Iterated Learning: A Framework for the Emergence of Language

Kenny Smith*
Simon Kirby
Henry Brighton
Language Evolution and Computation Research Unit
Theoretical and Applied Linguistics
School of Philosophy, Psychology and Language Sciences
University of Edinburgh
Adam Ferguson Building
40 George Square
Edinburgh, EH8 9LL
United Kingdom
{kenny,simon,henryb}@ling.ed.ac.uk

Abstract Language is culturally transmitted. Iterated learning, the process by which the output of one individual's learning becomes the input to other individuals' learning, provides a framework for investigating the cultural evolution of linguistic structure. We present two models, based upon the iterated learning framework, which show that the poverty of the stimulus available to language learners leads to the emergence of linguistic structure. Compositionality is language's adaptation to stimulus poverty.

Keywords
Iterated learning, cultural evolution, language, compositionality

1 Introduction

Linguists traditionally view language as the consequence of an innate "language instinct" [17]. It has been suggested that this language instinct evolved, via natural selection, for some social function—perhaps to aid the communication of socially relevant information such as possession, beliefs, and desires [18], or to facilitate group cohesion [9]. However, the view of language as primarily a biological trait arises from the treatment of language learners as isolated individuals. We argue that language should be more properly treated as a culturally transmitted system. Pressures acting on language during its cultural transmission can explain much of linguistic structure. Aspects of language that appear baffling when viewed from the standpoint of individual acquisition emerge straightforwardly if we take the cultural context of language acquisition into account. While we are sympathetic to attempts to explain the biological evolution of the language faculty, we agree with Jackendoff that "[i]f some aspects of linguistic behavior can be predicted from more general considerations of the dynamics of communication [or cultural transmission] in a community, rather than from the linguistic capacities of individual speakers, then they should be" [11, p. 101].

We present the *iterated learning model* as a tool for investigating the cultural evolution of language. Iterated learning is the process by which one individual's competence is acquired on the basis of observations of another individual's behavior, which is determined by that individual's competence.[1] This model of cultural transmission has proved particularly useful in studying the evolution of language. The primary goal of this article is to introduce the notion of iterated learning and demonstrate that it pro-

[*] To whom all correspondence should be addressed.

[1] There may be some confusion about the use of the terms "culture" and "observation" here. For our purposes, the process of iterated learninggives rise to culture. We use "observation" in the sense of *observational learning* and to contrast with other forms of learning such as reinforcement learning.

vides a new adaptive mechanism for language evolution. Language itself can adapt on a cultural time scale, and the process of language adaptation leads to the characteristic structure of language. To this end, we present two models. Both attempt to explain the emergence of *compositionality*, a fundamental structural property of language. In doing so they demonstrate the utility of the iterated learning approach to the investigation of language origins and evolution.

In a compositional system the meaning of a signal is a function of the meaning of its parts and the way they are put together [15]. The morphosyntax of language exhibits a high degree of compositionality. For example, the relationship between the string *John walked* and its meaning is not completely arbitrary. It is made up of two components: a noun (*John*) and a verb (*walked*). The verb is also made up of two components: a stem and a past-tense ending. The meaning of *John walked* is thus a function of the meaning of its parts.

The syntax of language is recursive—expressions of a particular syntactic category can be embedded within larger expressions of the same syntactic category. For example, sentences can be embedded within sentences—the sentence *John walked* can be embedded within the larger sentence *Mary said John walked*, which can in turn be embedded within the sentence *Harry claimed that Mary said John walked*, and so on. Recursive syntax allows the creation of an infinite number of utterances from a small number of rules. Compositionality makes the interpretation of previously unencountered utterances possible—knowing the meaning of the basic elements and the effects associated with combining them enables a user of a compositional system to deduce the meaning of an infinite set of complex utterances.

Compositional language can be contrasted with noncompositional, or *holistic*, communication, where a signal stands for the meaning as a whole, with no subpart of the signal conveying any part of the meaning in and of itself. Animal communication is typically viewed as holistic—no subpart of an alarm call or a mating display stands for part of the meaning "there's a predator about" or "come and mate with me." Wray [25] suggests that the protolanguage of early hominids was also holistic. We argue that iterated learning provides a mechanism for the transition from holistic protolanguage to compositional language.

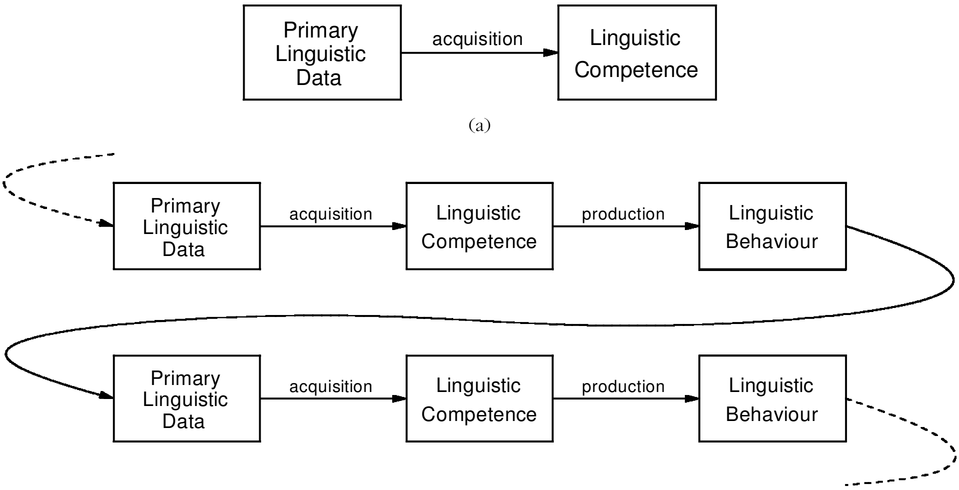
In the first model presented in this article, insights gained from the iterated learning framework suggest a mathematical analysis. This model predicts when compositional language will be more stable than noncompositional language. In the second model, techniques adopted from artificial life are used to investigate the transition, through purely cultural processes, from noncompositional to compositional language. These models reveal two key determinants of linguistic structure:

STIMULUS POVERTY: *The poverty of the stimulus available to language learners during cultural transmission drives the evolution of structured language—without this stimulus poverty, compositional language will not emerge.*

STRUCTURED SEMANTIC REPRESENTATIONS: *Compositional language is most likely to evolve when linguistic agents perceive the world as structured—structured prelinguistic representation facilitates the cultural evolution of structured language.*

2 Two Views of Language

In the dominant paradigm in linguistics (formulated and developed by Noam Chomsky [5, 7]), language is viewed as an aspect of individual psychology. The object of interest is the internal linguistic competence of the individual, and how this linguistic competence is derived from the noisy fragments and deviant expressions of speech children observe.



(a)

(b)

Figure 1. (a) The Chomskyan paradigm. Acquisition procedures, constrained by universal grammar and the language acquisition device, derive linguistic competence from linguistic data. Linguistic behavior is considered to be epiphenomenal. (b) Language as a cultural phenomenon. As in the Chomskyan paradigm, acquisition based on linguistic data leads to linguistic competence. However, we now close the loop—competence leads to behavior, which contributes to the linguistic data for the next generation.

External linguistic behavior (the set of sounds an individual actually produces during their lifetime) is considered to be epiphenomenal, the uninteresting consequence of the application of this linguistic competence to a set of contingent communicative situations. This framework is sketched in Figure 1a. From this standpoint, much of the structure of language is puzzling—how do children, apparently effortlessly and with virtually universal success, arrive at a sophisticated knowledge of language from exposure to sparse and noisy data? In order to explain language acquisition in the face of this poverty of the linguistic stimulus, the Chomskyan program postulates a sophisticated, genetically encoded language organ of the mind, consisting of a *universal grammar*, which delimits the space of possible languages, and a *language acquisition device*, which guides the "growth of cognitive structures [linguistic competence] along an internally directed course under the triggering and partially shaping effect of the environment" [6, p. 34]. Universal grammar and the language acquisition device impose structure on language, and linguistic structure is explained as a consequence of some innate endowment.

Following ideas developed by Hurford [10], we view language as an essentially cultural phenomenon. An individual's linguistic competence is derived from data that is itself a consequence of the linguistic competence of another individual. This framework is sketched in Figure 1b. In this view, the burden of explanation is lifted from the postulated innate language organ—much of the structure of language can be explained as a result of pressures acting on language during the repeated production of linguistic forms and induction of linguistic competence on the basis of these forms. In this article we will show how the poverty of the stimulus available to language learners is the cause of linguistic structure, rather than a problem for it.

3 The Iterated Learning Model

The iterated learning model [13, 3] provides a framework for studying the cultural evolution of language. The iterated learning model in its simplest form is illustrated in

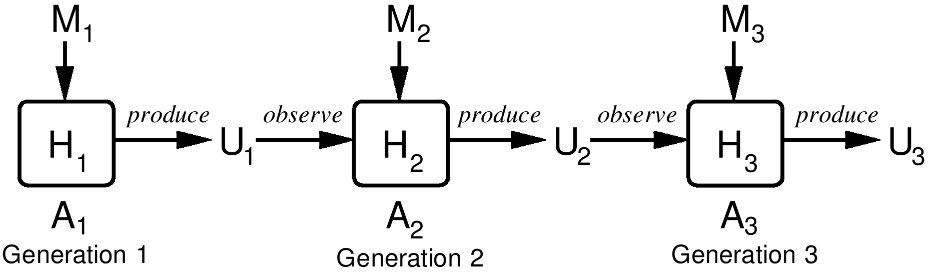


Figure 2. The iterated learning model. The *i*th generation of the population consists of a single agent A_i who has hypothesis H_i. Agent A_i is prompted with a set of meanings M_i. For each of these meanings the agent produces an utterance using H_i. This yields a set of utterances U_i. Agent A_{i+1} observes U_i and forms a hypothesis H_{i+1} to explain the set of observed utterances. This process of observation and hypothesis formation constitutes learning.

Figure 2. In this model the hypothesis H_i corresponds to the linguistic competence of individual i, whereas the set of utterances U_i corresponds to the linguistic behavior of individual i and the primary linguistic data for individual $i + 1$.

We make the simplifying idealization that cultural transmission is purely vertical—there is no horizontal, intragenerational cultural transmission. This simplification has several consequences. Firstly, we can treat the population at any given generation as consisting of a single individual. Secondly, we can ignore the intragenerational communicative function of language. However, the iterated learning framework does not rule out either intra-generational cultural transmission (see [16] for an iterated learning model with both vertical and horizontal transmission, or [1] for an iterated learning model where transmission is purely horizontal) or a focus on communicative function (see [22] for an iterated learning model focusing on the evolution of optimal communication within a population).

In most implementations of the iterated learning model, utterances are treated as meaning-signal pairs. This implies that meanings, as well as signals, are observable. This is obviously an oversimplification of the task facing language learners, and should be treated as shorthand for the process whereby learners infer the communicative intentions of other individuals by observation of their behavior. Empirical evidence suggests that language learners have a variety of strategies for performing this kind of inference (see [2] for a review). We will assume for the moment that these strategies are error-free, while noting that the consequences of weakening this assumption are a current and interesting area of research (see, for example, [23, 20, 24]).

This simple model proves to be a powerful tool for investigating the cultural evolution of language. We have previously used the iterated learning model to explain the emergence of particular word-order universals [12], the regularity-irregularity distinction [13], and recursive syntax [14]; here we will focus on the evolution of compositionality. The evolution of compositionality provides a test case to evaluate the suitability of techniques from mathematics and artificial life in general, and the iterated learning model in particular, to tackling problems from linguistics.

4 The Cultural Evolution of Compositionality

We view language as a mapping between meanings and signals. A compositional language is a mapping that preserves neighborhood relationships—neighbouring meanings will share structure, and that shared structure in meaning space will map to shared structure in the signal space. For example, the sentences *John walked* and *Mary walked* have parts of an underlying semantic representation in common (the notion of someone having carried out the act of walking at some point in the past) and will be near one another in semantic representational space. This shared semantic structure leads to shared signal structure (the inflected verb *walked*)—the relationship between the two sentences in semantic and signal space is preserved by the compositional mapping from meanings to signals. A holistic language is one that does not preserve such relationships—as the structure of signals does not reflect the structure of the underlying meaning, shared structure in meaning space will not necessarily result in shared signal structure.

In order to model such systems we need representations of meanings and signals. For both models outlined in this article meanings are represented as points in an F-dimensional space where each dimension has V discrete values, and signals are represented as strings of characters of length 1 to $l_{\max}$, where the characters are drawn from some alphabet Σ. More formally, the meaning space $\mathcal{M}$ and signal space $\mathcal{S}$ are given by

$$\mathcal{M} = \left\{ \left(f_1 \quad f_2 \quad \ldots \quad f_F \right) : 1 \leq f_i \leq V \text{ and } 1 \leq i \leq F \right\}$$

$$\mathcal{S} = \{ w_1 w_2 \ldots w_l \colon w_i \in \Sigma \text{ and } 1 \leq l \leq l_{\max} \}$$

The world, which provides communicatively relevant situations for agents in our models, consists of a set of N objects, where each object is labeled with a meaning drawn from the meaning space $\mathcal{M}$. We will refer to such a set of labeled objects as an *environment*.

In the following sections two iterated learning models will be presented. In the first model a mathematical analysis shows that compositional language is more stable than holistic language, and therefore more likely to emerge and persist over cultural time, in the presence of stimulus poverty and structured semantic representations. In the second model, computational simulation demonstrates that compositional language can emerge from an initially holistic system. Compositional language is most likely to evolve given stimulus poverty and a structured environment.

4.1 A Mathematical Model

We will begin by considering, using a mathematical model,[2] how the compositionality of a language relates to its stability over cultural time. For the sake of simplicity, we will restrict ourselves to looking at the two extremes on the scale of compositionality, comparing the stability of perfectly compositional language and completely holistic language.

4.1.1 Learning Holistic and Compositional Languages

We can construct a holistic language L_h by simply assigning a random signal to each meaning. More formally, each meaning $m \in \mathcal{M}$ is assigned a signal of random length l ($1 \leq l \leq l_{\max}$) where each character is selected at random from Σ. The meaning-signal mapping encoded in this assignment of meanings to signals will not preserve neighborhood relations, unless by chance.

Consider the task facing a learner attempting to learn the holistic language L_h. There is no structure underlying the assignment of signals to meanings. The best strategy here is simply to memorize meaning-signal associations. We can calculate the expected number of meaning-signal pairs our learner will observe and memorize. We will assume that each of the N objects in the environment is labeled with a single meaning selected

[2] This model is described in greater detail in [3].

randomly from the meaning space $\mathcal{M}$. After R observations of randomly selected objects paired with signals, an individual will have learned signals for a set of O meanings. We can calculate the probability that any arbitrary meaning $m \in \mathcal{M}$ will be included in O, $\Pr(m \in O)$, with

$$\begin{aligned}\Pr(m \in O) = \sum_{x=1}^{N} &\text{(probability that } m \text{ is used to label } x \text{ objects)}\\ &\times \text{ (probability of observing an utterance being produced}\\ &\quad \text{for at least one of those } x \text{ objects after } R \text{ observations)}\end{aligned}$$

In other words, the probability of a learner observing a meaning m paired with a signal is simply the probability that m is used to label one or more of the N objects in the environment *and* the learner observes an utterance being produced for at least one of those objects.

When called upon to produce utterances, such learners will only be able to reproduce meaning-signal pairs they themselves observed. Given the lack of structure in the meaning-signal mapping, there is no way to predict the appropriate signal for a meaning unless that meaning-signal pair has been observed. We can therefore calculate E_h, the expected number of meanings an individual will be able to express after observing some subset of a holistic language, which is simply the probability of observing any particular meaning multiplied by the number of possible meanings:

$$E_h = \Pr(m \in O) \cdot V^F$$

We can perform similar calculations for a learner attempting to acquire a perfectly compositional language. As discussed above, a perfectly compositional language preserves neighborhood relations in the meaning-signal mapping. We can construct such a language L_c for a given set of meanings $\mathcal{M}$ using a lookup table of subsignals (strings of characters that form part of a signal), where each subsignal is associated with a particular feature value. For each $m \in \mathcal{M}$ a signal is constructed by concatenating the appropriate subsignal for each feature value in m.

How can a learner best acquire such a language? The optimal strategy is to memorize feature-value–signal-substring pairs. After observing R randomly selected objects paired with signals, our learner will have acquired a set of observations of feature values for the ith feature, O_{f_i}. The probability that an arbitrary feature value v in included in O_{f_i} is given by $\Pr\left(v \in O_{f_i}\right)$:

$$\begin{aligned}\Pr\left(v \in O_{f_i}\right) = \sum_{x=1}^{N} &\text{(probability that } v \text{ is used to label } x \text{ objects)}\\ &\times \text{ (probability of observing an utterance being produced}\\ &\quad \text{for at least one of those } x \text{ objects after } R \text{ observations)}\end{aligned}$$

We will assume the strongest possible generalization capacity. Our learner will be able to express a meaning if it has viewed all the feature values that make up that meaning, paired with signal substrings. The probability of our learner being able to express an arbitrary meaning made up of F feature values is then given by the combined probability of having observed each of those feature values:

$$\Pr\left(v_1 \in O_{f_1} \wedge \cdots \wedge v_F \in O_{f_F}\right) = \Pr\left(v \in O_{f_i}\right)^F$$

We can now calculate E_c, the number of meanings our learner will be able to express after viewing some subset of a compositional language, which is simply the probability of being able to express an arbitrary meaning multiplied by N_{used}, the number of meanings used when labeling the N objects:

$$E_c = \Pr\left(v \in O_{f_i}\right)^F \cdot N_{\text{used}}$$

We therefore have a method for calculating the expected expressivity of a learner presented with L_h or L_c. This in itself is not terribly useful. However, within the iterated learning framework we can relate expressivity to *stability*. We are interested in the dynamics arising from the iterated learning of languages. The stability of a language determines how likely it is to persist over iterated learning events.

If an individual is called upon to express a meaning they have not observed being expressed, they have two options. Firstly, they could simply not express. Alternatively, they could produce some random signal. In either case, any association between meaning and signal that was present in the previous individual's hypothesis will be lost—part of the meaning-signal mapping will change. A shortfall in expressivity therefore results in instability over cultural time. We can relate the expressivity of a language to the stability of that language over time by $S_h \propto E_h/N$ and $S_c \propto E_c/N$. Stability is simply the proportion of meaning-signal mappings encoded in an individual's hypothesis that are also encoded in the hypotheses of subsequent individuals.

We will be concerned with the *relative stability* S of compositional languages with respect to holistic languages, which is given by

$$S = \frac{S_c}{S_c + S_h}$$

When $S = 0.5$, compositional languages and holistic languages are equally stable and we therefore expect them to emerge with equal frequency over cultural time. When $S > 0.5$, compositional languages are more stable than holistic languages, and we expect them to emerge more frequently, and persist for longer, than holistic languages. $S < 0.5$ corresponds to the situation where holistic languages are more stable than compositional languages.

4.1.2 The Impact of Meaning-Space Structure and the Bottleneck

The relative stability S depends on the number of dimensions in the meaning space (F), the number of possible values for each feature (V), the number of objects in the environment (N), and the number of observations each learner makes (R). Unless each learner makes a large number of observations (R is very large), or there are few objects in the environment (N is very small), there is a chance that agents will be called upon to express a meaning they themselves have never observed paired with a signal. This is one aspect of the poverty of the stimuli facing language learners—the set of utterances of any human language is arbitrarily large, but a child must acquire their linguistic competence based on a finite number of sentences. We will refer to this aspect of the poverty of stimulus as the *transmission bottleneck*. The severity of the transmission bottleneck depends on the number of observations each learner makes (R) and the number of objects in the environment (N). It is convenient to refer instead to the degree of object coverage (b), which is simply the proportion of all N objects observed after R observations—b gives the severity of the transmission bottleneck.

Together F and V specify the degree of what we will term *meaning-space structure*. This in turn reflects the sophistication of the semantic representation capacities of agents—we follow Schoenemann in that we "take for granted that there are fea-

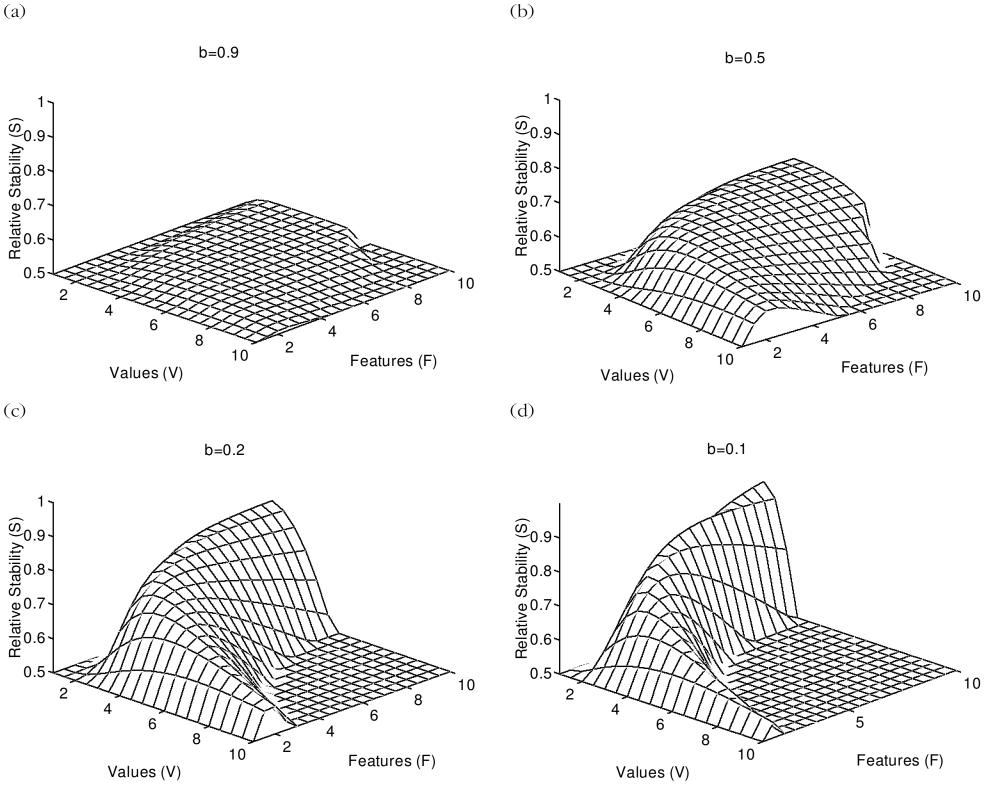


Figure 3. The relative stability of compositional language in relation to meaning-space structure (in terms of *F* and *V*), and the transmission bottleneck *b* (note that low *b* corresponds to a tight bottleneck). The relative stability advantage of compositional language increases as the bottleneck tightens, but only when the meaning space exhibits certain kinds of structure (in other words, for particular numbers of features and values). *b* gives the severity of transmission bottleneck, with low *b* corresponding to a tight bottleneck.

tures of the real world which exist regardless of whether an organism perceives them . . . [d]ifferent organisms will divide up the world differently, in accordance with their unique evolved neural systems . . . [i]ncreasing semantic complexity therefore refers to an increase in the number of divisions of reality which a particular organism is aware of" [19, p. 318]. Schoenemann argues that high semantic complexity can lead to the emergence of syntax. The iterated learning model can be used to test this hypothesis. We will vary the degree of structure in the meaning space, together with the transmission bottleneck *b*, while holding the number of objects in the environment (*N*) constant. The results of these manipulations are shown in Figure 3.

There are two key results to draw from these figures:

1. The relative stability *S* is at a maximum for small bottleneck sizes. Holistic languages will not persist over time when the bottleneck on cultural transmission is tight. In contrast, compositional languages are generalizable, due to their structure, and remain relatively stable even when a learner only observes a small subset of the language of the previous generation. The poverty-of-the-stimulus "problem" is in fact required for linguistic structure to emerge.

2. A large stability advantage for compositional language (high *S*) only occurs when the meaning space exhibits a certain degree of structure (i.e., when there are many features and/or values), suggesting that structure in the conceptual space of language learners is a requirement for the evolution of compositionality. In such

meaning spaces, distinct meanings tend to share feature values. A compositional system in such a meaning space will be highly generalizable—the signal associated with a meaning can be deduced from observation of other meanings paired with signals, due to the shared feature values. However, if the meaning space is too highly structured, then the stability S is low, as few distinct meanings will share feature values and the advantage of generalization is lost.

The first result outlined above is to some extent obvious, although it is interesting to note that the apparent poverty-of-the-stimulus problem motivated the strongly innatist Chomskyan paradigm. The advantage of the iterated learning approach is that it allows us to quantify the degree of advantage afforded by compositional language, and investigate how other factors, such as meaning-space structure, affect the advantage afforded by compositionality.

4.2 A Computational Model

The mathematical model outlined above, made possible by insights gained from viewing language as a culturally transmitted system, predicts that compositional language will be more stable than holistic language when (1) there is a bottleneck on cultural transmission and (2) linguistic agents have structured representations of objects. However, the simplifications necessary to the mathematical analysis preclude a more detailed study of the dynamics arising from iterated learning. What happens to languages of intermediate compositionality during cultural transmission? Can compositional language emerge from initially holistic language, through a process of cultural evolution? We can investigate these questions using techniques from artificial life, by developing a multi-agent computational implementation of the iterated learning model.

4.2.1 A Neural Network Model of a Linguistic Agent

We have previously used neural networks to investigate the evolution of holistic communication [22]. In this article we extend this model to allow the study of the cultural evolution of compositionality.[3] As in the mathematical model, meanings are represented as points in F-dimensional space where each dimensions has V distinct values, and signals are represented as strings of characters of length 1 to $l_{\max}$, where the characters are drawn from the alphabet Σ.

Agents are modeled using networks consisting of two sets of nodes. One set represents meanings and partially specified components of meanings ($\mathcal{N}_M$), and the other represents signals and partially specified components of signals ($\mathcal{N}_S$). These nodes are linked by a set $\mathcal{W}$ of bidirectional connections connecting every node in $\mathcal{N}_M$ with every node in $\mathcal{N}_S$.

As with the mathematical model, meanings are sets of feature values, and signals are strings of characters. Components of a meaning specify one or more feature values of that meaning, with unspecified values being marked as a wildcard $*$. For example, the meaning $(2 \quad 1)$ has three possible components: the fully specified $(2\ 1)$ and the partially specified $(2 \quad *)$ and $(*\ 1)$. These components can be grouped together into ordered sets, which constitute an analysis of a meaning. For example, there are three possible analyses of the meaning $(2 \quad 1)$—the one-component analysis $\{(2\ 1)\}$, and two two-component analyses, which differ in order, $\{(2\ *), (*\ 1)\}$ and $\{(*\ 1), (2\ *)\}$. Similarly, components of signals can be grouped together to form an analysis of a signal. This representational scheme allows the networks to exploit the structure of meanings and signals. However, they are not forced to do so.

[3] We refer the reader to [21] for a more thorough description of this model.

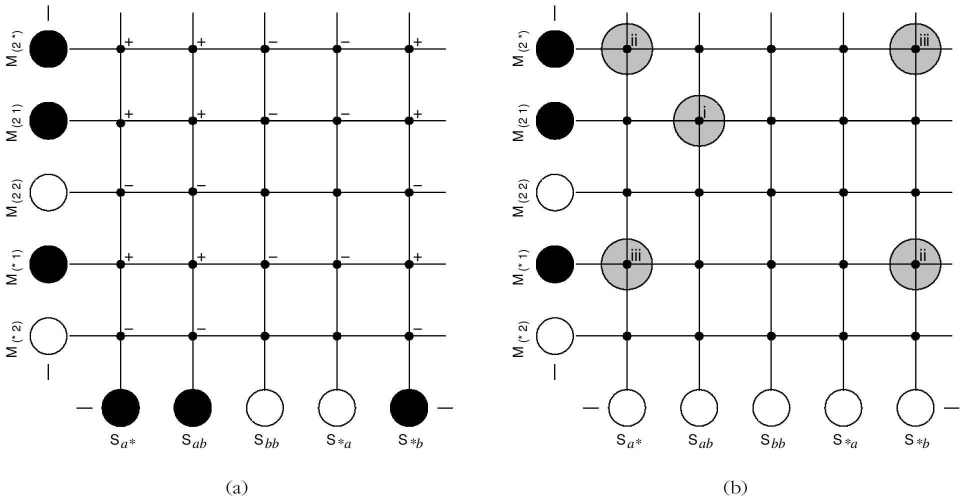


Figure 4. Nodes with an activation of 1 are represented by large filled circles. Small filled circles represent weighted connections. (a) Storage of the meaning-signal pair $\langle (2\ \ 1), ab\rangle$. Nodes representing components of (2 1) and *ab* have their activations set to 1. Connection weights are then either incremented (+), decremented (−), or left unchanged. (b) Retrieval of three possible analyses of $\langle (2\ 1), ab\rangle$. The relevant connection weights are highlighted in gray. The strength *g* of the one-component analysis $\langle \{(2\ 1)\}, \{ab\}\rangle$ depends of the weight of connection i. The strength *g* for the two-component analysis $\langle \{(2\ *), (*\ 1)\}, \{a*, *b\}\rangle$ depends on the weighted sum of two connections, marked ii. The *g* for the alternative two-component analysis $\langle \{(2\ *), (*\ 1)\}, \{*b, a*\}\rangle$ is given by the weighted sum of the two connections marked iii.

Learners observe meaning-signal pairs. During a single learning episode a learner will store a pair $\langle m, s\rangle$ in its network. The nodes in $\mathcal{N}_M$ corresponding to all possible components of the meaning m have their activations set to 1, while all other nodes in $\mathcal{N}_M$ have their activations set to 0. Similarly, the nodes in $\mathcal{N}_S$ corresponding to the possible components of s have their activations set to 1. Connection weights in $\mathcal{W}$ are then adjusted according to the rule

$$\Delta W_{xy} = \begin{cases} +1 & \text{if } a_x = a_y = 1 \\ -1 & \text{if } a_x \neq a_y \\ 0 & \text{otherwise} \end{cases}$$

where W_{xy} gives the weight of the connection between nodes x and y and a_x gives the activation of node x. The learning procedure is illustrated in Figure 4a.

In order to produce an utterance, agents are prompted with a meaning m and required to produce a signal s. All possible analyses of m are considered in turn with all possible analyses of every $s \in \mathcal{S}$. Each meaning-analysis–signal-analysis pair is evaluated according to

$$g(\langle m, s\rangle) = \sum_{i=1}^{C} \omega(c_{mi}) \cdot W_{c_{mi}c_{si}}$$

where the sum is over the C components of the analysis, c_{mi} is the ith component of m, and $\omega(x)$ is a weighting function that gives the non-wildcard proportion of x. This process is illustrated in Figure 4b. The meaning-analysis–signal-analysis pair with the highest g is returned as the network's utterance.

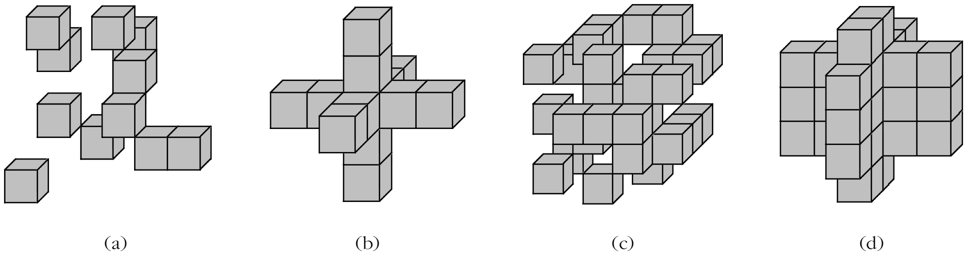


Figure 5. We will present results for the case where $F = 3$ and $V = 5$. This defines a three-dimensional meaning space. We highlight the meanings selected from that space with gray. Meaning space (a) is a low-density, unstructured environment. (b) is a low-density, structured environment. (c) and (d) are unstructured and structured high-density environments.

4.2.2 Environment Structure

In the mathematical model outlined above, the environment consisted of a set of objects labeled with meanings drawn at random from the space of possible meanings. In the computational model we can relax this assumption and investigate how nonrandom assignment of meanings to objects affects linguistic evolution. As before, an environment consists of a set of objects labeled with meanings drawn from the meaning space $\mathcal{M}$. The number of objects in the environment gives the *density* of that environment—environments with few objects will be termed low-density, whereas environments with many objects will be termed high-density. When meanings are assigned to objects at random, we will say the environment is *unstructured*. When meanings are assigned to objects in such a way as to minimize the average inter-meaning Hamming distance, we will say the environment is *structured*. Sample low- and high-density environments are shown in Figure 5. Note the new usage of the term "structured"—whereas in the mathematical model we were concerned with structure in the meaning space, given by F and V, we are now concerned with the degree of structure in the environment. Different levels of environment structure are possible within a meaning space of a particular structure.

4.2.3 The Effect of Environment Structure and the Bottleneck

The network model of a language learner-producer is plugged into the iterated learning framework. We will manipulate three factors—the presence or absence of a bottleneck, the density of the environment, and the degree of structure in the environment.

Our measure of compositionality is simply the degree of correlation between the distance between pairs of meanings and the distance between the corresponding pairs of signals. In order to measure the compositionality of an agent's language we first take all possible pairs of meanings from the environment, $\langle m_i, m_{j \neq i} \rangle$. We then find the signals these meanings map to in the agent's language, $\langle s_i, s_j \rangle$. This yields a set of meaning-meaning pairs, each of which is matched with a signal-signal pair. For each of these pairs, the distance between the meanings m_i and m_j is taken as the Hamming distance, and the distance between the signals s_i and s_j is taken as the Levenstein (string edit) distance.[4] This gives a set of distance pairs, reflecting the distance between all possible pairs of meanings and the distance between the corresponding pairs of signals. A Pearson product-moment correlation is then run on this set, giving the correlation between the meaning-meaning distances and the associated signal-signal distances. This correlation is our measure of compositionality. Perfectly compositional languages have a compositionality of 1, reflecting the fact that compositional languages

[4] Levenstein distance is a measure of string similarity. It is defined as the size of the smallest set of edits (replacements, deletions, or insertions) that could transform one string to the other.

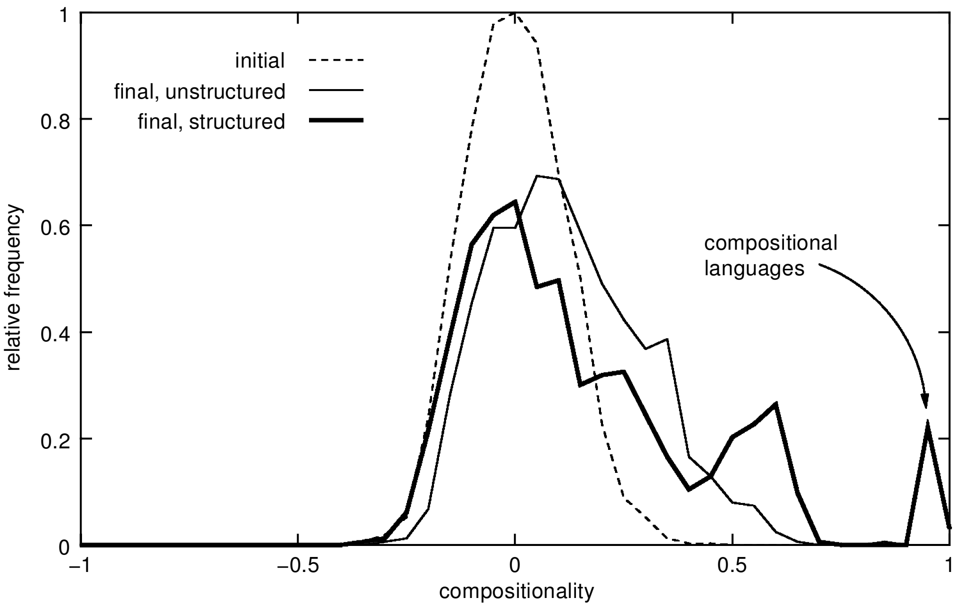


Figure 6. The relative frequency of initial and final systems of varying degrees of compositionality when there is no bottleneck on cultural transmission. The results shown here are for the low-density environments given in Figure 5. The initial languages are generally holistic. Some final languages exhibit increased levels of compositionality. Highly compositional languages are infrequent.

preserve distance relationships when mapping between meanings and signals. Holistic languages have a compositionality of approximately 0—holistic mappings are random, and therefore fail to preserve distance relationships when mapping between meaning space and signal space.

Figure 6 plots the frequency by compositionality of initial and final systems in 1000 runs of the iterated learning model, in the case where there is no bottleneck on cultural transmission. The initial agent has the maximum-entropy hypothesis—all meaning-signal pairs are equally probable. The learner at each generation is exposed to the complete language of the previous generation—the adult is required to produce utterances for every object in the environment. Each run was allowed to proceed to a stable state.

Two main results are apparent from Figure 6:

1. The majority of the final, stable systems are holistic.
2. Highly compositional systems occur infrequently, and only when the environment is structured.

In the absence of a bottleneck on cultural transmission, the compositionality of the final systems is sensitive to initial conditions. The majority of the initial holistic systems are stable. This can be contrasted with the result shown in Figure 3a, where compositional languages have a slight stability advantage for most meaning spaces when the transmission bottleneck is very wide ($b = 0.9$). When there is *no* bottleneck on transmission ($b = 1.0$), most holistic systems are perfectly stable. However, the initial system may exhibit, purely by chance, a slight tendency to express a given feature

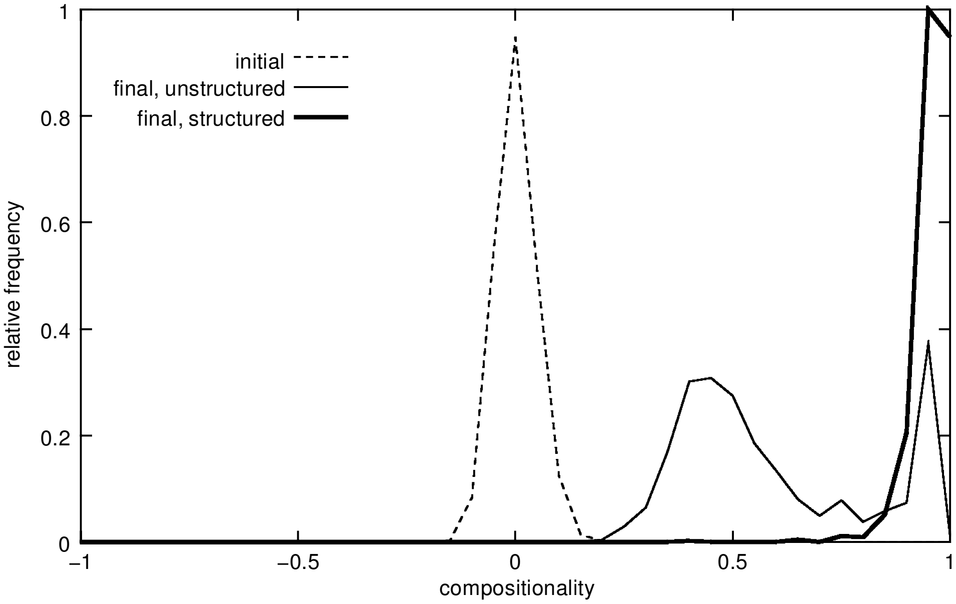


Figure 7. Frequency by compositionality when there is a bottleneck on cultural transmission. The results shown here are for the high-density environments given in Figure 5c and d. The initial languages are holistic. The final languages are compositional, with highly compositional languages occurring frequently.

value with a certain substring. This compositional tendency can spread, over iterated learning events, to other parts of the system, which can in turn have further knock-on consequences. The potential for spread of compositional tendencies is greatest in structured environments—in such environments, distinct meanings are more likely to share feature values than in unstructured environments. However, this spread of compositionality is unlikely to lead to a perfectly compositional language.

Figure 7 plots the frequency by compositionality of initial and final systems in 1000 runs of the iterated learning model, in the case where there is a bottleneck on cultural transmission ($b = 0.4$). Learners will therefore only see a subset of the language of the previous generation. Whereas in the no-bottleneck condition each run proceeded to a stable state, in the bottleneck condition runs were stopped after 50 generations. There is no such thing as a truly stable state when there is a bottleneck on cultural transmission. For example, if all R utterances an individual observes refer to the same object, then any structure in the language of the previous generation will be lost. However, the final states here were as close as possible to stable. Allowing the runs to continue for several hundred more generations results in a very similar distribution of languages.

Two main results are apparent from Figure 7:

1. When there is a bottleneck on cultural transmission, highly compositional systems are frequent.
2. Highly compositional systems are more frequent when the environment is structured.

As discussed with reference to the mathematical model, only highly compositional systems are stable through a bottleneck. The results from the computational model

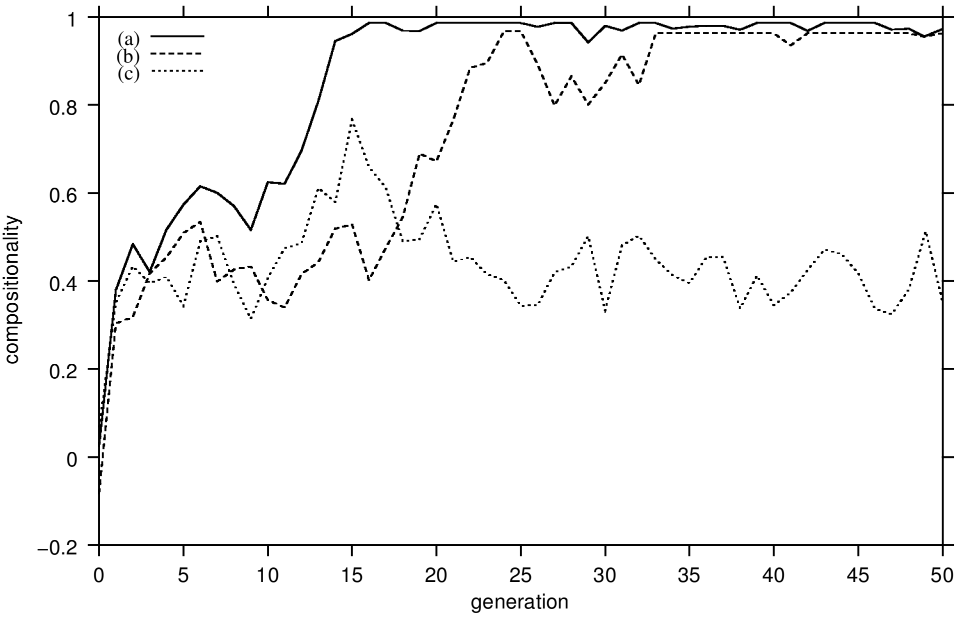


Figure 8. Compositionality by time (in generations) for three runs in high-density environments. The solid line (a) shows the development from an initially holistic system to a compositional language for a run in a structured environment. Thes dashed and dotted lines (b) and (c) show the development of systems in unstructured environments. The language plotted in (b) eventually becomes highly compositional, whereas the system in (c) remains partially compositional. Only the first 50 generations are plotted here, in order to focus on the development of the systems from the initial holistic state.

bear this out—over time, language adapts to the pressure to be generalizable, until the language becomes highly compositional, highly generalizable, and highly stable. Highly compositional languages evolve most frequently when the environment is structured, because in a structured environment the advantage of compositionality is at a maximum—each meaning shares feature values with several other meanings, and a language mapping these feature values to a signal substring is highly generalizable.

Figure 8 plots the compositionality by generation for three runs of the iterated learning model. The behavior of these runs is characteristic of the majority of simulations. Figure 8a and b show the development from initially random, holistic systems to compositional languages in structured and unstructured environments. In both these runs a partly compositional, partly irregular language rapidly develops, resulting in a rapid increase in compositionality. This partially compositional system persists for a short time, before developing into a highly regular compositional language where each feature value maps consistently to a particular subsignal. The transition is more rapid in the structured environment. In the structured environment, distinct meanings share feature values with several other meanings and as a consequence compositional languages are highly generalizable. Additionally, distinct meanings vary along a limited number of dimensions, which facilitates the spread of consistent, regular mappings from feature values to signal substrings. In Figure 8c a partially compositional language develops from the initial random mapping, but fails to become fully compositional. The lack of structure in the environment hinders the development of consistent compositional mappings and allows unstable, idiosyncratic meaning-signal mappings to persist.

5 Conclusions

Language can be viewed as a consequence of an innate language organ. This view of language has been advanced to explain the near-universal success of language acquisition in the face of the poverty of the stimulus available to language learners. The innatist position solves this apparent conundrum by attributing much of the structure of language to the language organ—an individual's linguistic competence develops along an internally determined course, with the linguistic environment simply triggering the growth of the appropriate cognitive structures. If we take this view, we can form an evolutionary account that explains linguistic structure as a biological adaptation to social function—language is socially useful, and the language organ yields a fitness payoff.

However, we have presented an alternative approach. We focus on the cultural transmission of language. We can then form an account that explains much of linguistic structure as a cultural adaptation, by language, to pressures arising during repeated production and acquisition of language. This kind of approach highlights the *situatedness* of language-using agents in an environment—in this case, a socio-cultural environment made up of the behavior of other agents. We have presented the iterated learning model as a framework for studying the cultural evolution of language in this context, and have focused here on the cultural evolution of compositionality. The models presented reveal two key factors in the cultural evolution of compositional language.

Firstly, compositional language emerges when there is a bottleneck on cultural transmission—compositionality is an adaptation by language that allows it to slip through the transmission bottleneck. The transmission bottleneck constitutes one aspect of the poverty-of-the-stimulus problem. This result is therefore surprising. The poverty of the stimulus motivated a strongly innatist position on language acquisition. However, closer investigation within the iterated learning framework reveals that the poverty of the stimulus does not force us to conclude that linguistic structure must be located in the language organ—on the contrary, the emergence of linguistic structure through cultural processes *requires* the poverty of the stimulus.

The second key factor is the availability of structured semantic representations to language learners—Schoenemann's semantic complexity [19]. The advantage of compositionality is at a maximum when language learners perceive the world as structured. If objects are perceived as structured entities and the objects in the environment relate to one another in structured ways, then a generalizable, compositional language is highly adaptive.

Of course, biological evolution still has a role to play in explaining the evolution of language. The iterated learning model is ideal for investigating the cultural evolution of language on a fixed biological substrate, and identifying the cultural consequences of a particular innate endowment. The origins of that endowment then need to be explained, and natural selection for a socially useful language might play some role here. We might indeed then find, as suggested by Deacon, that "the brain has co-evolved with respect to language, but languages have done most of the adapting" [8, p. 122]. The poverty of the stimulus faced by language learners forces language to adapt to be learnable. The transmission bottleneck forces language to be generalizable, and compositional structure is language's adaptation to this problem. This adaptation yields the greatest payoff for language when language learners perceive the world as structured.

References

1. Batali, J. (2002). The negotiation and acquisition of recursive grammars as a result of competition among exemplars. In [4, pp. 111–172].

2. Bloom, P. (2000). *How children learn the meanings of words*. Cambridge, MA: MIT Press.
3. Brighton, H. (2002). Compositional syntax from cultural transmission. *Artificial Life*, *8*, 25–54.
4. Briscoe, E. (Ed.) (2002). *Linguistic evolution through language acquisition: Formal and computational models*. Cambridge, UK: Cambridge University Press.
5. Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
6. Chomsky, N. (1980). *Rules and representations*. London: Basil Blackwell.
7. Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.
8. Deacon, T. (1997). *The symbolic species*. London: Penguin.
9. Dunbar, R. (1996). *Grooming, gossip and the evolution of language*. London: Faber and Faber.
10. Hurford, J. R. (1990). Nativist and functional explanations in language acquisition. In I. M. Roca (Ed.), *Logical issues in language acquisition* (pp. 85–136). Dordrecht, the Netherlands: Foris.
11. Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford, UK: Oxford University Press.
12. Kirby, S. (1999). *Function, selection and innateness: The emergence of language universals*. Oxford, UK: Oxford University Press.
13. 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*, 102–110.
14. Kirby, S. (2002). Learning, bottlenecks and the evolution of recursive syntax. In [4, pp. 173–203].
15. Krifka, M. (2001). Compositionality. In R. A. Wilson & F. Keil (Eds.), *The MIT encyclopaedia of the cognitive sciences*. Cambridge, MA: MIT Press.
16. Livingstone, D., & Fyfe, C. (1999). Modelling the evolution of linguistic diversity. In D. Floreano, J. D. Nicoud, & F. Mondada (Eds.), *Advances in artificial life: Proceedings of the 5th European Conference on Artificial Life* (pp. 704–708). Berlin: Springer.
17. Pinker, S. (1994). *The language instinct*. London: Penguin.
18. Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences*, *13*, 707–784.
19. Schoenemann, P. T. (1999). Syntax as an emergent characteristic of the evolution of semantic complexity. *Minds and Machines*, *9*, 309–346.
20. Smith, A. D. M. (2003). Intelligent meaning creation in a clumpy world helps communication. *Artificial Life*, *9*, 559–574.
21. Smith, K. (2002). *Compositionality from culture: The role of the environment structure and learning bias* (Technical report). Language Evolution and Computation Research Unit, University of Edinburgh.
22. Smith, K. (2002). The cultural evolution of communication in a population of neural networks. *Connection Science*, *14*, 65–84.
23. Steels, L. (1998). The origins of syntax in visually grounded robotic agents. *Artificial Intelligence*, *103*, 133–156.
24. Steels, L., Kaplan, F., McIntyre, A., & Van Looveren, J. (2002). Crucial factors in the origins of word-meaning. In A. Wray (Ed.), *The transition to language* (pp. 252–271). Oxford, UK: Oxford University Press.
25. Wray, A. (1998). Protolanguage as a holistic system for social interaction. *Language and Communication*, *18*, 47–67.