Spontaneous Evolution of Linguistic Structure—An Iterated Learning Model of the Emergence of Regularity and Irregularity

Simon Kirby

***Abstract*—A computationally implemented model of the transmission of linguistic behavior over time is presented. In this model [the iterated learning model (ILM)], there is no biological evolution, natural selection, nor any measurement of the success of the agents at communicating (except for results-gathering purposes). Nevertheless, counter to intuition, significant evolution of linguistic behavior is observed. From an initially unstructured communication system (a protolanguage), a fully compositional syntactic meaning-string mapping emerges. Furthermore, given a nonuniform frequency distribution over a meaning space and a production mechanism that prefers short strings, a realistic distribution of string lengths and patterns of stable irregularity emerges, suggesting that the ILM is a good model for the evolution of some of the fundamental features of human language.**

***Index Terms*—Cultural selection, evolution, grammar induction, iterated learning, language.**

I. Introduction

ONE striking feature of human languages is the structure-preserving nature of the mapping from meanings to signals (and *vice versa*).[1] This feature can be found in every human language, but arguably in no other species' communication system [1]. Structure preservation is particularly striking if we consider sentence structure. The syntax of English, for example, is clearly compositional—that is, the meaning of a sentence is some function of the meanings of the parts of that sentence. For example, we can understand the meaning of the sentence "the band played at Carlops" to the extent that we can understand the constituents of that sentence, such as *the band* or *Carlops*. Similarly, in the morphological paradigms of languages, *regularity* is pervasive. We can see this too as a structure-preserving mapping.[2]

An obvious goal for evolutionary linguistics is, therefore, an understanding of the origins of compositionality and regularity in morphosyntax. Ultimately, we would like to derive these properties from some other feature external to the syntax of human language. For example, one approach is to argue that structure preservation is actually a property of our innate biological endowment and can be explained by appealing to fitness enhancing mutations that were retained in our species through natural selection [3]. Ultimately, these types of explanation typically derive features of the meaning-string mapping from communicative pressures that influenced our protohuman ancestors [4].

This paper follows on from recent computational work that takes a different approach [5]–[13]. Instead of concentrating on the biological evolution of an innate language faculty, this line of research places more explanatory emphasis on languages themselves as adaptive systems. Human languages are arguably unique not only for their compositionality, but also in the way they persist over time through iterated observational learning. That is, information about the mapping between meanings and signals is transmitted from generation to generation of language users through a repeated cycle of use, observation, and induction. Rather than appealing to communicative pressures and natural selection, the suggestion is that structure-preserving mappings emerge from the dynamics of iterated learning.

Using computational models of iterated learning, earlier work demonstrated that structure-preserving maps emerge from unstructured random maps when learners are exposed to a subset of the total range of meanings. The problem with this kind of approach, which this paper addresses, is that it predicts that there should be no stable *ir*regularity in language. This does not map well onto what we know about natural language morphology, where historically stable irregularity is common. The results reported in this paper, however, show that both regularity *and* irregularity are predicted by a more sophisticated version of the iterated learning model (ILM).

II. Overview of the Iterated Learning Model

In order to model the cultural/historical transmission of language from generation to generation, the ILM has four components:

1) a meaning space;
2) a signal space;
3) one or more learning agents;
4) one or more adult agents.

The simulations reported here use only one adult and one learner. Each iteration of the ILM involves the adult agent being given a

Manuscript received July 4, 2000; revised October 30, 2000. This work was supported by the British Academy.

The author is with the Language Evolution and Computation Research Unit, Department of Linguistics, University of Edinburgh, Edinburgh, U.K. (e-mail: simon@ling.ed.ac.uk).

Publisher Item Identifier S 1089-778X(01)03281-7.

[1]By "structure-preserving," here I mean simply that similar meanings tend to map to similar signals. We could formally define degree of structure preservation in terms of a correlation of distances between points in meaning-space and signal-space (i.e., a *topographic* mapping), but this is unnecessary for our purposes here.

[2]Interestingly, recent computational analysis of very large corpora of English usage [2] suggests that this structure preservation can even be seen within the monomorphemic lexicon, whose phonetic structure was previously thought to be arbitrarily related to meaning.

set of randomly chosen meanings to produce signals for. The resulting meaning-signal pairs form training data for the learning agent. After learning, this agent becomes a new adult agent, the previous adult agent is removed, and a new learning agent is introduced. Typically, this cycle is repeated thousands of times or until a clearly stable end-state is reached. Furthermore, the simulation is usually initialized with no language system in place. In other words, the initial agents have no representation of a mapping from meanings to signals at all.

Clearly, the agents in the ILM need at least:

1) an internal representation of language that specifies the ways in which signals can be produced for particular meanings;
2) an algorithm for inducing this representation given examples of meanings and signals;
3) some means of generating signals for meanings that the induced language representation does not include (e.g., in the early stages of the simulation).

Apart from the meaning space, the various components of the simulation reported in this paper are the same as those reported in [7], [9], [10]. The following sections set out each in turn.

A. Meaning Space

In contrast to earlier work, which used a meaning space involving recursively structured hierarchical predicates, a much simpler set of meanings is employed here. Each meaning is simply a vector of values drawn from a finite set. Specifically, for the results reported here, a meaning consists of two components, a and b, both of which can range over five values.[3] Thus, there are 25 possible meanings, (a_0, b_0) to (a_4, b_4) and, within the space of meanings, there is some structure. We can imagine how this simple model of meanings could map onto the world in various ways. For example, one component of the meaning could be an object and the other an action—expressing a very simple one-place predicate. The interpretation of the meaning space is not important, however. It provides us with a very simple model of a structured space of concepts. Usefully, the space can be displayed as a table, as will be done later.

B. Signal Space

The signal space, in common with earlier work, is an ordered linear string of characters drawn from the letters a–z. There is no set limit on string length. These letters cannot be equated simply with the phonemes of natural languages. A better parallel would be with syllables, but even this is not a one-to-one correspondence. Basically, the letters that make up signals are the atomic elements of the language that the grammatical system cannot break down.

C. Language Representation

Obviously there are many possible representations of the mapping between meanings and signals in the spaces described above. To keep the model as general as possible and retain compatibility with earlier work, the language is represented as a simplified form of definite-clause grammar (DCG). Specifically, the grammars consist of context-free rewrite rules in which nonterminals may have a single argument attached to them that conveys semantic information. Thus, a rule has the following form:

$$C\colon \mu \rightarrow \lambda$$

where

- C category label;
- μ meaning structure;
- λ string of terminals (characters from the alphabet) and nonterminals.

There is a special category label S that signifies the start symbol for production. In other words, every legal meaning-string pair must expand an S rule in the language. For these simulations, the meaning structure has one of the following forms:

$$(a_i, b_j) \quad \text{or} \quad a_i \quad \text{or} \quad b_i$$

but, in general, any structure can be used. Each of the meaning elements may either be specified directly or inherited from the value of the meaning structure attached to one of the nonterminals in λ using variable unification.

For example, this rule specifies that the string abc maps to the meaning (a_0, b_0)

$$S\colon (a_0, b_0) \rightarrow \mathrm{abc}$$

as do the following set of rules:

$$S\colon (x, y) \rightarrow A\colon yB\colon x$$
$$A\colon b_0 \rightarrow \mathrm{ab}$$
$$B\colon a_0 \rightarrow \mathrm{c}.$$

An adult agent that had either of these sets of rules would output abc if asked to produce a signal for the meaning (a_0, b_0).

D. Induction Algorithm

In order that the simulation of iterated learning over many generations be a practical prospect, the induction algorithm of the learning agents must be computationally cheap. The heuristic incremental algorithm developed in [7] is used here.

Induction takes place in two steps incrementally for every input meaning-signal pair.

Incorporation: A single rule that covers the input pair is added to the grammar. In this model, a rule is only added if the learning agent cannot already parse the signal component of the pair. This is included to impose a preference for unambiguous languages because the learner will not acquire multiple meanings for one string.

Generalization: The algorithm iteratively tries to integrate the new rule into the existing grammar by looking for possible generalizations. These generalizations are essentially subsumptions over pairs of rules. In other words, the algorithm takes a pair of rules from the grammar and tries to find a more general rule to replace them with (within a set of heuristic constraints). This may be done, for example, by merging category labels, or discovering common strings of right-hand-side constituents. After subsumption, duplicate rules are deleted, and the process is repeated until no more heuristics apply. At this stage the learning agent accepts another input pair to be incorporated.

[3]To check that the model scales up, the results were replicated with a meaning space of 20×20. These are not reported because the large size seems to add little to the simulation and the results are harder to visualize. Note also that previously reported results relating to the emergence of recursion (but not of irregularity) have, in principle, an infinite meaning space.

The details of the algorithm are outlined below (adapted from [9]).

Induction algorithm: Given a meaning μ, a string s, and a grammar g:
i.1 parse s using g, **if** the parse is successful, **then return** g.
i.2 form g', the union of g and $S: \mu \rightarrow s$.
i.3 apply a generalization algorithm to g':
  g.1 take a pair of rules (r_1, r_2) from g'.
  g.2 **if** there is a category label substitution c to c', that would make r_1 identical to r_2, **then** rewrite all c in g' with c', **go to g.5**.
  g.3 **if** r_1 and r_2 could be made identical by "chunking" a substring on either or both their right-hand sides into a new rule or rules, **then** create the new rules, and delete the old ones in g', **go to g.5**.
  g.4 **if** r_1's right-hand side is a proper substring of r_2's and r_1's semantics is identical to either the top level predicate or one of the arguments of r_2's semantics, **then** rewrite r_2 in g' to refer to r_1, **go to g.5**.
  g.5 delete all duplicate rules in g'.
  g.6 if any rules in g' have changed, **go** to g.1
i.4 **return** g'.

The general chunking method for the DCG-type representations is not given here for lack of space, but can be found in [7]. However, for the meaning space used here, practically chunking is pretty simple. Given a pair of rules

$$C: (a_i, b_j) \rightarrow \lambda_1 \lambda_2 \lambda_3$$
$$C: (a_i, b_k) \rightarrow \lambda_1 \lambda_4 \lambda_3$$

where λ_i is any string of terminals/nonterminals, these are deleted and replaced with

$$C: (a_i, x) \rightarrow \lambda_1 C_{\text{new}}: x \lambda_3$$
$$C_{\text{new}}: b_j \rightarrow \lambda_2$$
$$C_{\text{new}}: b_k \rightarrow \lambda_4.$$

The constraints on this process are that λ_2 *and* λ_4 must be nonnull and that λ_1 *or* λ_3 must be nonnull. The obvious equivalent chunking process is applied if it is the b component of the semantics that the rules have in common.

E. Invention Algorithm

At the start of the simulation especially, the adult agents will frequently not have a way of expressing particular meanings. In other words, their grammar will not be able to generate any string-meaning pair for some meanings. In the very first generation of the ILM, the adult agent has no grammar at all and, therefore, no way to express any meaning. In order for any innovation to emerge in the linguistic system that is being transmitted, the agents need some form of creativity—an ability to invent new strings.

The simplest approach would be to generate strings at random whenever an agent must produce a signal for a meaning that its grammar cannot generate. For the most part, in fact, this is the strategy employed in the simulation. However, in some cases this approach is rather implausible. Consider the case of an agent that has a completely regular compositional language, but does not have a word for a particular meaning component. For example, the agent might have words for all the meaning components a_0 to a_4 and words for b_0 to b_3 and might produce signals for whole meanings by concatenating the relevant words for the a component and then the b component. This is obviously a very regular syntactic and humanlike language. However, the agent does not have a way to produce a string for, say, (a_0, b_4). It seems counterintuitive for the agent to generate a completely random new string for this meaning given that it has a word for a_0 and has already induced a compositional rule for concatenating words together to produce sentences. Instead, it seems sensible that an agent in this state should simply invent a new word for b_4.

On the other hand, it is undesirable for the model to allow for agents to introduce compositionality *de novo* in the invention process. To give another example, consider an agent who has signals for the same range of meanings as the agent in the previous example, but instead of using a regular rule to produce sentences, simply lists each meaning and its associated signal unanalyzed in its grammar. In this case, we do not want the invention algorithm to introduce any compositionality itself—instead, an entirely random innovation seems more plausible.

What is needed, then, is an *invention algorithm* that in producing a novel string preserves what structure is already existing in the agent's grammar, but does not introduce any new structure. The general algorithm for any meaning space is informally summarized below: (N.B. If the grammar is empty, then this algorithm cannot be applied, and a random string is generated.)

Invention algorithm: Given a meaning μ and a grammar g that cannot generate a meaning-string pair (μ, s):
i.1 find the μ' most similar to μ for which the pair (μ', s') can be generated (i.e., initially try all meanings with one meaning component the same, then all meanings).
i.2 form μ'', the intersection of μ and μ', with any meaning slots that are different replaced with a unique "difference-flag."
i.3 generate a string s corresponding to μ'' using g, but include a pseudorule of the sort:
  any-category: difference-flag→random-string
i.4 **return** s.

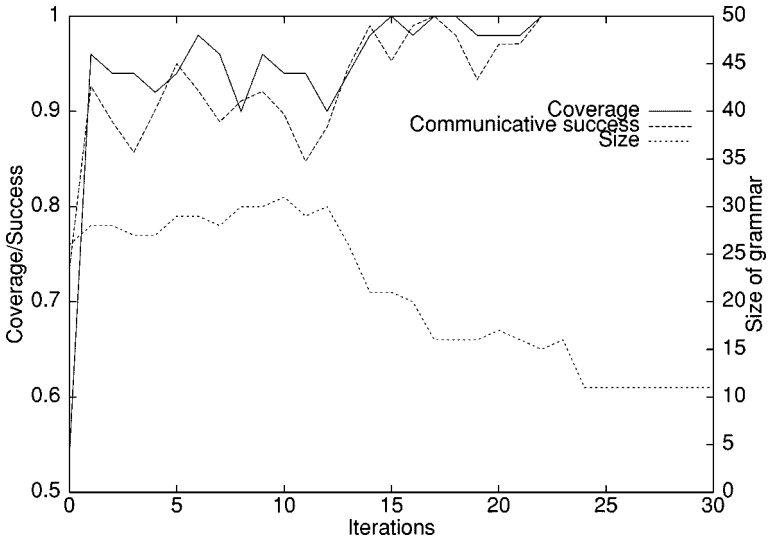

Fig. 1. Emergence of a stable system in the simple simulation. Coverage is a measure of the proportion of utterances that are made without recourse to invention. Success is calculated by testing adult and learner *after* learning in a set of communication games, half of the time with the adult as speaker and half of the time with the learner as the speaker. Number reflects the proportion of successful games (i.e., when the hearer can parse the string with the same meaning as the speaker).

The random string generated as part of the invention algorithm can be produced with various different lengths. In the simulations reported here, the strings were of a random length between one and ten. To encourage consistency in output, the agent uses the meaning-string pair produced by invention as input to one iteration of the induction algorithm. In other words, agents learn from their own innovations.

III. Results

In this section, the results of running the ILM with various different conditions are presented. Typical results are analyzed here, but qualitatively similar behaviors are observed in all replications.

A. Emergence of Regular Compositionality

The simulation described in the previous section was set up with an initial population of one adult and one learner, both with no linguistic knowledge. The adult produces 50 utterances for the learner, each of which is a randomly chosen meaning. Notice that although there are only 25 meanings, the chances of all 25 being exemplified in the learner's input are less than one.[4] This feature of the ILM has been termed the "bottleneck" in linguistic transmission [8]. Although the bottleneck is not particularly tight here, it plays a more important role in later experiments.

Fig. 1 shows a plot of the simulation converging from the initial condition on a language that is stable, covers the complete meaning space, and allows for perfect communication.

Because the meaning space is essentially a two-dimensional vector, we can visualize the language used at a particular point in time in the simulation as a table of strings with the columns corresponding to the a component of the meaning and the rows to the b component. The table below was produced by the first adult agent in the simulation at the end of its "life." Notice that each cell in the table has only one string associated with it. This does not specify the language of that agent completely because there may be more than one way of producing a signal for each meaning. In this experiment, the agents make a random choice whenever that happens.

	a_0	a_1	a_2	a_3	a_4
b_0	s	sq	-	pnj	bjmjimsq
b_1	n	avvcf	jlimgttztp	pclcfho	kebae
b_2	ebhzyuyrl	afeeyykokz	-	pyuhu	hwrpg
b_3	rqbvtggjac	zrdleab	rxktywr	rbq	rkhxpbmx
b_4	drnlblwmo	afqjghvuw	gnbyq	pquztpi	wf

In this first iteration of the ILM, we are essentially seeing the language that is produced from iteratively inducing 50 random inventions. There is no clear regularity in the system and the agent even at the end of life has no way to produce a couple of meanings (a_2, b_0) and (a_2, b_2). We can term this type of system a *protolanguage* system (in fact, it matches well with Wray's definition of protolanguage [14] rather better than the more well-known Bickertonian definition [15][5]).

Once the language has converged, however, the system looks quite different. Here is the system after 30 generations (it looks the same after another 1000).

[4]The probability of a particular meaning being in the training input to a learner is $1 - (1 - 1/n)^r$, where n is the size of the meaning space and r is the number of utterances. So, the chances of every meaning being represented are $(1 - (1 - 1/n)^r)^n$. So, only three times in every 100 generations should a learner hear all meanings in the simulations reported here.

[5]Wray argues that protolanguage utterances were more likely to be *holistic* unanalyzed expressions as opposed to the short concatenations of words (rather like asyntactic minisentences) that Bickerton envisages.

	a_0	a_1	a_2	a_3	a_4
b_0	wcpalsdqu	asdqu	hnqmxsdqu	gpmhmsdqu	bsdqu
b_1	wcpalp	ap	hnqmxp	gpmhmp	bp
b_2	wcpaiihm	aihm	hnqmxihm	gpmhmihm	bihm
b_3	rkhxpwcpalmx	rkhxpamx	rkhxphnqmxmx	rkhxpgpmhmmx	rkhxpbmx
b_4	cswcpalbf	csabf	cshnqmxbf	csgpmhmbf	csbbf

This language is quite clearly compositional in a completely regular way. There appears to be substrings for each component of the meaning. For example, for the most part, a_0 is expressed using the string wcpal at the start of the utterance, whereas b_0 is expressed using the string sdqu at the end of the utterance. The only exceptions to this pattern are the meanings involving b_3 and b_4. These appear to have *circumfixes* like rkhxp- -mx and cs- -bf.

The grammar for this language is shown below.

$$
\begin{aligned}
S\text{: }(x, y) &\rightarrow A\text{: }xB\text{: }y\\
S\text{: }(x, b_3) &\rightarrow \text{rkhxp}A\text{: }x\text{mx}\\
S\text{: }(x, b_4) &\rightarrow \text{cs}A\text{: }x\text{bf}\\
A\text{: }a_0 &\rightarrow \text{wcpal}\\
A\text{: }a_1 &\rightarrow \text{a}\\
A\text{: }a_2 &\rightarrow \text{hnqmx}\\
A\text{: }a_3 &\rightarrow \text{gpmhm}\\
A\text{: }a_4 &\rightarrow \text{b}\\
B\text{: }b_0 &\rightarrow \text{sdqu}\\
B\text{: }b_1 &\rightarrow \text{p}\\
B\text{: }b_2 &\rightarrow \text{ihm}.
\end{aligned}
$$

This compositionality is just what was expected to emerge given earlier results with more complex meaning spaces. In every run of the simulation, this sort of behavior always emerges. The exact structure of the language is different every time, but structure preservation in the mapping between meanings and strings appears to be inevitable. We will return to an explanation for this in a later section.

B. Problems with Regularity

Although the primary target for explanation in this line of work has been the pervasive structure-preserving nature of human linguistic behavior, the results outlined above and in previous work by myself and others do not appear to capture the nature of regularity in human languages accurately. This is because a completely stable compositional system is actually unusual in real languages, whereas there does not seem to be any other possible attractor in the ILM dynamic. If we are to claim that structure preservation in human languages arises out of the dynamics of the transmission of learned behavior, it seems important that we should be able to show partial but stable *irregularity* emerging as well.

C. Pressures on Language Use

One possible reason that the models presented so far appear to be *too* perfectly structure-preserving is that there is no pressure on the language that is being transmitted other than a pressure to be learnable. Real languages have to be used by speakers who are not always accurate and whose performance may be influenced by least-effort principles. To test the hypothesis that features of linguistic *performance* influence the emergence of regularity, the experiment described earlier was repeated with two modifications.

1) If the agent speaking has more than one way of expressing a meaning, then instead of picking one at random, the shortest string is always used.
2) With a certain probability per character, the speaker may fail to pronounce characters in a string. In other words, there is a chance of random noiselike erosion of the strings transmitted from generation to generation. For the simulation results here, the erosion probability per character was 0.001.[6]

In many ways, the behavior of the simulation under these conditions is similar to the previous simulation. The initial language is completely noncompositional (irregular), but regularity increases gradually over time. The major differences are that the system obviously never reaches complete stability, since there is always a chance that a random erosion will disrupt linguistic transmission. Fig. 2 shows the evolution of the system over 1000 generations.

Again, it is easy to see language evolution in progress in the model by tabulating strings from adult agents in the simulation at different points in time. Here is a language from early in the simulation (generation 13).

	a_0	a_1	a_2	a_3	a_4
b_0	qbkgofetfv	mpqr	kyfnfz	knj	wgvick
b_1	wkjwjk	usdptfzoq	lrl	lrifj	lz
b_2	xwlua	tvnakitga	fjpginnza	fja	kbtlakgyoa
b_3	lyxzd	qesgsqgyfoq	liuc	ifjiuc	lhsmy
b_4	pplj	lvtjoq	ubvqsj	yj	f

As before, this early system is clearly a protolanguage as opposed to a fully compositional mapping from meanings to strings. Again, however, a structure-preserving system does eventually emerge (generation 223).

	a_0	a_1	a_2	a_3	a_4
b_0	qda	bguda	lda	kda	ixcda
b_1	qr	bgur	lr	kr	ixcr
b_2	qa	bgua	la	ka	ixca
b_3	qu	bguu	lu	ku	ixcu
b_4	qp	bgup	lp	kp	ixcp

[6]If the erosion resulted in a string of length zero, the speaking agent was considered not to have said anything. In this case, no information was passed to the learning agent.

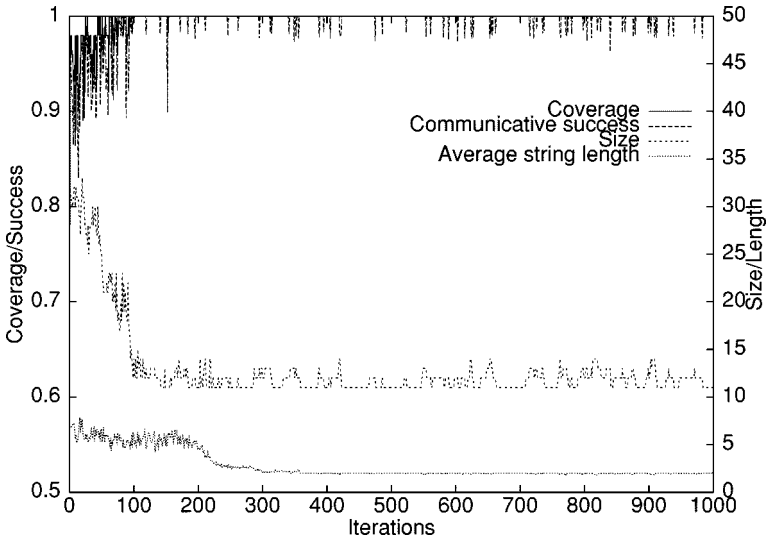

Fig. 2. Evolution of the language in an ILM with pressures for short strings. In addition to coverage, communicative success and size of grammar mean string length of utterances is also plotted.

Just as in the previous simulation, a clearly regular encoding has evolved. This is not what we were looking for, however. There is still no irregularity in this language. In later generations, some irregularity does emerge (shown in bold in this example from generation 763).

	a_0	a_1	a_2	a_3	a_4
b_0	qd	gd	ld	kd	xd
b_1	qr	gr	lr	**k**	xr
b_2	qa	ga	la	ka	xa
b_3	qu	gu	**u**	ku	xu
b_4	qp	gp	lp	kp	xp

These irregulars are not stable, however. They typically only last one or two generations, being rapidly reregularized by learners. So, although the performance pressures clearly influence the evolution of the system (in that the string length is much reduced), realistic irregulars have not emerged.

D. Nonuniform Frequency Distributions

What else might be needed to model both the emergence of structure-preserving regularity and stable irregularity in languages? A clear indication of what is missing from the ILMs presented so far is given if we look at where in real languages irregulars appear most stable. For example, here are some of the verbs in English that have an irregular past tense: *be, have, do, say, make, go, take, come, see, get, . . .*.

Strikingly, these verbs are also the ten most frequent verbs in English usage [16]. In fact, it is recognized by linguists that irregularity (i.e., noncompositionality) correlates closely with frequency in natural language [17]. The frequency with which meanings need to be expressed in the ILM (and, hence, indirectly the frequency of use of particular strings) is uniform. In contrast, the frequency of use of words in natural languages approximates a Zipfian distribution [18]; that is, the frequency of use of a particular word is inversely proportional to its frequency rank. While we cannot infer the frequency distribution of particular *meanings* in real languages from this directly, it strongly suggests that a uniform distribution is an unrealistic idealization.

Consequently, the simulation in the previous section is rerun with a nonuniform distribution over meanings (shown in Fig. 3) based on a Zipfian surface. This means that when, in the ILM, meanings are chosen at random for the adult agent to produce strings, the probability of picking a particular meaning is weighted so that the frequency of use of meanings approximates the function shown in Fig. 3.

The results of a run with this distribution of meanings is shown in Fig. 4. It is noticeable that the system appears far less stable than others, suggesting that the process of language change is ongoing throughout the simulation (as it is in real language history). The most important result, however, can be seen by looking at a snapshot of the language taken at one point in time in the simulation (generation 256).

	a_0	a_1	a_2	a_3	a_4
b_0	**g**	**s**	kf	jf	uhlf
b_1	**y**	jgi	ki	ji	uhli
b_2	yq	jgq	kq	jq	uhlq
b_3	ybq	jgbq	kbq	jbq	uhlbq
b_4	yuqeg	jguqeg	kuqeg	juqeg	uhluqeg

As before, there are some irregular forms (shown in bold), but in contrast with the previous result, they are highly stable. For example, this particular cluster of irregulars appeared in generation 127 and lasts until generation 464, at which point y is regularized to yi. Indeed, the irregular g appears constant throughout all 1000 generations of the simulation. Furthermore, just as in the real case, the irregular forms are all highly frequent. It is also interesting to note that length appears to correlate inversely with

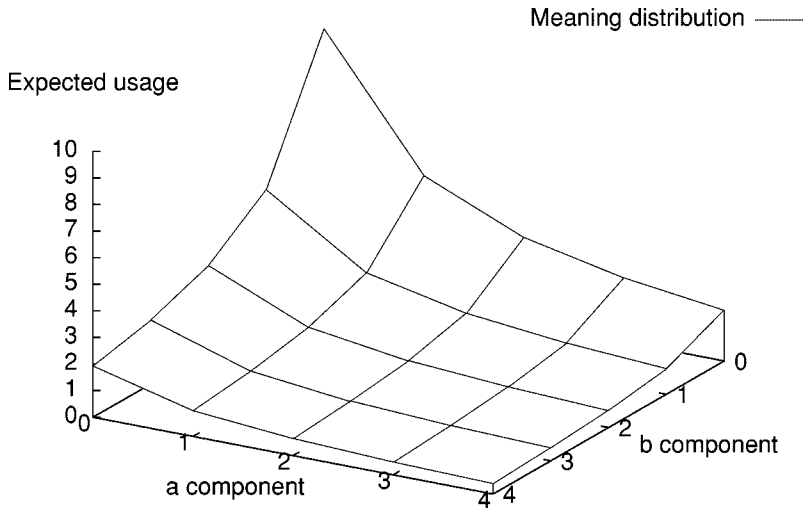

Fig. 3. Expected number of each meaning in the input to a learner. Probability of a meaning (a_i, b_j) is proportional to $(i+1)^{-1}(j+1)^{-1}$ and, as before, the total number of utterances is 50.

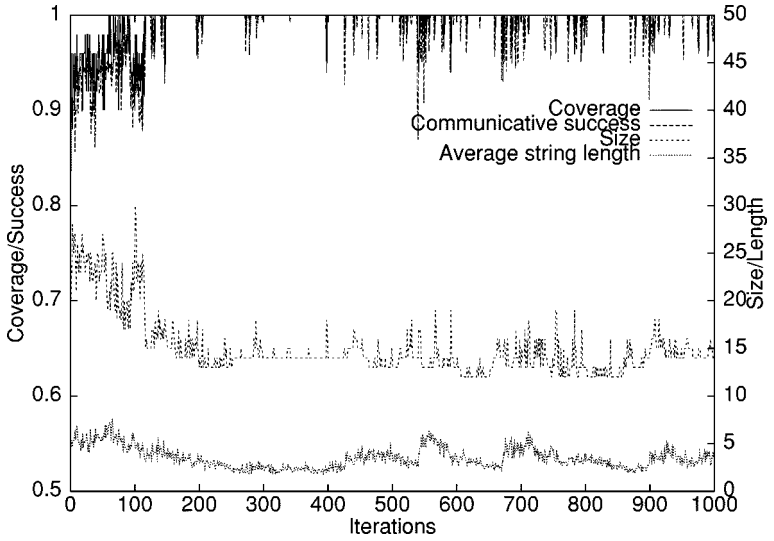

Fig. 4. Evolution of language with performance pressures and a nonuniform distribution of meanings.

frequency (although quantitative results have yet to be obtained). This correlation is also well known in human language [18].

IV. Discussion

Why does language emerge in these simulations? There is no biological evolution in the ILM—the agent architecture is the same throughout the simulation. Nowhere in the simulation is the communicative system of the agents measured against an objective function. To put it crudely, the agents do not care if they can communicate or not. Nor does the agent architecture put any hard constraints on whether the language that they can use should be structure-preserving or not (as witnessed by the languages in the early stages of the simulation). Given these facts, one might suspect that there would be no evolution in the ILM at all. Counter to intuition, the system appears to adapt. In even the simplest instantiation of the model, structure emerges in the meaning-signal mapping. Words/morphemes spontaneously emerge that correspond to subparts of the meaning and regular rules evolve for combining these into complete sentences. Where there are length pressures placed on the signal channel, the language adapts to shorter codes. When the distribution of meanings is not smooth, a system evolves with a realistic pattern of frequent short irregulars and infrequent regular forms.

The key to understanding the behavior in the model lies in seeing the language (as opposed to the language users) as adapting to improve its own survival. In a standard evolutionary simulation, a model of natural selection would lead the agents to adapt to some fitness function. However, in this case there is no natural selection; agents do not adapt, but rather we can see the process of transmission in the ILM as imposing a cultural linguistic selection on features of the language that the agents use.

The survival of a language, or rather a feature of a language like a rule or word, over time relies on it being repeatedly replicated in the ILM. Fig. 5 shows how linguistic information flows through time, both in the model and in reality. It is clear from this that in order for a linguistic feature of any kind to be successful, it must survive the two transformations between its internal representation (I-language in the Chomskian parlance [19]) and its external representation (E-language or utterances). We can,

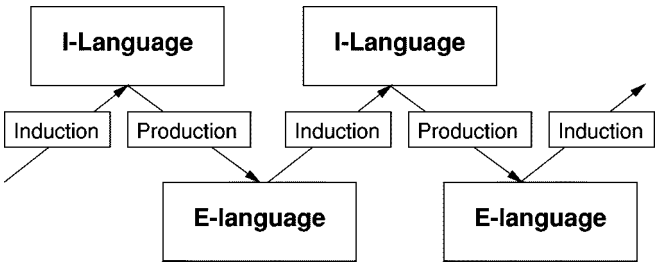

Fig. 5. Process of linguistic transmission. I-language refers to the internal representation (e.g., grammar) of language, whereas E-language is the external form of language as sets of utterances. For language to persist, it must be transformed repeatedly from one domain to the other through the processes of induction and production.

therefore, see the processes of language induction and language production as imposing endogenous selection pressures on languages. To put it another way, these transformations act as *bottlenecks* on the persistence of linguistic variation (see also [8] and [11]).

Taking these in turn, it is clear that the optimal system with regard to these two bottlenecks is rather different.

Induction: The induction or learning bottleneck appears to favor languages that are maximally structure-preserving. More generally, it has been argued that the learning bottleneck favors linguistic generalizations [11] or compressed internal representations [20]. If we consider generalizations as replicators, it is clear why this might be the case. For example, assume in a hypothetical language that there are two ways of expressing a particular meaning. The first is noncompositional: it is simply a completely idiosyncratic holistic word for that whole meaning. The second, on the other hand, is compositional, produced by some more general rule that can also produce a variety of other meanings. What are the chances that each of these ways of producing that meaning will survive? For the first way of expressing the meaning to survive, that meaning-string pair must be produced by the adult and heard by the learner. However, for the second, it does not necessarily need to be produced. In some sense, the actual meaning-string pair is not the relevant replicator; rather, it is the generalization that is replicating. By definition, a generalization will cover more meaning space and, therefore, have more chance of being expressed in the input to the learner. The learning bottleneck, therefore, will tend to produce a selective pressure for generalizations. Notice that the precise form of generalization that is possible must depend on the prior bias of the learner—in the case of this model, for example, this corresponds to the induction heuristics and the choice of representation language (i.e., DCGs). That said, however, the argument about the relative replicability of generalizations is quite general. It is an ongoing research project to discover how similar emergent languages are with a range of representational and search biases (see, e.g., [13]).

Production: The pressure imposed by the production bottleneck in this simulation is very clear. Shorter strings are more likely to show up in the languages in the simulation. Ultimately, we might expect that if it were possible for pressures from production alone to influence the simulation, languages would tend toward minimal length codes for the meaning space.

The language for a 5×5 meaning space and a 26-letter alphabet cannot both be minimal length and structure-preserving because the shortest language would have one letter for each whole meaning, which makes compositionality impossible. The two selection pressures on language are, therefore, in competition.[7] What is interesting about this competition is that the relative pressure varies according to the frequency of use of the meanings. The induction pressure becomes more severe for low-frequency meanings since these will have less likelihood of being expressed in the learner's training set. The low frequency forms, therefore, need to behave in regular paradigms. Conversely, the higher frequency forms are much more likely to replicate without the need to be part of a general pattern. This means that they can succumb to the pressure for shortness and, hence, irregularity.

Various directions for future work are possible given the results described in this paper, for example.

1) The role of induction bias: This has already been mentioned; essentially we need to replicate these results with a range of different learning algorithms to uncover the (nontrivial) relationship between bias and emergent structure of systems in the ILM.
2) The role of invention: I have argued for an invention algorithm that never increases the degree of compositionality inherent in the learner's grammar at the time of invention. This is clearly only one point on a scale of possible approaches, however. For example, different dynamics would be observed given a purely random invention process, or one which maximizes compositionality. An interesting project would be to look at the types of language-change that are observed with different algorithms and compare these with real language change.
3) Different degrees of irregularity: In some languages (such as isolating languages like Chinese), there is little or no irregularity. A challenge for modeling could be to isolate the possible mechanisms that might lead to different degrees of irregularity in different languages. This could involve investigating more closely the processes of, for example, phonological erosion, which here is modeled very crudely (as the random elision of segments).
4) The comprehension bottleneck: The results described here are due to an interaction of induction and production bottlenecks. Other work looks at the role of comprehension bottlenecks on emergent structure (e.g., [21]). A possible extension to the model might be to integrate this third bottleneck into the ILM.

V. Conclusion

A central and unique feature of human language—structure preservation—has been explained in terms of the evolution of languages themselves as opposed to language users. Within the framework of an ILM, it has been shown that general pressures on the transmission of language over time give rise not only to compositional systems of meaning-signal mapping, but also

[7]Appealing to pressures in competition is a well recognized explanatory principle in linguistic functionalism, where the term "competing motivations" is used. See [21] and [22] for discussion.

realistic patterns of language dynamics, utterance length, and irregularity.

This research suggests that if we are to understand the origins of human linguistic behavior, we may need to concentrate less on the way in which we as a species have adapted to the task of using language and more on the ways in which languages adapt to being better passed on by us.

ACKNOWLEDGMENT

The author would like to thank J. Hurford, members of the Language, Evolution, and Computation Research Unit, the three anonymous reviewers, and the audiences of the Evolutionary Computation in Cognitive Science 2000 workshop in Melbourne and the Paris Evolution of Language Conference 2000 for their constructive comments.

REFERENCES

[1] M. Oliphant, "The learning barrier: Moving from innate to learned systems of communication," *Adaptive Behav.*, to be published.
[2] R. Shillcock, S. McDonald, C. Brew, and S. Kirby, Human languages evolve to fit a structure-preserving mental lexicon, unpublished manuscript.
[3] S. Pinker and P. Bloom, "Natural language and natural selection," *Behav. Brain Sci.*, vol. 13, no. 4, pp. 707–784, 1990.
[4] M. A. Nowak, J. B. Plotkin, and V. A. A. Jansen, "The evolution of syntactic communication," *Nature*, vol. 404, no. 6777, pp. 495–498, Mar. 30, 2000.
[5] J. Batali, "Computational simulations of the emergence of grammar," in *Approaches to the Evolution of Language*, J. Hurford, M. Studdert-Kennedy, and C. Knight, Eds. Cambridge, U.K.: Cambridge Univ. Press, 1998.
[6] J. Batali, "The negotiation and acquisition of recursive grammars as a result of competition among exemplars," in *Linguistic Evolution through Language Acquisition: Formal and Computational Models*, E. Briscoe, Ed. Cambridge, U.K.: Cambridge Univ. Press.
[7] S. Kirby, "Learning, bottlenecks and the evolution of recursive syntax," in Linguistic Evolution Through Language Acquistion: Formal and Computational Models, E. Briscoe, Ed. Cambridge, U.K.: Cambridge Univ. Press, to be published.
[8] S. Kirby, "Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners," in *The Evolutionary Emergence of Language*, C. Knight, M. Studdert-Kennedy, and J. R. Hurford, Eds. Cambridge, U.K.: Cambridge Univ. Press, to be published.
[9] ——, "Syntax out of learning: The cultural evolution of structured communication in a population of induction algorithms," in *Advances in Artificial Life*. ser. Lecture Notes in Computer Science, D. Floreano, J. D. Nicoud, and F. Mondada, Eds. New York: Springer-Verlag, 1999.
[10] ——, "Learning, bottlenecks and infinity: A working model of the evolution of syntactic communication," in *Proceedings of the AISB'99 Symposium on Imitation in Animals and Artifacts*, K. Dautenhahn and C. Nehaniv, Eds: Society for the Study of Artificial Intelligence and the Simulation of Behavior, 1999.
[11] J. R. Hurford, "Expression/induction models of language evolution: Dimensions and issues," in *Linguistic Evolution Through Language Acquistion: Formal and Computational Models*, E. Briscoe, Ed. Cambridge, U.K.: Cambridge Univ. Press, to be published.
[12] ——, "Social transmission favors linguistic generalization," in *The Emergence of Language: Social Function and the Origins of Linguistic Form*, C. Knight, M. Studdert-Kennedy, and J. Hurford, Eds. Cambridge: Cambridge Univ. Press, pp. 324–352, to be published.
[13] B. Tonkes and J. Wiles, "Methodological issues in simulating the emergence of language," in *Proc. 3rd Conf. Evolution Language*, Paris, France, Apr. 2000, submitted for publication.
[14] A. Wray, "Protolanguage as a holistic system for social interaction," *Lang. Commun.*, vol. 18, no. 1, pp. 47–67, 1998.
[15] D. Bickerton, *Language and Species*. Chicago, IL: Univ. of Chicago Press, 1990.
[16] N. Francis and H. Kucera, *Frequency Analysis of English Usage: Lexicon and Grammar*. Boston, MA: Hoghton Mifflin, 1982.
[17] S. Pinker, *Words and Rules*. London, U.K.: Weidenfeld & Nicolson, 1999.
[18] G. K. Zipf, *The Psycho-Biology of Language*. London, U.K.: Routledge, 1936.
[19] N. Chomsky, *Knowledge of Language*. New York: Praeger, 1986.
[20] T. Teal and C. Taylor, "Compression and adaptation," in *Advances in Artificial Life*. ser. Lecture Notes in Computer Science, D. Floreano, J. D. Nicoud, and F. Mondada, Eds. New York: Springer-Verlag, 1999.
[21] S. Kirby, *Function, Selection and Innateness: The Emergence of Language Universals*. London, U.K.: Oxford Univ. Press, 1999.
[22] F. J. Newmeyer, *Language Form and Language Function*. Cambridge, MA: MIT Press, 1998.

Simon Kirby received the M.A. degree in artificial intelligence and linguistics and the Ph.D. degree in linguistics from the University of Edinburgh, Edinburgh, U.K., in 1992 and 1996, respectively.

He is currently a British Academy Research Fellow in the Language Evolution and Computation Research Unit at the Department of Theoretical and Applied Linguistics, University of Edinburgh, Ediburgh, U.K. He was also a Visiting Research Fellow at the Collegium Budapest Institute for Advanced Study, Budapest, Hungary. His research interests include the use of computational modeling to understand the origins of human language and the impact of evolution on linguistic structure.

Dr. Kirby received the Howe Prize for Artificial Intelligence in 1992.