled with a probability distribution over the possible transition from each language L_i to each other (possibly different) L_j. But in the present setting, the only incorrect/imperfect learning allowed is a single parameter change per reproductive event. We refer to this constraint as *gradual learning*. The rationale behind this approach is that learning errors do not typically result in the learner acquiring a radically different language. This single parameter change constraint on incorrect/incomplete learning is analogous to only allowing single point mutations to the linear sequence representation of the language. As such, it defines the "sequence space"[20] through which the population moves through the Darwinian evolutionary process.

We study language change in an ideal population using a simulation, running the following algorithm:

`for each` individual i in a population P

```
        assign i a language L_i chosen randomly from the set of all possible languages ℒ
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 second random individual j from the population
    replace individual j with an offspring of individual k
        the offspring becomes an individual j
        if the offspring is mutant( mutation rate = μ)
          change a random parameter of L_j
        else
          Lj = L_k
        end if
        update fitness of the individual j
end do
```

We measure the *dominant* language frequency directly each time step by counting the number of individuals speaking of each language. The dominant language at any given time is simply the language that is most frequent at that time. This will typically change unless the population has strongly converged.

The linguistic coherence of the population, denoted ϕ, is defined as follows:

$$\phi = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij} \tag{2}$$

Counting the actual number of languages that exist in the population may disguise the degree of variation when some of the languages disproportionately dominate. Consequently, we used an analogue to the effective number of alleles in a population, which we will refer to as the effective number of languages in the population, n_e [23]:

$$n_e = \left(\sum_{i=1}^{N} p_i^2 \right)^{-1}, \tag{3}$$

where p_i is the frequency of each language.

Table 1 shows the parameter settings for the experimental setup. We used a population size N of 500, a base fitness f_0 of 0.001, and we let the number of different possible languages n be 64. Each language in a set can be represented as a linear sequence of length l with elements drawn from a set of d possible values. For set A, the similarity between languages that are not the same is set to a constant value a equal to 0.5. For all other sets, a_{ij} is the Hamming distance divided by sequence length as described above. The reproduction cycle repeated for 100,000 times to make each run long enough to reach an equilibrium. A set of 20 replica runs, varying only the random number generator seed, were done at each q between 0.5 and 1 at 0.02 intervals.

3   Analytic Model

Given a uniform similarity, a, between n different languages, and the learning fidelity of q, three equilibrium solutions, X_0 and $X_\pm$, for language frequency were derived by Komarova *et. al.* [21] for a family of single-component languages:

$$X_0 = 1/n \tag{4}$$

$$X_\pm = \frac{(a-1)(1+(n-2)q) \mp \sqrt{D}}{2(a-1)(n-1)} \tag{5}$$

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 certain learning fidelity of q_1, only the symmetric solution X_0 exists and no single language dominates. Solving for q when $D = 0$ determines the critical leaning fidelity threshold q_1, which corresponds to the *error threshold* in molecular evolution.

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

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 language dominates the population is stable. This q_2 value is the point where $X_0 = X_-$, giving:

$$q_2 = \frac{n^2(f_0 + a) + (n+1)(1-a)}{n^2(f_0 + a) + 2n(1-a)} \tag{7}$$

Komarova *et. al.* provide much more detail and proofs[21].

By introducing a regularity in language, we effectively change the transition matrix of a_{ij}. To compare our findings with the analytical result, we use the average language similarity $\bar{a}$ for calculating q_1 and q_2, where $\bar{a}$ is calculated using the equation below:

$$\bar{a} = \frac{\sum_{k=1}^{l-1} \frac{l-k}{l} (d-1)^k \binom{n}{k}}{d^l - 1} \tag{8}$$

We consider 4 settings A-D, varying in the "amount of structure." The four cases are listed in Table 2 together with the calculated $\bar{a}$ for each case.

Setting	l	d	n($= d^l$)	$\bar{a}$	q_1	q_2
A	1	64	64	0.500	0.830	0.985
B	2	8	64	0.111	0.516	0.902
C	3	4	64	0.238	0.662	0.955
D	6	2	64	0.492	0.826	0.985

Table 2. System settings, average language similarity ($\bar{a}$), q_1 and q_2. When $l = 1$, we use $a = 0.5$. A: one component with 64 options, B: two components with 8 options, C: three components with 4 options, D: 6 components with 2 options. Each setup has exactly the same number of possible languages.

4 Results

We plot the experimental and analytic results for comparison in Figure 1. The empirical results for the uniform similarity of $a = 0.5$ between two different languages closely followed the expectation from the analytic results arrived at by Komarova *et. al.*[21] as shown in Figure 1 A, which we have previously described in detail *(Reference removed)*. In the region where only the symmetric solution is stable ($q < q_{1A}$), the average language frequency stays at the value of $1/n$. The dominant language frequency appears higher than expectation because it is the upper end of a distribution of language frequencies which has a non-zero variance due to the finite population size.

In the bi-stability region ($q_{1A} < q < q_{2A}$), a discrepancy between the analytic and empirical results is manifested in 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 time step is stochastic. This combined with stochastic learning errors appear to be sufficient perturbation to make the asymmetric solution unstable empirically for lower q values in this region.

The results of the multi-component languages (Figure 1 B, C and D) do not show the clear transition from symmetric to asymmetric solution. The trend is considerably smoother, with nothing but an increased variance in results at the point of the phase transition for parameter sets C and D. Parameter set B shows a region where both symmetric and asymmetric solutions appear stable for q values between 0.6 and 0.7, but it is notable that the empirical asymmetric dominant abundance is significantly below the analytical expectation for this set as well as C and D.

Since the setup A and D have similar $\bar{a}$ values ($\bar{a}_A \simeq \bar{a}_D$), they provide a better example of what difference the multi-parameter language brings to the language evolution scenario. Figure 2 compares the number of languages and the effective number of languages (n_e), calculated using the equation (3). In the single-parameter language case, all the possible languages exist in the population in the region where $q < q_{1A}$. On the other hand, the 6-parameter case has only half of the all possible languages at $q = q_{1D}$.
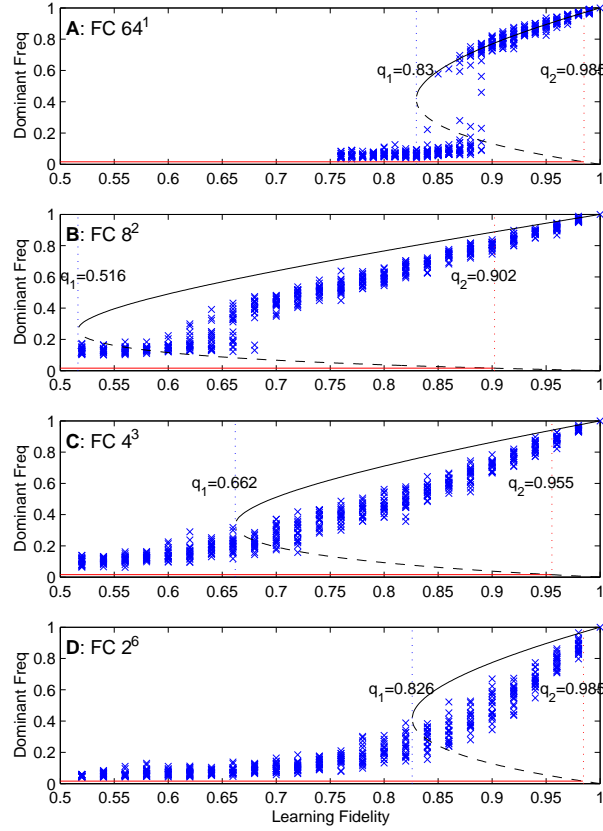
Fig. 1. The dominant($\times$) language frequency after 100,000 time steps overlaid with symmetric (horizontal dotted line) and asymmetric (curved line) solutions for a(or $\bar{a}$), $n = 64$, $f_0 = 0.001$. Each point is an independent replica.

Figure 1 A shows that if the learning fidelity is greater than 0.9, one language dominates in the population. That trend is illustrated clearly by the average effective number of languages in Figure 2 A. There are still over half of all possible languages remaining in the population at $q = 0.9$. This number overestimates the true variation in the population since some languages disproportionately dominate while most are at very low frequency. The effective number of languages n_e for the set A at $q = 0.9$ is close to 1 ($n_e = 1.68$), which indicates that the population has converged to one language, and the rest of languages exist a very low frequency . Incomplete/incorrect learning provides a constant influx of variants, but these variants do not propagate to any appreciable frequency due to their inferior fitness.

In contrast, Figure 2 D shows a gradual decline in number of languages as learning fidelity increases. For this set, the number of languages in the population
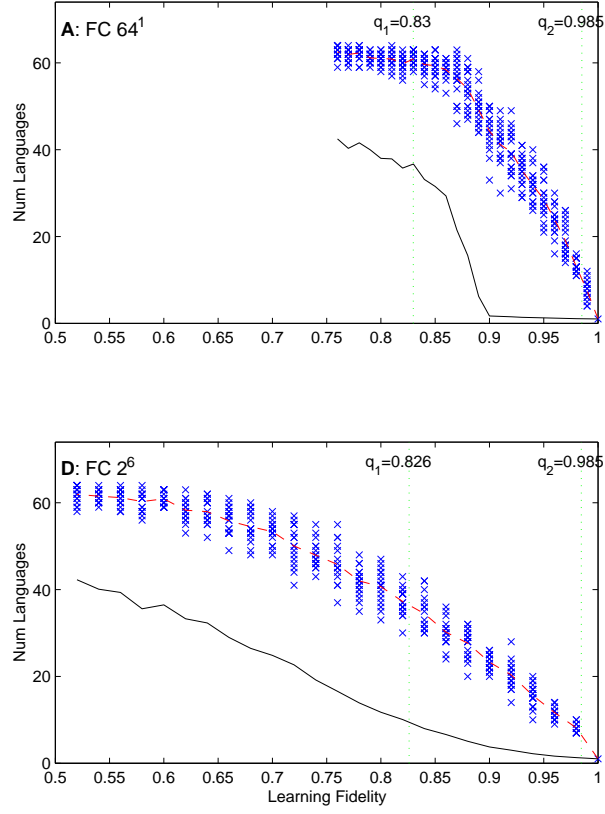
Fig. 2. The number of languages($\times$) and the average effective number of languages(—). The dash line indicates the mean of the number of languages.

starts decreasing noticeably for q values above 0.55, and the effective number of languages n_e decreases over the entire range. However, at q values above 0.9, set D shows a higher n_e value of 3.75 at $q = 0.9$ than set A, indicating that there are more relatively high abundance languages in set D despite the fact that the total number of languages is lower.

In set A, all possible languages are a single step away in sequence space; in other words, all possible languages are reachable by a single incorrect/incomplete learning event. In set D, however, only a small subset of possible languages are producible as single step variants from the dominant language. These single-step variants of the dominant will account for the majority of non-dominant languages in the population. Additionally, these variants will have a high fitness relative $\bar{a}$, and a higher equilibrium frequency in mutation-selection balance.

5 Discussion

For the set of single-component languages, these empirical results closely match the analytic results produced by Komarova *et al.* A finite population size combined with stochastic scheduling and learning errors leads to some level of perturbations. If we want to model language change of the sort exhibited by human languages, we see that with high fidelity learner-driven change could only occur just above the critical error threshold q_1, near the bifurcation point.

These simulations also show that substantial levels of linguistic coherence can be achieved with lower learning fidelity if structure is introduced. All four settings explored here have language spaces of exactly the same size, and yet two of these settings (in particular, B and C) allow fairly stable asymmetric solutions even with quite low learning fidelity. The structured languages show a much more gradual approach to coherence.

We conclude that a simple regularity combined with gradual learning can dramatically reduce the number of languages that exist in the population, even in regions where analytic results indicate that only symmetric solutions will be stable. Gradual learning used in this experiment seems a more realistic approximation to reality than the "memoryless" learning used in previous work. The qualitatively different dynamics with respect to the critical learning fidelity suggests that convergence to a set of closely related languages is significantly easier than previously thought.

These results are in keeping with the expectations of a quasi-species interpretation. Gradual learning maps the grammars into a sequence space where some grammars are fewer mutational (incomplete/incorrect learning) steps from others. Calculating the similarity between grammars which determines fitness as one minus Hamming distance divided by sequence length ensures that grammars that are close in the sequence space have similar fitness values. This produces a smooth fitness landscape.

The upshot of this smooth fitness landscape is that selection operates on the quasi-species formed by the dominant grammar and its close variants. At learning fidelity values below q_1, the population converges not to a single dominant grammar with all other grammars equally represented, but instead to a family of similar grammars. The fidelity of this family is higher than the nominal learning fidelity because a sizable proportion of incomplete/incorrect learning events among members of the quasi-species result in other members of the quasi-species. At still lower q values, that family of grammars (the quasi-species) spreads farther out in sequence space, until at some point it includes all possible grammars and is identical to the symmetric analytical solution provided by Nowak and Komarova's model [6, 21].

For learning fidelity values higher than q_1, we note that a structured grammar space weakens the selection against the minor variants of the dominant grammar in comparison to unstructured or single component grammar models. This effect causes the population to display a dominant abundance below the analytical model's expectations because the close variants of the dominant have a higher equilibrium abundance in mutation-selection balance.

It is our assertion that natural languages belong to a highly structured set even if they are not reducible to a simple sequence representation. As such, the qualitatively different dynamics explored here are important to understanding how human language evolves through time. Additionally, 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.

Clearly, this study is abstract enough to generalize to many situations where agents adapt by learning from one another and convergence is desirable. For further study, it would be natural to explore how structuring languages and gradual learning affect diversification in a structured population

Acknowledgments

References

1. Sutherland, W.J.: Parallel extinction risk and global distribution of languages and species. Nature **423** (2003) 276–279
2. Staff: Endangered languages: Babel runs backwards. Economist **374** (2005)
3. Hashimoto, T., Ikegami, T.: Evaluation of symbolic grammar systems. Springer-Verlag:Berlin (1995)
4. Hashimoto, T., Ikegami, T.: Emergence of net-grammar in communicating agents. BioSystems **38** (1996) 1–14
5. Steels, L.: Self-organizing vocabularies. In Langton, C., Shimohara, T., eds.: Artificial Life V, Nara, Japan (1996) 179–184
6. Nowak, M.A., Komarova, N.L., Niyogi, P.: Evolution of universal grammar. Science **291** (2001) 114–118
7. Kirby, S.: Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity. IEEE Transactions on Evolutionary Computation **5** (2001) 102–110
8. Smith, K., Brighton, H., Kirby, S.: Complex systems in language evolution: the cultural emergence of compositional structure. Advances in Complex Systems **6** (2003) 537–558
9. Niyogi, P.: The Computational Nature of Language Learning and Evolution. MIT Press, Cambridge, Massachusetts (2003) `http://people.cs.uchicago.edu/~niyogi/Book.html`.
10. Chomsky, N.: Aspects of the Theory of Syntax. MIT Press, Cambridge, MA (1965)
11. Chomsky, N.: Rules and Representations. Basil Blackwell, London (1980)

12. Pinker, S., Bloom, P.: Natural language and natural selection. Behavioral and Brain Sciences **13** (1990) 707–784
13. Pinker, S.: The Language Instinct. Penguin, London (1994)
14. Hauser, M.D., Chomsky, N., Fitch, W.T.: The faculty of language: what it is, who has it, and how did it evolve? Science **298** (2002) 1569–1579
15. Fitch, W.T., Hauser, M.D., Chomsky, N.: The evolution of the language faculty: Clarifications and implications. Cognition **Forthcoming** (2005)
16. Nowak, M.A., Komarova, N., Niyogi, P.: Computational and evolutionary aspects of language. Nature **417** (2002) 611–617
17. Joseph, B.: Historical linguistics. In Aronoff, M., Rees-Miller, J., eds.: The Handbook of Linguistics. Blackwell, Oxford (2000)
18. Lamb, S.M., Mitchell, E.D.: Sprung from a Common Source: Investigations into the Prehistory of Languages. Stanford University Press, Stanford, California (1991)
19. Eigen, M., Schuster, P.: The hypercycle: A principle of natural self-organization. Springer Verlag: Berlin (1979)
20. Eigen, M., McCaskill, J., Schuster, P.: The molecular quasi-species. Adv. Chem. Phys. **75** (1989) 149–263
21. Komarova, N.L., Niyogi, P., Nowak, M.A.: Evolutionary dynamics of grammar acquisition. Journal of Theoretical Biology **209** (2001) 43–59
22. Livingstone, D., Fyfe, C.: Modelling the evolution of linguistic diversity. In D. Floreano, J.N., Mondada, F., eds.: ECAL99, Berlin, Springer-Verlag (1999) 704–708
23. Crow, J.F., Kimura, M.: An Introduction to Population Genetics Theory. Harper & Row Publishers, New York, Evanston and london (1970)