

Understanding Linguistic Evolution by Visualizing the Emergence of Topographic Mappings

Henry Brighton*

Simon Kirby

Language Evolution and Computation
Research Unit

School of Philosophy, Psychology,
and Language Sciences

The University of Edinburgh

Edinburgh, UK

hbrighton@mpib-berlin.mpg.de

simon@ling.ed.ac.uk

Abstract We show how cultural selection for learnability during the process of linguistic evolution can be visualized using a simple iterated learning model. Computational models of linguistic evolution typically focus on the nature of, and conditions for, stable states. We take a novel approach and focus on understanding the process of linguistic evolution itself. What kind of evolutionary system is this process?

Using visualization techniques, we explore the nature of replicators in linguistic evolution, and argue that replicators correspond to local regions of regularity in the mapping between meaning and signals.

Based on this argument, we draw parallels between phenomena observed in the model and linguistic phenomena observed across languages. We then go on to identify issues of replication and selection as key points of divergence in the parallels between the processes of linguistic evolution and biological evolution.

Keywords

Language, evolution, visualization,
replicators, learning

1 Introduction

Linguistic evolution is the process by which languages themselves evolve as they are transmitted from one generation to another [5]. But to what degree does the process of linguistic evolution mirror that of biological evolution? At first glance the two processes share similarities. Both result in the transmission of information from one generation to another, both result in adaptation, and both appear to be driven by the differential retention of some unit of replication. In this article we explore these parallels using visualization techniques: we aim to understand the degree to which language itself can be considered an evolutionary system. Previous work in this area has focused on the construction of computational models (e.g., [16, 2]). These models aim to explain the conditions under which linguistic structure emerges. Here, stable states¹ represent adaptations to the problem of cultural transmission, and it is these states that are taken as the focus of explanation. In this article we take a slightly different approach, and focus on the *process* of cultural adaptation. What kind of evolutionary system do models of linguistic evolution represent?

In order to address this question, we present a simple model of linguistic evolution designed solely to enable us to visualize the process of cultural adaptation. This approach provides a unique way of understanding what we term the *middle scale* properties of the evolutionary system. Typically, models of linguistic evolution are analyzed with respect to either (1) global measures that conflate

* Corresponding author. Current address: Center for Adaptive Behavior and Cognition, Max Planck Institute for Human Development, Lentzealle 94, D-14195 Berlin, Germany.

¹ When we refer to stability in this discussion, we are referring to “start near, stay near” (Liapounov) stability, where subsequent states need not be identical, but rather remain within some neighborhood of the state space [12].

several adaptive processes, or (2) detailed analyses of particular states of the model. Using visualization techniques, we show how models of linguistic evolution can be understood in terms of competition between replicators. One key problem in understanding linguistic evolution, as well as biological evolution, is formalizing precisely what kind of entities we regard as replicators. The approach we take here is geared toward this endeavor: visualization allows us to begin to shed some light on the middle scale between the very specific and the very general. Our conclusion is that, after examining computational models and their relation to the process of language change and evolution, the notion of selection during linguistic evolution is best regarded as an analogue process of replication in which replicators are localized mappings between meanings and signals. We argue that the degree to which the notion of linguistic evolution parallels that of biological evolution remains very much an open problem.

2 Background

The complex structural properties of human language make it quite unlike any other communication system found in nature. Although languages differ substantially, they nevertheless exhibit universal tendencies. One of the achievements of modern linguistics is to show how all languages, past and present, can be described as particular configurations of linguistic parameters. In this sense, all languages are essentially the same, and exhibit only superficial differences. We are interested in why language has these very specific structural properties: we seek an explanation for how this system, exhibiting unsurpassed complexity in relating sound and meaning, came to exist.

One approach to identifying why languages have their characteristic structure is to ask why the innate human faculty for language has the structure that it does. Such an enterprise could proceed with an attempt to identify the cognitive structures responsible for processing language. But would such an analysis tell us the whole story? In this section we argue that it would not, and we discuss a substantially revised conceptual framework. We will argue that universal tendencies in language are to some degree adaptive solutions to the problem of cultural transmission. This view is central to the remainder of the article: we will improve our understanding of the notion of cultural adaptation using visualization techniques.

2.1 Language as an Expression of the Genes

Among those interested in language, a widespread hypothesis is that language, like the visual system, is an expression of the genes (e.g., [7]). Differences between languages, according to this position, are tightly constrained by innately specified dimensions of variation. So, returning to the original question, if we are interested in the structural hallmarks of language, then we should seek a wholly psychological (i.e., cognitive, mentalistic, or internalist) explanation. In short, by understanding those parts of the human cognitive system relevant to language we can understand why languages have certain structural characteristics and not others. This position is largely substantiated by the *argument from the poverty of the stimulus*, where justification for the notion of linguistic nativism is based on the observation that the yield of the process of language acquisition (knowledge of a specific language) is significantly underdetermined by the input available to the learner [6]. Language, therefore, should be considered part of our biological endowment, just like the visual system. The intuition is that one would not want to claim that we learn to see, and the same way, we should not claim that we learn to speak.

2.2 Cultural Selection for Learnability

Linguistic nativism is far from accepted in the extreme form presented above (e.g., [9, 13]). An alternative to this hypothesis is that the structure of language, to some extent, is learned by children: humans can arrive at complex knowledge of language without the need to have hardwired (innately specified) expectations of *all* dimensions of linguistic variation. This is the view that we will adopt in this article. We assume that to some degree language is learned through inductive generalizations from linguistic data, but to what degree it is learned is unclear. Previously, we have argued that the

degree to which language is learned through a process of inductive generalization has a profound effect on the explanatory framework used to understand why language has the structure that it does [4]. To do this we take an evolutionary perspective, and seek to explain *how*, from a non-linguistic environment, linguistic structure can emerge through cultural evolution. In short, this view casts doubt on the view that the hallmarks of language are, as Chomsky states, “coming from inside, not from outside” [7].

In this article we examine the process by which structure arises through cultural selection for learnability. What exactly does this mean? Consider how languages persist from one generation to another. Firstly, languages are transmitted via human minds. Minds act as the conduit for the transmission of particular languages. Secondly, linguistic transmission is mediated through learning: Humans learn language from conspecifics. Transmission of a particular language from one mind to another occurs through cultural transmission. The *capacity* for processing language persists as a result of genetic transmission. It is the process of cultural transmission that we have argued has a profound effect on the structure of languages we see (see also [8, 11, 14]). To summarize, we claim that universal tendencies in language cannot be explained entirely in terms of a psychological theory. We also require a theory of how certain linguistic forms are adaptive with respect to constraints on cultural transmission.

2.3 Natural Language from Artificial Life

We have laid down the theoretical basis for studying language as a system exhibiting structural properties that exist as adaptations to cultural transmission. To test this hypothesis we build agent-based computational models. We term these models *iterated learning models* (see [17] for a review). An iterated learning model comprises a series of agents arranged into generations. These agents are initially identical. Each agent, in turn, learns an abstract language on the basis of the linguistic behavior of agents in the previous generation. This modeling framework captures the fact that humans learn language on the basis of input from an existing speech community. As one might expect, progress in understanding linguistic structure as a set of cultural adaptations is often sought by constructing more elaborate models. We might seek accurate models of language acquisition, plausible models of semantic structure, or more informed models of population dynamics. The more we experiment with concepts such as these, the more elaborate are the theories we can propose. In this article, rather than elaborate on existing models, we will strip the iterated learning process bare, and shed some light on an understanding of cultural selection for learnability. The model we develop is designed solely to facilitate visualization: We will make several simplifying assumptions in order to frame the problem in such a way that visualization is possible. Where appropriate, we will highlight these assumptions.

3 Iterated Instance-Based Learning

In this section we will develop a basic iterated learning model. Although the model will address the issue of the emergence of linguistic structure, the design of the model will be geared toward visualization. At each stage in the construction of the model we will seek the most basic set of computational modeling decisions. First, we will describe a model of language based on a mapping between two Euclidean spaces. Second, we will develop a model of language learning based on the instance-based learning paradigm. With these components in place, along with a basic model of cultural transmission, we will then be in a position to visualize the process of iterated learning.

3.1 Language as a Mapping

Language is a relationship between sound and meaning. Language, therefore, can be regarded as mapping between two spaces. The precise structure of these spaces is little understood, but what we will term the *meaning space* corresponds to the set of structures relating to an internal representation interfacing with the conceptual/intentional aspects of the cognitive system (also known as LF, or

logical form, in Chomskyan terms). The *signal space* corresponds to external signals. With respect to the traditional view of the language faculty, signals are derived from another internal linguistic representation interfacing with the articulatory-perceptual systems (known as PF, or phonetic form). Figure 1a depicts the relationship between these linguistic representations and the surrounding (non-linguistic) cognitive system. We will discuss language as a mapping between the meaning and the signal space, but abstract away from the details of the structure of these spaces. In fact, we will treat both the meaning space and the signal space as bounded Euclidean spaces in \mathbb{R}^2 . This abstraction may appear to wash away any notion of linguistic plausibility, but a fundamental linguistic property of interest, *compositionality*, remains intact. Compositionality is a property of the relationship between sound and meaning. A compositional language is one where the meaning of a signal is a function of the meaning of its parts. Other compositional communication systems can be found in nature, but language is the only one that is learned. For example, ants [19] and honeybees [22] have communication systems that are both compositional and innate.

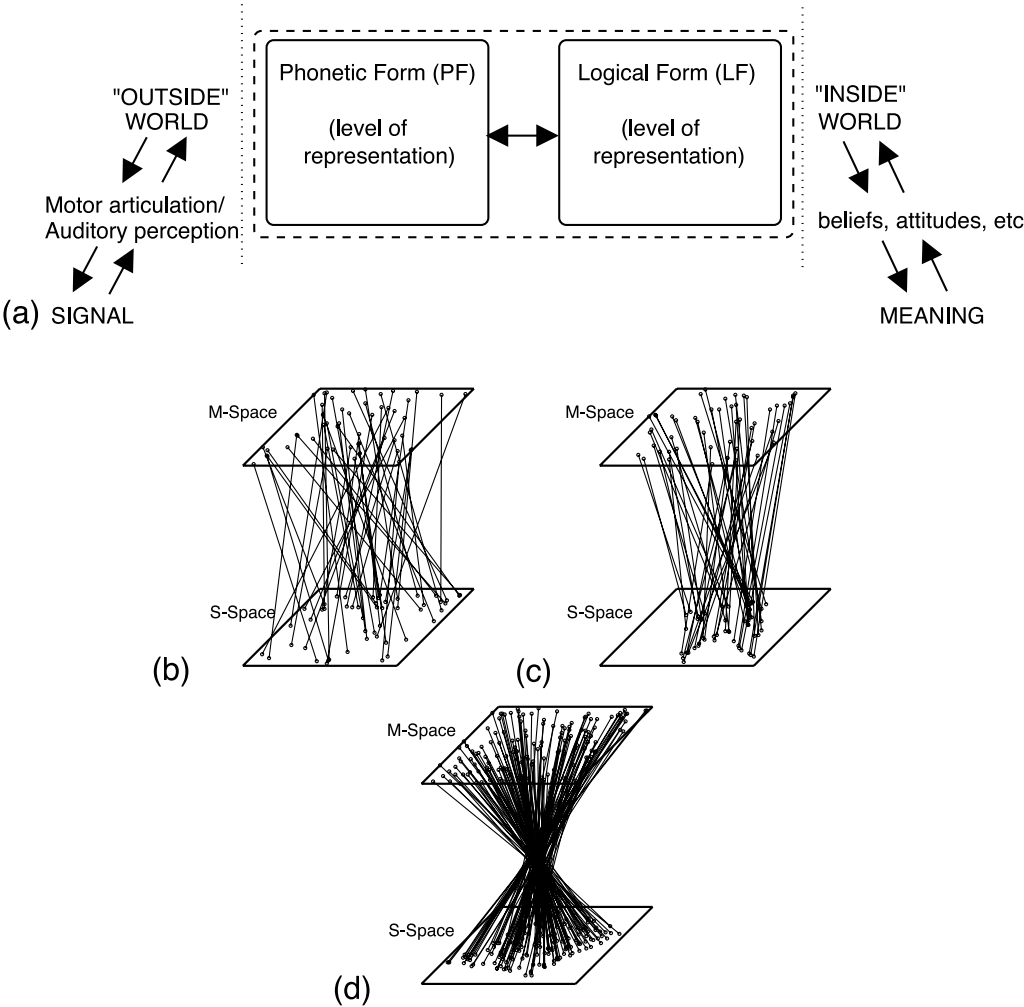


Figure 1. Modeling language. In (a), the relationship between meanings, signals, and the traditional view of the language faculty is detailed. The mapping shown in (b) is random; the neighbors surrounding a point in the meaning space suggest nothing about the location of the corresponding signal. In (c) and (d), similar meanings map to similar signals; these two languages represent a topographic relation between meanings and signals.

As an example of compositionality in language, consider the compositional utterance “large olive pike”. Parts of the signal correspond to parts of the meaning. Similarly, the utterance “large olive tench” shares parts of the signal, and therefore, parts of the meaning. If language were non-compositional, arbitrary signals would be associated with each meaning. These examples illustrate the fact that much of language is neighborhood related: nearby meanings tend to map to nearby signals. Although this is a rather crude characterization of language as a whole, the property of compositionality can be considered an absolute language universal: *all* languages exhibit compositionality.

Given a model of language in which meanings and signals are points in \mathfrak{R}^2 , Figure 1b–d illustrates three degrees of compositionality. In Figure 1b, which represents a non-compositional language, meanings are randomly associated with signals: similar meanings are unlikely to be associated with similar signals. In contrast, Figure 1c represents a language for which similar meanings do map to similar signals. Figure 1d represents a fully distance-correlated mapping, and like Figure 1c, we can consider it a topographic mapping [21]. Given the language model presented here, compositionality is a matter of degree.

To summarize, in the model of language used in the following experiments, meanings and signals are drawn from two-dimensional real-valued number spaces, denoted by \mathcal{M} and \mathcal{S} , respectively. More precisely, some meaning $m \in \mathcal{M}$ is the pair (x, y) where $x, y \in \mathfrak{R}$ such that both $0 < x \leq 100$ and $0 < y \leq 100$. Similarly, signals are points in the space \mathcal{S} where some $s \in \mathcal{S}$ is the pair (x, y) where $x, y \in \mathfrak{R}$ such that both $0 < x \leq 100$ and $0 < y \leq 100$.

3.2 An Instance-Based Associative Memory

Within the iterated learning model each agent will observe the linguistic performance of another agent. On the basis of this performance, which is represented by a series of meaning-signal pairs, the agent is expected to derive a competence, that is, the ability to produce signals for all meanings. The agents must therefore generalize beyond the data they experience: They are required to yield signals for meanings they have never observed. Each agent can be thought of as a learning algorithm. The learning algorithm is required to take a finite set of meaning-signal pairs $O = \{p_1, p_2, \dots, p_r\}$ where each pair $p_i = \langle m_i, s_i \rangle$ is such that $m_i \in \mathcal{M}$ and $s_i \in \mathcal{S}$. The algorithm must then induce a function $F : m \in \mathcal{M} \mapsto s \in \mathcal{S}$ defined for all $m \in \mathcal{M}$. In other words, given a finite subset of some infinitely large language, the learning algorithm must induce a hypothesis that relates an infinite number of meanings to their appropriate signals.

These are the functional requirements of an agent. First, the agent learns by observing the behavior of the agent in the previous generation. Second, once the learner has learned through observation, it is expected to produce signals for any given meaning. To perform these two tasks, we use an instance-based learning algorithm (see [1]) augmented so that, rather than producing classification decisions, it produces signals. The operational details are as follows:

- *Learning.* The learner observes r instances of the mapping (language) of the agent in the previous generation. These instances are stored as is, without any further processing.
- *Production.* When called to express some arbitrary meaning m , the learner finds the three nearest neighbors among the r stored instances. These nearest neighbors are examples of production, and as such can be used to inform the production of m . Using vector geometry, an appropriate signal s for m can be calculated as a linear combination of the vectors relating m 's three nearest neighbors. Note that if m has already been observed, which is extremely unlikely, then the agent will just produce the signal observed in conjunction with m .

So, rather than as a classifier, the instance-based learning algorithm is acting as an associative memory. We describe this process in more detail in the Appendix, but for the purposes of this discussion, the key point is that the learning algorithm is in the spirit of the instance-based learning

approach: we assume that the nearest neighbors of m should inform the production of m . The associative memory therefore contains a bias: The assumption is that signals for novel meanings will conform to any local transformation between nearby meanings and their corresponding signals. To summarize, each agent memorizes a finite set of observed meaning-signal associations. Then, armed with a nearest-neighbor decision rule, it can produce signals for all meanings on the basis of local transformations between the stored meaning-signal associations.

3.3 Putting It All Together

A model of language and a model of learning are now in place. In this section the remaining details required for a fully specified iterated learning model will be fleshed out. To retain the simplicity of the model, we will consider an iterated learning model where each generation contains a single agent. A simulation of n iterations of the model will therefore comprise a series of agents A_1, A_2, \dots, A_n . Each agent, in turn, first observes linguistic behavior and then produces linguistic behavior. Figure 2 illustrates this process. The agents act as a conduit for the evolving language. The agents themselves do not evolve in any way; it is only the language that changes. Recall that each agent induces its linguistic competence on the basis of r observed meaning-signal pairs. The agent will then be prompted to produce signals for r random meanings. These production decisions, which also represent a series of r meaning-signal pairs, will form the input to the agent in the next generation. The first agent is always given r meaning-signal pairs drawn from a random language in which each meaning maps to a random signal.

A crucial component of the model, which so far has only been implicitly introduced, is the notion of the *transmission bottleneck*. The transmission bottleneck refers to the fact that agents never observe the entire language of the previous generation. This restriction is paralleled in the learning of natural language by human infants, and is closely related to the notion of the poverty of stimulus, discussed above. Because each agent has infinite generative capacity, and the transmission of this capacity is transmitted on the basis of a finite set, the transmission bottleneck is unavoidable. Because an infinite set is being transmitted over a channel with a finite capacity, the precise number of observations each agent learns from is of little interest to this discussion. The parameter r determines the rate at which a structure appears in the mapping. The qualitative outcome of the model will therefore be invariant over sensible values of r : The initial random language will change to reflect the bias present in the learners: the remnants of the first-generation language will be washed out as the evolving language passes through each agent. We will show that, as the simulation progresses, the language will begin to exhibit compositional structure. Each generation is therefore passed linguistic data containing the structural residue laid down by the production behavior of previous agents.

The experimental details are now in place. We have a model of language in which both meanings and signals are points in a two-dimensional real-number space. A model of learning takes examples of language performance and derives language competence. This model of learning contains a bias toward neighborhood preservation. We also have a model of cultural transmission: a single-agent-per-generation iterated learning model.

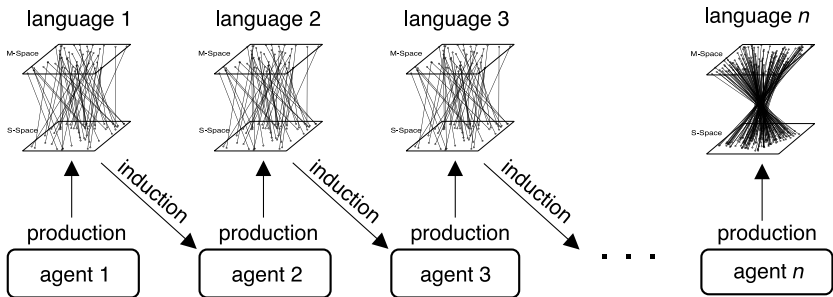


Figure 2. The iterated learning model. The language evolves as it passes through each agent.

4 Visualizing Iterated Learning

Typically, an analysis of an iterated learning model will proceed by tracing some measure of interest, such as the degree of compositionality in the language [3], the stability of the system [2], or the generative capacity or size of the evolved grammars [16]. An analysis of this sort conflates several properties of the system into a single measure and provides a useful indication of the global state of the system over time. We can also take a complementary approach where we analyze individual languages or agents. By focusing on individual states we gain an understanding of the details of the system. However, performing this kind of analysis over the course of a simulation is intractable for simulations lasting hundreds of thousands of generations. In this section, we will use visualization techniques to analyze the iterated learning model at a middle scale lying between an analysis of global measures and an analysis of individual states.

Before conducting an analysis, it is worth considering how we go about using iterated learning models to inform an explanation. An individual simulation run can be thought of as a trajectory through the language space. If this trajectory corresponds to a random walk, then the model can tell us very little. Of more interest is a situation in which multiple runs of the model lead to some subset of the language space being consistently visited. In this situation we can conclude that this region of the language space is occupied by languages exhibiting adaptive properties. Ultimately, by running the model multiple times with differing initial conditions, the hope is that the model can shed light on the occurrence of these properties in nature. For example, we may seek to explain compositional and recursive structure [16], or languages with irregular and regular forms [15]. Here, the stable states of the model are of interest: they are interpreted as a model of the phenomena being explained.

Another source of explanatory force relates to the *process* of iterated learning. For example, in what sense is iterated learning an evolutionary system? To achieve any degree of explanatory force, this question must be resolved. As well as identifying the occurrence of particular states, we also need to explain how and why these states are reached. Now, the design of the experiment detailed above was driven by the need to trade off the ability to draw insightful conclusions about the stable states of the model in favor of being able to gain a firmer grasp on the process of iterated learning. It is a feature of desirable explanations that both of these issues are fully understood: we seek to understand how certain states are emergent properties of the dynamics, and we seek a satisfactory conceptual framework for understanding the dynamics. This is the desired output of the research program. So, in this section, we attempt a foray into understanding the process of iterated learning through visualization. Visualization is important because we need to gain a grasp on the middle scale between global behavior and the behavior of individual agents.

4.1 Coarse-Grained Visualization: Interpolated Grid Projection

Visualizing the behavior of the iterated learning developed above will require the visualization of a mapping over time. Specifically, we are interested in the formation of topographic mappings over time. The visualization of topographic mappings is a feature of research into the formation of self-organizing maps in associative memories such as the Kohonen network [21, 18]. As a first cut, we will adopt a similar procedure. First, we take a set of points in the meaning space and, using the learning algorithm represented by the agent, map these points to the signal space. The result will be some distortion of the relationship between the points in the meaning space. The points in the meaning space we project are the 81 points defining a 9 by 9 grid of equally spaced locations in the meaning space. This grid represents a sample of the meaning space. Figure 3a illustrates this process using a 3 by 3 grid. The projected points are joined by lines to indicate how the signal space topology is related to meaning space topology. In the case depicted in Figure 3a, the mapping between meanings and signals is neighborhood preserving: neighbors in the meaning space map to neighbors in the signal space.

Figure 3b depicts the application of this visualization method over the course of 140 iterations (generations) of the iterated learning model discussed in the previous section. The structure of the

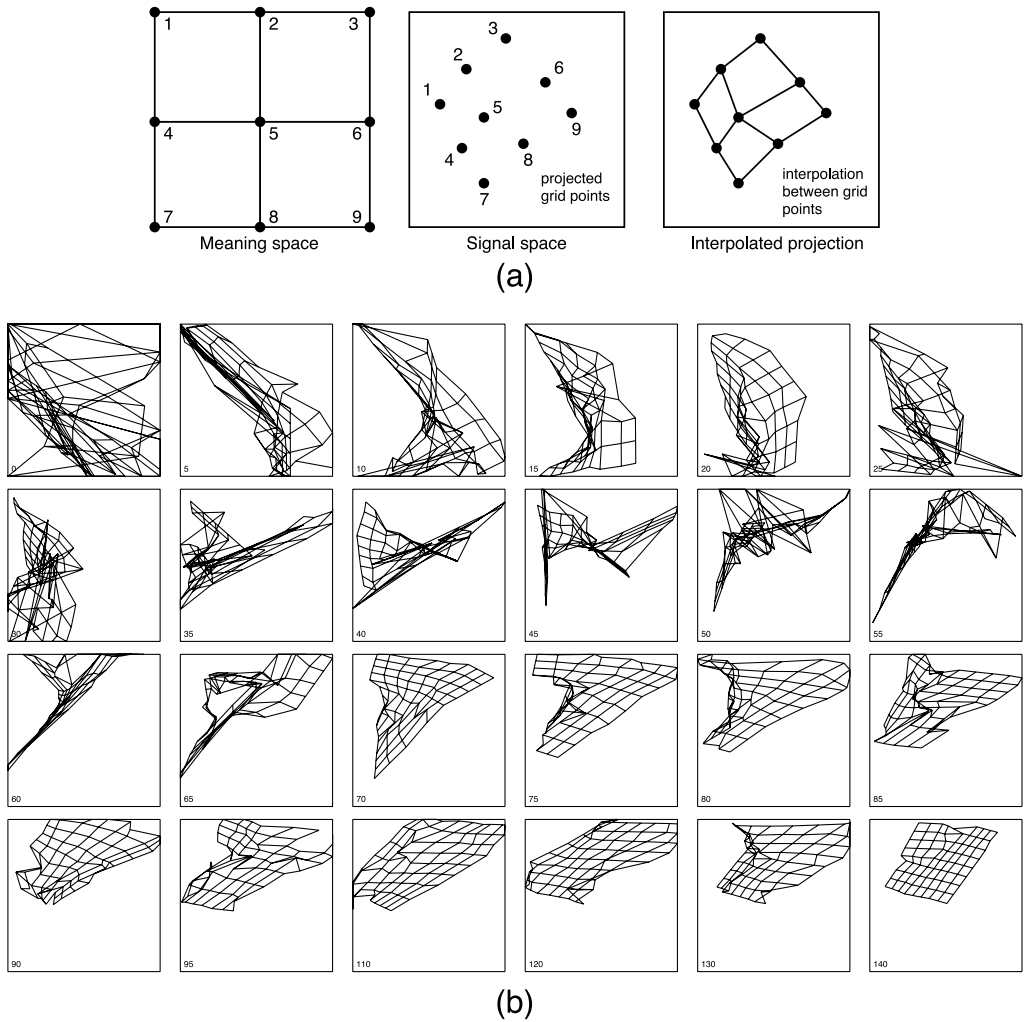


Figure 3. Coarse-grained visualization. Part (a) illustrates how the visualization procedure is used in the projection of a 3 by 3 grid. In (b), this method is used to visualize 140 iterations of the model using a 9 by 9 grid.

grid becomes increasingly apparent as the simulation progresses, through a process of the untangling of competing *submappings*. What we will refer to as a submapping, or a *fragment* of the mapping, is a mapping from some neighborhood of the meaning space to some set of signals in the signal space. These signals, of course, will only occupy a neighborhood themselves if the submapping is topographic. Figure 3b illustrates how fragments of the mapping establish themselves—a local topographic relation—but may disappear due to the pressure imposed by some more dominant and globally coherent topographic relation. After the 90th iteration, the mapping is untangled in the sense that all competing submappings have roughly the same orientation. This structure is stable, and persists for some time afterwards with the mapping suffering from local distortions. The mapping is never entirely stable. Although these results represent the behavior of a single simulation run, the behavior is typical. The main feature of interest is the stability of the topographic relation between meanings and signals. This relationship persists, and in this sense is propagating itself. We can conclude that topographic mappings, in this context, are adaptations to the problem of iterated learning.

Before discussing this process in more detail, it is worthwhile noting some deficiencies in the visualization method. First, there is no way of telling, early on in the emergence of the mapping,

the relative orientation of the submappings. Second, the process of interpolation hides interesting discontinuities in the mapping. Note that, in contrast, the application of this visualization method to understanding the development of self-organizing maps does not entail any interpolation: each activation unit is represented in the grid. Unfortunately, when visualizing the language model developed here, a significant amount of information is lost due to the process of interpolation.

4.2 Fine-Grained Visualization: Image Projection

We now extend the basic visualization method in two ways. First, in order to visualize the mapping with minimal interpolation, we increase the rate at which we sample the meaning space. To do this, instead of projecting the intersecting points of the grid, we project all points drawn from some test image. These points are not interpolated in any way once they are projected into the signal space. Because the meaning space is infinitely large, the projected test image will still represent an approximation to the “true” mapping. The second extension to the basic visualization method is to assign colors to the points being projected. By doing this, we can use a color test image with some discernible structure such that the orientation of the fragments of the mapping can be recovered visually.

This procedure, along with the test image, is shown in Figure 4. The test image is a color map over which we have superimposed (a) a grid and (b) arrows pointing to each of the four corners of the test image. Figure 5 depicts the state of a single simulation run over 500 generations. By far the most insightful application of this visualization method, however, is gained by viewing successive states in the form of an animation. We have a number of MPEG movie files illustrating the evolution of structure in the mappings using this visualization method.²

We can identify the following phases in the evolution of topographic maps:

- **Phase 1.** In the initial state, every point in the meaning space is mapped randomly to a point in the signal space.
- **Phase 2.** Structure begins to emerge in the mapping between meanings and signals. Competition begins between many dramatically different transformations.
- **Phase 3.** We see increasing dominance of a collection of mutually supporting contiguous submappings. Rogue fragments of the mapping continue to compete, but are highly unlikely to establish themselves or grow, due to the continued growth of the dominant mapping.
- **Phase 4.** One observes the eventual dominance of a continuous but fluid collection of mappings sharing transformational identity.
- **Phase 5.** A global topographic tendency remains, with the projected image continually floating and being repelled by the edges of the space.

Visualizing the system in this way illustrates how iterated learning can result in the cumulative evolution of structure. The language, as it passes down each generation, progressively acquires a general tendency toward structure preservation in the mapping. But in what way, if any, is this an evolutionary system? What kind of conceptual framework should we invoke to describe the development of the stable, structured states? These experiments raise some problematic questions, which we will now discuss.

4.3 Analysis

Part of our motivation for employing visualization techniques is the need to further our understanding of iterated learning as an evolutionary system. More specifically, in what way does information transmission through learning constitute an adaptive system, and in what way is this system

² These are available on the web; see <http://www.ling.ed.ac.uk/~henryb>.

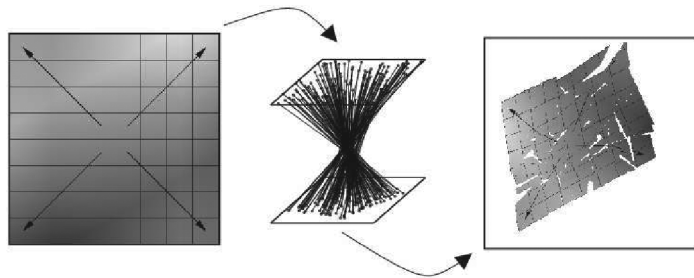


Figure 4. The fine-grained visualization method.

related to biological evolution? Before considering these questions it is worthwhile relating the behavior of the model to the linguistic phenomenon we are interested in.

4.3.1 Relating Model Behavior to Language

The model we have developed can only superficially be related to specific examples of linguistic phenomena. However, the utility of the approach we take concerns a more general analysis of information transmission through learning. Given this level of abstraction, the relevant parallel between language and the model is that both can be understood as examples of a culturally transmitted system relating two structured spaces. In light of this similarity, we believe fairly strong parallels can be drawn. For example, the initial phase in linguistic evolution, where we see a transition from a system of relating meanings to signals arbitrarily to a system of relating meanings to signals compositionally, mirrors the transition from protolanguage to full human language discussed by Wray [23]. However, the principal parallel we draw relates to individual states in the model that, when visualized, depict competing and inconsistent structured regions of the mapping between meanings and signals. We also see this in language. For example, Latin has five noun declensions: every noun belongs to one of these declensions, and each declension has an associated system of case endings.

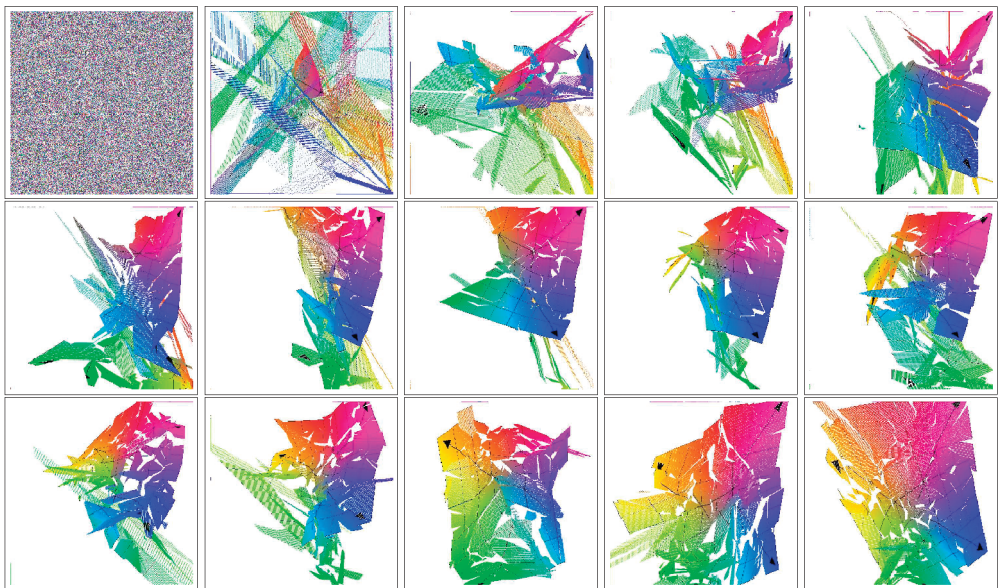


Figure 5. Fine-grained visualization of the evolving language over 500 generations.

These systems of case endings are regular but, importantly, differ across declensions. Thus we see *competing systems of regularity* being transmitted in both language and the model. In short, both language and the model exhibit idiosyncrasies despite widespread regularity (e.g., [10]). It is at this level of abstraction that the model can provide insight.

An analysis of the system over many generations reveals how imperfections in the process of transmission can profoundly alter the evolutionary trajectory, the result being an occasional wholesale reorientation of the mapping. In language, similar processes are at work. For example, the rich case structure of Latin, through phonological change, became ambiguous during the history of French. As a result of this ambiguity, development from Old to New French saw the abandonment of the case system in favor of word order [20]. This example, where small changes lead to subsequent restructuring, is a characteristic feature of the process of language change. We can draw a strong parallel between that process and the behavior of the simple iterated learning model we present here. Our model tells us that a patchwork of local transformations can retain a general tendency despite imperfect transmission. Inaccuracies are inevitable during the transmission of this mapping, and we see that deformations introduced can lead to either first, a fleeting discontinuity or, second, a more profound restructuring of the mapping. As this discussion suggests, the model allows us to visualize (1) the emergence of regularity from an initial state of irregularity, and (2) subsequent transmission and variation of this regularity. We will now use these points to inform a discussion focusing on the evolutionary nature of the system.

4.3.2 Iterated Learning as an Evolutionary System

When considering language as an evolutionary system, to begin with we must consider the notion of selection. If we define the notion of selection as the differential retention of some unit of replication, then the first problem we arrive at is that it is far from clear what is being selected in the model presented here. In other words, what is the unit of replication? Two observations are important. Firstly, it is clear that features of the mapping persist for lengthy periods of the simulation. A feature, in this context, could be defined as one or more contiguous regions of the mapping between meanings and signals. These regions must represent approximately the same relationship between meanings and signals. For example, they might represent approximately the same scaling and a rotation of the meaning space. Secondly, neither the nature of this relationship between meanings and signals in one of these regions, nor the size of the region, is ever entirely identical from one generation to the next. To begin to explain this behavior in terms of replication, we must then question the essentially binary notion that these features of the mapping are either passing unaltered between two agents, or being selectively blocked. This brings us to one fundamental difference between linguistic and biological evolution. In the case of biological evolution we can speak of relatively direct copying of DNA. In the case of, for example, asexual selection, this DNA is either transmitted or it is not. In the case of linguistic evolution we do not see this direct copying. The mapping is first translated into externalized utterances, and then, for this information to then be transmitted successfully, these utterances need to be *reverse translated* back into a grammar. Put another way, the notion that a property of the mapping can be copied or transmitted directly does not apply in the case of linguistic evolution. Instead, during linguistic evolution copying occurs as a result of the *combination* of learning and production, and any notion of selection will turn on the nature of these processes. In light of this argument, we note that replication in this model is matter of degree: It is more informative to regard replication as an analogue process exhibiting varying degrees of copying fidelity. Copying fidelity will necessarily be subject to the vagaries of the processes of learning and production, rather than a digital process within which features of the mapping are either copied unaltered or not copied at all.

As we have suggested, replicators in the model are structured compositional regions of the mapping between meanings and signals. These regions represent some local regular transformation in much the same way as a noun declension represents a local system of regularity in Latin. Because these regions are compositional, they can be generalized from: A signal for a novel meaning in this region can be calculated that conforms to the local transformation. Regularity in a region of the

mapping will therefore increase the probability that the region will be transmitted accurately. We can draw a parallel here with language, as the regularity of the case endings within noun declension applies to many nouns, and therefore this regularity is more likely to be transmitted than some idiosyncratic system related to few nouns.

Through visualization, the analysis of the model and its parallels with language suggest that we can rightfully speak of the differential retention of units of replication. These units are local systems of regularity—a hallmark of language. These systems of regularity compete with each other and undergo deformation as they are transmitted across generations. This analysis has also raised what we believe to be a key point of divergence between linguistic and biological evolution. In contrast to the transmission of DNA, linguistic structure is not copied directly but undergoes translation and reverse translation. It is far from clear how the analogue process of selection we have argued for in linguistic evolution is related to the digital process of selection we see in biological evolution. We view this issue as one requiring further investigation.

5 Summary

First we outlined the theoretical basis for treating language as a complex adaptive system in which certain structural properties are adaptations to the problem of cultural transmission. We argued that rather than a purely psychological theory, an explanation for the universal tendencies in language requires a theory explaining why these structural properties are adaptive. Typically, models of linguistic evolution focus on explaining how specific properties of language emerge, and the range of conditions for emergence. We have taken a slightly different approach, one where we try to characterize the process of linguistic evolution in evolutionary terms. To investigate this issue we then developed a simple iterated learning model. The construction of this model was guided by the need to visualize the dynamics of iterated learning. In order to visualize the evolving system, language was modeled as a mapping between a two-dimensional meaning space and a two-dimensional signal space. Coordinates in these space are real numbers. Linguistic agents were modeled as augmented instance-based learners capable of mapping meanings to signals. Designing the iterated learning model in this way allowed us to visualize the evolution of linguistic structure.

Using two visualization methods, we demonstrated how linguistic evolution can be understood in terms of competition between local transformations. Stable states were characterized by a fluid collection of locally consistent transformations. When invoking an evolutionary explanation for development of these states, the notion of selection is, at first sight, problematic. Subsequent states in the model only superficially suggest replication. In order to account for this process, we argued that an analogue process of replication occurs. This analogue process differs from the essentially binary nature of selection of DNA in that linguistic structure is both translated to, and reverse-translated from, externalized utterances. Any change introduced as a result of these processes will be transmitted and inherited by the next generation of language users. Drawing a parallel with language, we discussed how a similar process underlies the phenomenon of language change where the retention of structural tendencies occurs while at the same time language is never fully stable. To our knowledge, this is the first time visualization techniques have been used to understand the process of linguistic evolution. Several important issues remain in understanding this process, and we regard the experiments reported here as a first step in understanding linguistic evolution in the context of evolutionary theory.

Acknowledgment

Henry Brighton was supported by EPSRC studentship award 99303013 and ESRC Postdoctoral fellowship award T026271445.

References

1. Aha, D. W., Kibler, D., & Albert, M. K. (1991). Instance-based learning algorithms. *Machine Learning*, 6(1), 37–66.

2. Brighton, H. (2002). Compositional syntax from cultural transmission. *Artificial Life*, 8(1), 25–54.
3. Brighton, H., Kirby, S., & Smith, K. (2003). Situated cognition and the role of multi-agent models in explaining language structure. In D. Kudenko, E. Alonso, & D. Kazakov (Eds.), *Adaptive agents and multi-agent systems*. Berlin: Springer.
4. Brighton, H., Kirby, S., & Smith, K. (2005). Cultural selection for learnability: Three hypotheses concerning the characteristic structure of language. In M. Tallerman (Ed.), *Language Origins: Perspectives on Evolution*. Oxford, UK: Oxford University Press.
5. Briscoe, E. (Ed.), *Linguistic evolution through language acquisition: Formal and computational models*. Cambridge, UK: Cambridge University Press.
6. Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
7. Chomsky, N. (2002). *On nature and language*. Cambridge, UK: Cambridge University Press.
8. Christiansen, M. (1994). *Infinite languages, finite minds: Connectionism, learning and linguistic structure*. Ph.D. thesis, University of Edinburgh.
9. Cowie, F. (1999). *What's within? Nativism reconsidered*. Oxford, UK: Oxford University Press.
10. Culicover, P. W. (1999). *Syntactic nuts: Hard cases, syntactic theory, and language acquisition*. Oxford, UK: Oxford University Press.
11. Deacon, T. W. (1997). *The symbolic species*. New York: W. W. Norton.
12. Glendinning, P. (1994). *Stability, instability, and chaos: An introduction to the theory of nonlinear differential equations*. Cambridge, UK: Cambridge University Press.
13. Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford, UK: Oxford University Press.
14. Kirby, S. (1999). *Function, selection, and innateness: The emergence of language universals*. Oxford, UK: Oxford University Press.
15. Kirby, S. (2001). Spontaneous evolution of linguistic structure: An iterated learning model of the emergence of regularity and irregularity. *IEEE Journal of Evolutionary Computation*, 5(2), 102–110.
16. Kirby, S. (2002). Learning, bottlenecks and the evolution of recursive syntax. In E. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational models* (pp. 173–203). Cambridge, UK: Cambridge University Press.
17. Kirby, S. (2002). Natural language from artificial life. *Artificial Life*, 8, 185–215.
18. Kohonen, T. (2001). *Self-organizing maps* (3rd ed.). Berlin: Springer.
19. Reznikova, Z., & Ryabko, B. (1986). Analysis of the language of ants by information-theoretical methods. *Problems of Information Transmission*, 22(3), 245–249.
20. Trask, R. L. (1996). *Historical linguistics*. London: Arnold.
21. Van Hulle, M. M. (2000). *Faithful representations and topographic maps: From distortion- to information-based self-organization*. New York: Wiley-Interscience.
22. von Frisch, K. (1974). Decoding the language of the bee. *Science*, 185, 663–668.
23. Wray, A. (1998). Protolanguage as a holistic system for social interaction. *Language and Communication*, 18, 47–67.

Appendix: Signal Production

The most obvious production scheme is the following. Given the meaning m , the three nearest neighbors of m are found, along with their corresponding signals. Now, using these three observed examples of the mapping, given by $\langle m_1, s_1 \rangle$, $\langle m_2, s_2 \rangle$, and $\langle m_3, s_3 \rangle$, we can say that the relationship between m and its neighbors m_1 , m_2 , and m_3 in meaning space should somehow be reflected in the relationship between s , the signal to be produced for m , and the signals s_1 , s_2 , and s_3 . That is, any local relationship between meanings and signals should guide our choice of s . The problem of production,

characterized in this way, is in line with the motivation behind instance-based learning. To exploit the relationship suggested by this approach, we can first represent m as a linear combination of two vectors, denoted as p and q , derived from m_1 , m_2 , and m_3 :

$$p = m_2 - m_1 \quad (1)$$

$$q = m_3 - m_1 \quad (2)$$

Representing m as a linear combination of p and q , we use two scalar constants A and B :

$$m = Ap + Bq \quad (3)$$

Now, A and B represent the multiplicative factors of the vectors p and q that give m when added together. Rewriting Equation 3, we have

$$\begin{pmatrix} m_x \\ m_y \end{pmatrix} = A \cdot \begin{pmatrix} p_x \\ p_y \end{pmatrix} + B \cdot \begin{pmatrix} q_x \\ q_y \end{pmatrix}. \quad (4)$$

Rearranging, we get

$$A = \frac{q_y m_x - q_x m_y}{p_x q_y - p_y q_x} \quad (5)$$

$$B = \frac{p_y m_x - p_x m_y}{q_x p_y - q_y p_x} \quad (6)$$

Next, we perform a similar operation in the signal space. Using s_1 , s_2 , and s_3 , we construct another two vectors v and w , which mirror the vectors p and q used in the meaning space:

$$v = s_2 - s_1 \quad (7)$$

$$w = s_3 - s_1 \quad (8)$$

The process of induction used to postulate the signal s proceeds by using A and B in the signal space. As A and B capture the relationship between m and its neighbors induction occurs because we assume the same relationship holds between s and the corresponding signals of m 's neighbors:

$$s = Av + Bw \quad (9)$$

Any structured relationship between m and its neighbors will now be reflected in the relationship between s and its neighbors. This procedure for finding an appropriate signal given a meaning represents a bias toward structure preservation. Any existing relationship between the meaning space and the signal space is used to inform the choice of new signals.