The Emergence of Linguistic Structure: An Overview of the Iterated Learning Model

Simon Kirby & James R. Hurford
Language Evolution and Computation Research Unit
Department of Linguistics
University of Edinburgh
http://www.ling.ed.ac.uk/lec

March 21, 2001

> "The most basic principle guiding [language] design is not communicative utility but reproduction — theirs and ours ... Languages are social and cultural entities that have evolved with respect to the forces of selection imposed by human users. The structure of a language is under intense selection because in its reproduction from generation to generation, it must pass through a narrow bottleneck: children's minds." (Deacon, 1997, 110)

1 Introduction

As language users humans possess a culturally transmitted system of unparalleled complexity in the natural world. Linguistics has revealed over the past 40 years the degree to which the syntactic structure of language in particular is strikingly complex. Furthermore, as Pinker and Bloom point out in their agenda-setting paper *Natural Language and Natural Selection* "grammar is a complex mechanism tailored to the transmission of propositional structures through a serial interface" (Pinker and Bloom, 1990, 707). These sorts of observations, along with influential arguments from linguistics and psychology about the innateness of language (see, e.g. Chomsky, 1986; Pinker, 1994), have led many authors to the conclusion that an explanation for the origin of syntax must invoke neo-Darwinian natural selection.

> "Evolutionary theory offers clear criteria for when a trait should be attributed to natural selection: complex design for some function, and the absence of alternative processes capable of explaining such complexity. Human language meets these criteria." (Pinker and Bloom, 1990, 707)

Since Pinker and Bloom made these arguments there have been many attempts to put forward a coherent evolutionary story that would allow us to derive known features of syntax from communicative selection pressures (e.g. Nowak, Plotkin, and Jansen, 2000; Newmeyer, 1991 and discussion in Kirby, 1999a). One problem with this approach to evolutionary linguistics is that it often fails to take into account that biological natural selection is only one of the complex adaptive systems at work.

Language emerges at the intersection of three complex adaptive systems:

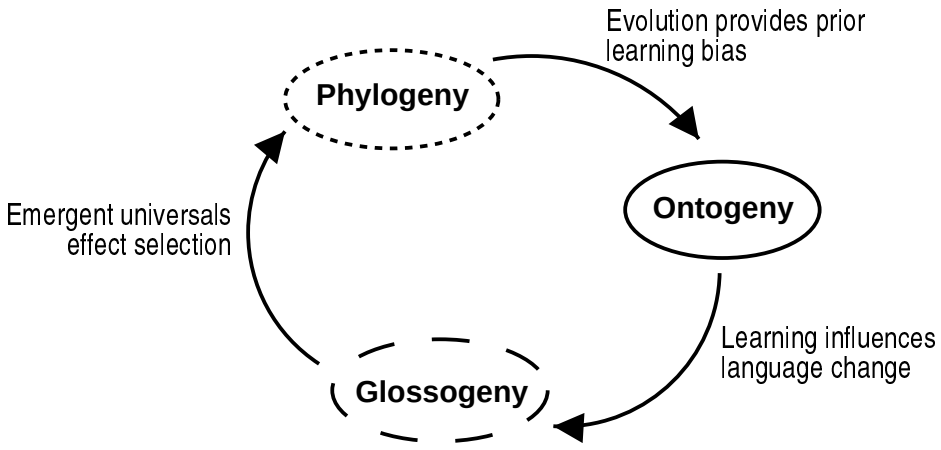

Figure 1: The three adaptive systems that give rise to language. Some of the interactions between these systems are shown.

Learning During *ontogeny* children adapt their knowledge of language in response to the environment in such a way that they optimise their ability to comprehend others and to produce comprehensible utterances.

Cultural evolution On a historical (or *glossogenetic*) timescale, languages change. Words enter and leave the language, meanings shift, and phonological and syntactic rules adjust.

Biological evolution The learning (and processing) mechanisms with which our species has been equipped for language, adapt in response to selection pressures from the environment, for survival and reproduction.

There are two problems with this multiplicity of dynamical systems involved in linguistic evolution. Firstly, we understand very little about how learning, culture, and evolution interact (though, see Belew, 1990; Kirby and Hurford, 1997; Boyd and Richerson, 1985), partly because language is arguably the only sophisticated example of such a phenomenon. There clearly *are* interactions: for example, biological evolution provides the platform on which learning takes place, what can be learnt influences the languages that can persist through cultural evolution, and the structure of the language of a community will influence the selection pressures on the evolving language users (see figure 1).

Secondly, it is not clear what methodology we should use to study this problem. Mathematical techniques for looking at the interaction of dynamical systems and linguistic behaviour are in their infancy (though, Nowak, Komarova, and Niyogi, 2001, take a potentially valuable approach). We feel that computational modelling is currently the most appropriate methodology, but although simulations of language learning have a long history, and there are many methods from the A-life field that can be used for modelling evolution, models of the cultural transmission of learned behaviour are relatively sparse (see Steels, 1997 for a review). This is unfortunate, since we will argue in this chapter that it is through this particular mechanism that the most basic features of human language syntactic structure can be explained.

To remedy this situation, we introduce here the Iterated Learning Model (ILM), a general approach to exploring the transmission over a glossogenetic timescale of observationally learned behaviour. We will illustrate the ILM with a few examples of simulations that lead to two conclusions:

- There is a non-trivial mapping between the set of learnable languages (i.e. the languages allowed by our innate language faculty), and the set of stable languages (i.e., the languages we can actually expect to see in the world).
- Under certain circumstances, cultural evolution leads inevitably to recursively compositional (i.e., syntactic) languages.

2 The Iterated Learning Model

The central idea behind the ILM is to model directly the way language exists and persists via two forms of representation (Chomsky, 1986):

I-language This is language as it exists internally, as patterns of neural connectivity, or more abstractly, as a grammar.

E-language This is the external form of language as actual sets of utterances.

For a language, or a pattern within a language, to persist from one generation of language users to the next it must be mapped from I-language to E-language (through use) and from E-language back to I-language again (through learning). This is the bottleneck on transmission that Deacon mentions in the quotation that starts this chapter.

In this chapter (and in previous work) we are interested in a particular property of language that must persist in this way: the structure of the mapping from meanings to signals (and *vice versa*). In order to model this, the ILM needs four basic components:

1. A meaning space.
2. A signal space.
3. One or more language-learning agents.
4. One or more language-using adult agents.

Each iteration of the ILM involves an adult agent being given a set of randomly chosen meanings to produce signals for. The resulting meaning-signal pairs form training data (E-language) for one or more learning agents. After a learning period (the *critical period* to use terminology from language acquisition), the learning agents form their individual I-languages and thus become adults. New learners are added to replace the learners, and adults are removed in order to maintain population size.[1] This cycle is typically repeated several thousand times or until a stable point attractor in the dynamical system is reached. Importantly, a normal ILM simulation will be initialised with no linguistic system in place whatsoever. To put it another way, the initial adults have no I-language and at the start the community of agents have no E-language.

[1] Note that this is a highly simplified population dynamic. See Briscoe (2000) for an alternative approach to population replacement. More sophisticated approaches to population dynamics are likely to be a prerequisite of a model of creolisation, for example.

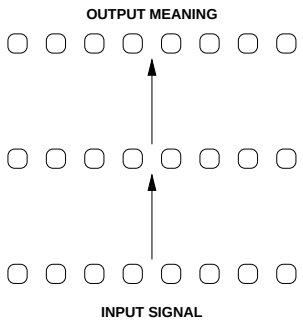

Figure 2: A network that maps from 8-bit binary signals to 8-bit binary meanings. Arrows indicate full connectivity.

2.1 A simple ILM

To exemplify the iterated learning model, here we describe a simple simulation using neural networks as learners. Although the setup is very simple, there are still interesting properties that spontaneously emerge in the languages in the system, which will point the way to the simulations which we will describe in later sections.

In line with much of the work that we have done with the ILM, we will trade off complex population dynamics against speed for the simulation. Accordingly, the population at any one time consists of a single adult and a single learner. The agents are feed-forward networks, with an $8 \times 8 \times 8$ structure. The networks map from an 8-bit binary number representing a signal to an 8-bit binary number representing a meaning, in other words, given appropriate input, the agents can learn to *parse* utterances (figure 2).

Because the networks are unidirectional, we need some *production* mechanism, mapping back from meanings to signals. One way to do this is to use a version of the *obverter* learning strategy discussed in Oliphant and Batali (1997), Smith (in press), (Smith, 2001b), (Smith, 2001a). The idea behind obverter is for a communicative agent to produce signals that maximise the chance of the hearer understanding the correct meaning. Given that a speaker will not have access to the signal-to-meaning mapping of the hearer, obverter makes the simplifying assumption that the speaker's own mapping approximates that of the hearer. In practical terms this means that, in order to decide which signal to produce for a meaning, we need to search for the signal that would result in the desired meaning if parsed by the speaker.

We want to find a signal s given a meaning m.

$$s_{\text{desired}} = \underset{s}{\operatorname{argmax}} P(s|m) \tag{1}$$

$$= \underset{s}{\operatorname{argmax}} \frac{P(m|s)P(s)}{P(m)} \tag{2}$$

$$= \underset{s}{\operatorname{argmax}} P(m|s) \tag{3}$$

$$= \underset{s}{\operatorname{argmax}} C(m|s) \tag{4}$$

Where $C(m|s)$ is the confidence that the network has the mapping $s \rightarrow m$. In other words,

find the signal that maximises the network's confidence in the given meaning.[2]

In order to calculate $C(m|s)$, we treat the real-numbered network outputs $o[1\ldots8]$ as a measure of confidence in the the meaning bits $m[1\ldots8]$.

$$C(m[1\ldots8]|o[1\ldots8]) = \prod_{i=1}^{8} C(m[i]|o[i]) \tag{5}$$

$$C(m[i]|o[i]) = \begin{cases} o[i] & \text{if } m[i] = 1, \\ (1 - o[i]) & \text{if } m[i] = 0. \end{cases} \tag{6}$$

In summary, the ILM in this case proceeds as follows:

1. An initial population is setup consisting of two randomly initialised networks, a speaker and a hearer.
2. A certain number of random meanings are chosen from the set of binary numbers 00000000 to 11111111, with replacement.
3. The speaker produces signals for each of these meanings by applying the obverter procedure.
4. This set of signal-meaning pairs is used to train the hearer network using the backpropagation of error learning algorithm.[3]
5. The speaker is removed, the hearer is designated a speaker, and a new hearer is added (with randomly initialised weights).
6. The cycle repeats.

What happens in such a model? It turns out that there are three types of behaviour. Which behaviour emerges depends entirely on the number of utterances the hearer learns from. With a very small training set, the language evolves as shown in figure 3. This graph shows the expressivity of the language (i.e. how much of the meaning space is covered by the 256 signals), and its stability (i.e. how different the hearer's language is from the speaker's after training). The graph shows the results with 20 randomly chosen meanings. The emergent language is inexpressive and unstable.

For a very large training set the behaviour is rather different. Figure 4 shows a longer run with 2000 randomly chosen meanings each generation. Eventually, a completely stable and completely expressive language is found. For a medium-sized training set, apparently similar behaviour emerges, albeit faster. Figure 5 shows results for 50 randomly chosen meanings.

There is more to these results than intially meets the eye, however. The runs with a medium-sized training set have a feature that does not emerge in runs with either the very

[2]This is potentially a very computationally costly operation. However, the efficiency of obverter need not be a big issue if careful *memoization* is used in the network computation. As long as the training phase and the production phase of the agent are strictly separated, then network weights will be fixed throughout production. This means that, as results are computed, they can be stored in a lookup table (or cache). Notice that this kind of optimisation is possible only because the ILM is generational. It would not work for the types of population model that Batali (2001), Batali (1998) employs, for example.

[3]The learning algorithm used has a learning rate of 0.1 and no momentum term. Each learner is presented with 100 randomised epochs of the data set.

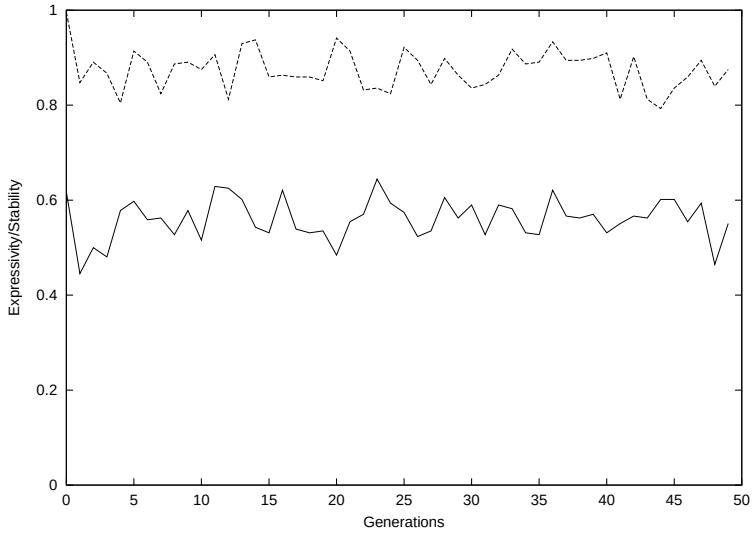

Figure 3: A simulation run where only 20 random meanings are produced by each generation's speaker. The dotted line corresponds to the size of the difference between the leaners and the adults language after training. The solid line represents the proportion of the meaning space covered by the language. In this case the language is inexpressive and unstable.

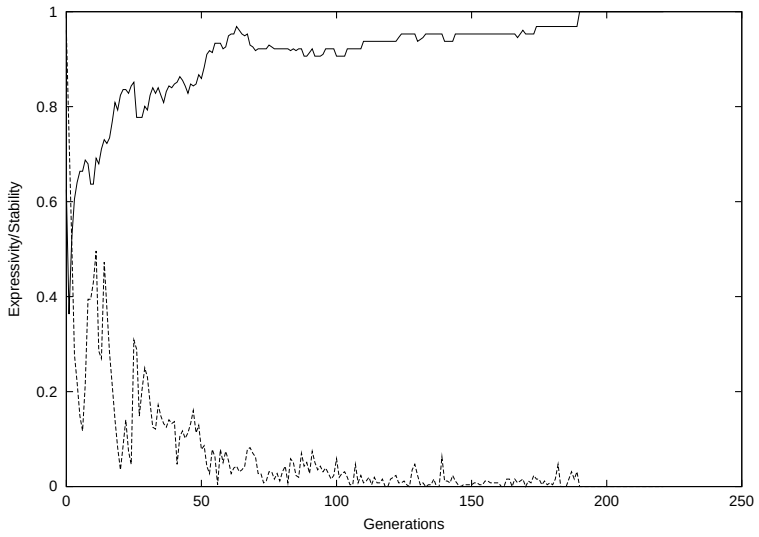

Figure 4: A simulation run where 2000 random meanings are produced by each generation's speaker. Here an expressive and stable system is eventually reached.

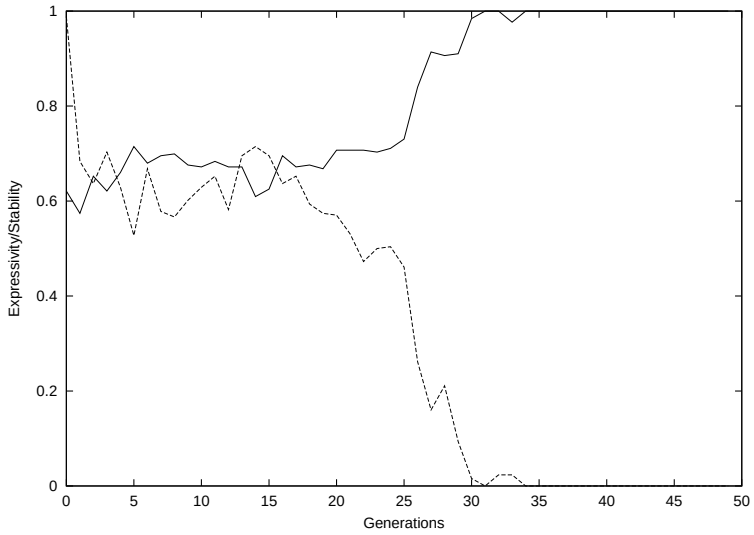

Figure 5: A simulation run where 50 random meanings are produced by each generation's speaker. In this run, an expressive stable system is reached relatively quickly.

large or very small numbers of meanings in training. Whereas the language in the run with the large training set is completely expressive, the mapping from meanings to signals is essentially random. That is, there is no structure in the pairings.[4]

Surprisingly, this random pairing never arises for the runs with a medium-sized training set. In the run with 50 random meanings, the emergent meaning-signal mappings are highly structured. In fact, for the run shown in figure 5, they can be captured by a very simple set of translation rules:

$$\begin{aligned} m_1 &\leftrightarrow \neg s_3 \\ m_2 &\leftrightarrow \neg s_5 \\ m_3 &\leftrightarrow \neg s_6 \\ m_4 &\leftrightarrow \neg s_1 \\ m_5 &\leftrightarrow s_4 \\ m_6 &\leftrightarrow s_8 \\ m_7 &\leftrightarrow \neg s_2 \\ m_8 &\leftrightarrow s_7 \end{aligned}$$

where m_n is the nth bit of the meaning, s_n is the nth bit of the signal, and $\neg s_n$ is the logical negative of the nth signal bit. For example, the meaning 00110100 is expressed as 01101001.

It turns out that this kind of result *always* emerges as long as the number of training examples is not too small and not too large (the particular numbers seem to depend on the structure of the networks, and the size of the meaning and signal spaces – see Brighton and

[4]This sort of system has been described as a protolonguage system by Wray (1998). She uses this term rather differently from Bickerton (1990). Whereas Bickertonian protolanguage has some structure in the mapping from signals to meanings, in that it has words that are concatenated in a non-structured way to form proto-sentences, Wray argues that a better model of protolanguage of which there are current "living fossils" is one in which the meaning-signal mapping is completely holistic. The sentences in a protolanguage according to Wray cannot be broken down in any way which will give a clue to what the meaning of the sentence is.

Kirby (2001a, Brighton and Kirby (2001b) for an approach to quantifying the critical values for the ILM).

3 Recursive compositionality out of iterated learning

Our second example of the ILM is covered in more detail elsewhere (Kirby, 2001; Kirby, 1999b; Kirby, 1999c), so we will only describe it briefly here.[5]

As we mentioned earlier, perhaps the fundamental feature of human language that sets it apart from other animals' communication systems is the unique way in which the structure of a signal can be decomposed into separate meaning-bearing parts. Linguists refer to this as *compositionality*.

Compositionality A compositional signalling system is one in which the meaning of a signal is some function of the meaning of the parts of that signal and the way in which they are put together.

Furthermore, the mapping from sentences to meanings in language is not only compositional, but recursive. Parts of sentences which can be ascribed a meaning can themselves be decomposed by the same compositional function. Ultimately, this leads to what has been called the "digital infinity" of language: as language users, we can make potentially infinite use of finite means by constructing meaning-bearing syntactic structures that themselves contain structures of the same type.

3.1 A symbolic approach

It should be clear that the ILM described in the previous section produced a language that was compositional by the definition above with the correct number of training examples. However, it is impossible for the linguistic system also to be recursive using the same modelling methodology (for a start, the signal and meaning spaces are finite). The simulations described in this section are an attempt to get round this limitation.

Rather than using bit-vectors to represent meanings and signals, here we use a simple variant of predicate logic for meanings, and strings of characters as signals. For example, two possible meanings and their corresponding signals might be:

$$\mathsf{loves(mary,john)} \leftrightarrow \texttt{marylovesjohn}$$
$$\mathsf{knows(gavin,loves(mary,john))} \leftrightarrow \texttt{gavinknowsmarylovesjohn}$$

In the simulations reported here, there are 5 possible predicates each of which take two arguments which may vary over 5 possible "people". There are also 5 predicates of propositional attitude which take one normal argument, and one recursive argument as in the example above.

In order to represent the mapping from meanings to strings we use a simple version of definite clause grammar (DCG). It is important here to note that this does not preclude non-compositional languages *a-priori*. A dictionary-like holistic protolanguage can equally well be expressed in DCG notation as a recursively compositional language. So, the examples above could be generated by the non-compositional grammar: (In this DCG representation, the material to the right of the slash indicates the meaning assigned to the syntactic category to the left of the slash.)

[5] The model described is an extension of an earlier model that was put forward in (Kirby, 2000).

S/loves(mary,john) → `marylovesjohn`
S/knows(gavin,loves(mary,john)) → `gavinknowsmarylovesjohn`

Obviously, in this case, the same strings could be generated by a compositional grammar.

In order to implement the ILM, we need an inductive learning algorithm for these grammars. An ideal method might involve a search for optimal grammars using some metric such as Minimum Description Length. Indeed, we are pursuing work along these lines (Brighton and Kirby, 2001b; Brighton and Kirby, 2001a). However, one problem with the ILM lies in the necessity for very efficient learning methods (since any learning problem is necessarily scaled up many thousand-fold when generations of learners are required). Accordingly, the work we have undertaken with DCGs relies upon a heuristic-driven incremental grammar induction method described in detail in Kirby (in press). Very briefly, this learning method initially *incorporates* rules into the grammar upon encountering each utterance, and then searches for possible generalisations over pairs of rules in the grammar that fit within a set of heuristic criteria which can replace those pairs of rules with a single one. In this way, the induction method ensures that the learner can always parse the data heard, but may also generalise to unseen examples if the generalisation is justified by the data. (See, Tonkes and Wiles (in prep) for some criticism of this approach.)

In addition to a learning mechanism, the ILM in this case also needs some way of innovating new signals. Given that the initial agents will have no I-language, they will have no way of generating strings for any of the meanings they are called upon to produce. (This is not true of the networks in the previous example because a network will always produce an output for any input. In other words, there is no equivalent of a failure to generate.) The technique for generating novel strings is also described elsewhere (Kirby, in press) — it is essentially random, but aims to be somewhat "smart" in that it avoids generating random strings wherever there is clear compositional structure in the grammar that the agent currently possesses.

3.2 From protolanguage to recursive syntax

In each iteration of the ILM, the speaker is required to produce strings for 50 simple meanings (i.e. with no embedding), 50 degree-1 meanings (with one level of embedding), and 50 degree-2 meanings. If invention is employed by the speaker, both the speaker and the hearer add the resultant meaning-string pair to their linguistic knowledge. Otherwise, only the hearer carries out the induction processes mentioned above.

In the initial stages of the simulation, a *protolanguage* emerges. The language that is culturally evolving has some words for some meanings, but no structure. Here are some example sentences from an early language (with the meanings given as English glosses):

(7) ldg
"Mary admires John"

(8) xkq
"Mary loves John"

(9) gj
"Mary admires Gavin"

(10) axk
"John admires Gavin"

(11) gb
"John knows that Mary knows that John admires Gavin"

We can see no obvious structure here. There appears to be no compositional encoding of the meanings. In fact, given the symbolic nature of the system, we can inspect the agents' I-language directly:

S/loves(john,mary) → sdx
S/admires(mary,gavin) → gj
S/admires(john,gavin) → axk
S/admires(gavin,heather) → nui
S/loves(john,heather) → my
S/loves(mary,john) → xkq
S/admires(mary,john) → ldg
S/thinks(john,loves(mary,gavin)) → fi
S/thinks(heather,loves(heather,gavin)) → ad
S/thinks(john,admires(heather,gavin)) → xuy
S/knows(gavin,loves(gavin,mary)) → k
S/knows(gavin,loves(john,mary)) → ysw
S/thinks(mary,knows(gavin,loves(heather,john))) → pq
S/thinks(mary,knows(heather,loves(heather,john))) → rr
S/knows(john,knows(mary,admires(mary,john))) → lr
... (plus another 101 rules)

Early on the simulation, the process of producing input for the next generation learner involves a lot of random invention. As a result, for many generations communication systems appear to stay random and idiosyncratic. Surprisingly, however, at some point in every simulation run, the language suddenly changes:[6]

(12) gj h f tej m
John Mary admires
"Mary admires John"

(13) gj h f tej wp
John Mary loves
"Mary loves John"

(14) gj qp f tej m
Gavin Mary admires
"Mary admires Gavin"

(15) gj qp f h m
Gavin John admires
"John admires Gavin"

(16) i h u i tej u gj qp f h m
John knows Mary knows Gavin John admires
"John knows that Mary knows that John admires Gavin"

In contrast to the previous example, there is obvious structure here. This is made even clearer by looking at the entire grammar of an agent in this simulation.

[6]The spaces are included here merely to aid comprehension. They are not available to the learner.

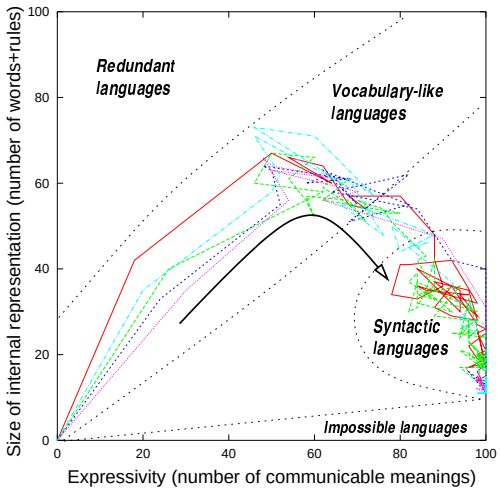

Figure 6: The results of several simulation runs plotted showing size and expressivity of the grammars. In order to plot this graph, meanings with embedding are ignored. In each run, the expressivity of the language increases to a maximum, but the size of the grammar initially grows during the protolanguage stage before reducing to a minimum, reflecting the emergence of syntax.

$$S/p(x,y) \rightarrow \texttt{gj}\ A/y\ \texttt{f}\ A/x\ \ B/p$$
$$S/p(x,q) \rightarrow \texttt{i}\ A/x\ \ D/p\ \ S/q$$
$$A/\mathsf{heather} \rightarrow \texttt{dl}$$
$$A/\mathsf{mary} \rightarrow \texttt{tej}$$
$$A/\mathsf{pete} \rightarrow \texttt{n}$$
$$A/\mathsf{gavin} \rightarrow \texttt{qp}$$
$$A/\mathsf{john} \rightarrow \texttt{h}$$
$$B/\mathsf{detests} \rightarrow \texttt{b}$$
$$B/\mathsf{loves} \rightarrow \texttt{wp}$$
$$B/\mathsf{hates} \rightarrow \texttt{c}$$
$$B/\mathsf{likes} \rightarrow \texttt{e}$$
$$B/\mathsf{admires} \rightarrow \texttt{m}$$
$$D/\mathsf{believes} \rightarrow \texttt{g}$$
$$D/\mathsf{knows} \rightarrow \texttt{u}$$
$$D/\mathsf{decides} \rightarrow \texttt{ipr}$$
$$D/\mathsf{says} \rightarrow \texttt{p}$$
$$D/\mathsf{thinks} \rightarrow \texttt{m}$$

Recursive compositionality has emerged in this simulation run after a few hundred generations, and is retained from that point onwards as a completely stable attractor in the dynamical system. Figure 6 gives a quantitative overview of the dynamics of the system.

4 Irregularity and frequency

Results like those described in the previous sections are encouraging in that they demonstrate the utility of the ILM as a mechanism for exploring how cultural transmission might explain some of the fundamental features of language that are the target of evolutionary linguistics.

Ultimately we would like to widen the remit of this sort of *glossogenetic* evolutionary explanation, in order to see how much can be explained without appealing directly to natural selection-type functional arguments. In particular, there are a couple of problems with the type of language that emerges in the simulations reported so far. Although compositional structure is clearly very common in language, it is not universal.

If we look at the morphological systems of many languages, for example, we find that there are often subparts of a paradigm that appear to flout the compositional structure obvious in the rest of the language. A classic example of this type of behaviour can be found by looking at the past tense of English verbs. Whilst the majority of verbs inflect quite regularly (by adding $-ed$), some have irregular past tenses (for example, the past tense of verbs like *go*, *have*, or *take*). This partial non-compositionality cannot be explained by the ILMs so far.

The second problem with the results shown previously is that the end results are always completely stable. However, one of the striking features of languages is that they are always changing. This has been termed "the logical problem of language change" (Niyogi and Berwick, 1995). Kirby (in press) tries to address these issues within the context of the iterated learning model.

4.1 Laziness

In this version of the model, we include two extra features, both of which are motivated by our understanding of real language use. Firstly, learning is not the only mechanism at work in the transmission of language from generation to generation. When producing utterances, speakers may have multiple choices to make about how to express a particular meaning. As a simplification we assume that speakers are motivated at least in some part by a principle of *least effort*. To model this, agents in this simulation will always produce the shortest string for a given meaning.[7] In addition, we allow that the production of language may not always be perfect. In this simulation, therefore, we add a parameter that expresses a per-character probability of a character in the output string being dropped from the utterance. For the results reported here, this probability is 0.001.

4.2 Frequency

A second modification that we include is motivated by looking at which English verbs are irregular in the past tense. As Pinker (1999) notes, the top ten verbs of English ranked by frequency (*be, have, do, say, make, go, take, come, see, get*) are all irregular (Francis and Kucera, 1982). In the previous simulations, however, all the meanings have identical probability of being picked for production by the speakers.

To include a frequency bias, we define a non-uniform probability distribution over the meaning space that reflects the chance that a meaning will be picked. To make this easier to do, and to make the results more visualisable, we have simplified the meaning space for this simulation. Rather than a complex of predicates and arguments, each meaning can be thought of as a combination of just two components. These could correspond to two properties of an object or situation, for example shape and colour, or action and tense. Here, we will refer to the meanings as having an a feature and a b feature, with the meanings ranging

[7] It is interesting to note that in an analysis of a simulation by Batali (1998), Goroll (1999) suggests that a least-effort principle is implicit in the production mechanism that Batali uses and may explain some of the features of his results.

from (a_0, b_0) to (a_4, b_4). The probability of a particular meaning (a_i, b_j) is proportional to $(i+1)^{-1}(j+1)^{-1}$. This frequency distribution is inspired by the recognition that word usage is inversely proportional to its frequency rank (Zipf, 1936).

4.3 Frequency correlates with irregularity

The ILM was run with the same induction procedure as in the previous example, but with fewer training examples — 50 in this case. This reflects the fact that the meaning space is much smaller, with only 25 possible combinations as opposed to the potentially infinite number of predicates that could be formed recursively in the last section. It should be noted, however, that although the meaning space is small, each learner is still not guaranteed to see every meaning during learning. In fact, given the frequency distribution, it is vanishingly unlikely that all meanings will be present in a particular speaker's output.

To summarise the results: after several generations during which the language remains inexpressive and idiosyncratic (i.e. a protolanguage), structure emerges in the simulation. Here is an example of the complete language of the 256th speaker presented as a table:

	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

Not only does this language have irregular forms (shown in bold in the table), but these correspond to the most frequent meanings (in the top left here). Furthermore, this language does not represent a stable attractor. There is a process of language change that continues throughout the simulation. Sometimes substrings corresponding to particular components of meaning will get shorter, other times completely new forms will emerge. However, the basic partially irregular compositionality is always in place. The irregular forms seem more stable than the regulars. For example, the irregular cluster here arose in generation 127 and lasted until generation 464, at which point y is reregularised to yi. It is also interesting that the length of the strings seems to be inversely correlated with frequency as has been observed in real languages (Zipf, 1936).

The only factor which can help an idiosyncratic meaning-form pair (i.e. one which does not conform to any generalization) to persist through the history of a language is increased frequency of use. An experiment described in Hurford (2000) corroborates the conclusion that frequency correlates with irregularity, but reveals an interesting difference from the example with irregular morphology just given. In this simulation, an iterated learning model was implemented, with a population of agents starting with no language at all, and over time a language emerged in the community in which there were completely general compositional rules for expressing a range of meanings represented as formulae in predicate logic. A variation on the basic simulation was then implemented in which one particular meaning was used with vastly greater frequency than any of the other available meanings. This inflated frequency held throughout the simulated history of the community. The result was that, as before, a language emerged in the population characterized by general compositional rules, but in addition, all speakers also had one special idiosyncratic stored fact pertaining to the highly frequent meaning.

The irregular treatment of the high-frequency item came about through the persistence of this one idiosyncratic meaning-form pairing from the earliest stages of the simulation. In the early stages, before any regular grammatical rules had emerged, *all* meanings were expressed idiosyncratically. Over time, the task of representing the other form-meaning pairings was taken over by general grammatical rules. But the single high-frequency meaning always (because of its frequency) happened to be encountered early in the learning experience of every agent, before the point where the agent had started to form any generalizations over observed examples. Thus the archaic idiosyncratic form-meaning pairing was perpetuated. No other meaning had the 'privilege' of assured early exposure to each new learner, and none of them ended up with an idiosyncratic irregular form. If this simulation were allowed to run on indefinitely, it is predicted that at some stage the eventual occurrence of an extremely improbable throw of the dice permitted by the random presentation of meanings would give a generation in which even this high-frequency meaning was unrepresented in examples given to the next generation, and its expression would then succomb to regularization.

This model of the origins of irregularity differs from the model dealing with morphological irregularity given earlier, and in fact seems less realistic. It predicts that languages inexorably lose irregularities, albeit perhaps very slowly in the case of high-frequency items. Besides a Zipfian frequency distribution, the morphological simulation incorporated the realistic factor of random erosion of signals by noise during transmission. In this way, new forms not conforming to previously established regularities are constantly liable to enter the language. This mechanism is plausible as modelling the entrance of at least some of the irregular verbs of Modern English, such as *had* (from the more regular OE *haefde*), *made* (cf the corresponding regular German *machte*, and *says* and *said* with their irregularly shortened vowels (i.e. not rhyming with the phonologically regular *plays* and *played*).

5 Why social transmission favours linguistic generalisation

A theory of language transmission through iterated acquisition needs to take into account the capacities brought to the acquisition task by the child and the nature of the data to which the child is exposed. Clearly, humans are capable of acquiring and storing economical recursive grammars that generate infinite sets of sentences. Equally clearly, humans are also capable of acquiring and storing vast, but of course finite, inventories of arbitrary facts.

However many examples of meaning-form pairs a learner is exposed to, the number of examples is always finite. We refer to the finiteness of the examples presented to each generation of learners as a 'bottleneck'. In fact, various kinds of bottleneck in the transmission of linguistic information between generations are conceivable, and these are discussed in more detail in Hurford (2001a). The kind of bottleneck most crucial to the iterated learning model is a 'semantic bottleneck'. This refers to the fact that only a small proportion of all the available meanings are ever expressed during the learning experience of any learner; the learner is not exposed to an exhaustive review of all the meaning-form pairings of the language. In the interesting case of an infinite language, this is necessarily the case.

It is worth considering for a moment what would happen in an evolving population whose members were not capable of acquiring grammars expressing generalizations over data, but could only resort to storage of arbitrary individual meaning-form mappings. Such agents could only acquire, and subsequently pass on to the next generation, the meaning-form pairs they had actually observed. If the agents in a given generation happened never to be motivated

to express a particular meaning for which they had acquired a signal, then that particular meaning-form pair would get lost in the historical process, and disappear from the language of the community. Only those meaning-form pairs which were guaranteed to be in the input to every new generation of learners would survive. If no meaning-form pairs were guaranteed such totally reliable exposure with each generation, no pairs would persist indefinitely through the history of the language. If some meanings were expressed more frequently than others each generation, these would tend to have historically more stable forms (not necessarily regular, as we have seen). Wherever a meaning-form mapping gets dropped in the historical process, the next time an agent is prompted to express that particular meaning, a new random form will be invented. The more such invention happens, the less stable is the language.

This simple scenario illustrates the idea of a language 'surviving' in a community of agents, and undergoing change which is predictable in terms of the learning capacity of the agents, the number of examples that each learner is exposed to, and the frequency distribution of the meanings. In this simple example, the agents only have recourse to a learning strategy which correlates examples observed during learning on a one-to-one basis with items of stored knowledge. Basically, such agents are capable only of internalizing a lexicon.

Now consider a situation in which agents have two kinds of strategy at their disposal. They can either store individual facts, as before, or they can *generalize* over examples, thus producing a many-to-one correlation between observed examples and items of stored knowledge. It does not matter at all how this generalization is done. Implicit in any notion of generalization in this context is the idea of *variables* ranging over (parts of) meanings or (parts of) forms. Also implicit, and essential here, is the idea that such generalization can be a form of *over*generalization, extending to some meaning-form pairs which the agent has not observed. This immediately builds into the agent a capacity to express meanings for which there were no exact precedents in its learning phase. Now there is the possibility of a particular form-meaning pair being absent from the language input to one generation of learners without it necessarily therefore being lost thereafter in the history of the language. A particular meaning-form pair can 'go underground', even for many generations, provided that all learners over the period acquire, from other examples, a generalization covering this particular meaning-form pair. But any meaning-form pair which is *not* covered by a generalization in the minds of learners/producers of the language, and which happens to be absent from the examples given to the next generation, will drop out of the language. Thus we see a difference in the survival potential of different types of meaning-form pairs in a language. Those that are covered by generalizations are much more likely to persist in the history of the language than those which are not covered by any generalization.

It follows that the strength or coverage of generalizations correlates with the survival potential of the meaning-form pairs they cover. Compare two generalizations made by a learning agent, one covering proportion n of the space of meanings and the other covering proportion m of the meaning space, where $0 < n < m < 1$. If this agent as an adult produces B distinct meaning-form pairs[8], in response to being prompted to express randomly chosen meanings, the probabilities that a learner in the next generation will observe at least one example covered by these generalizations are $1-(1-1/n)^B$ and $1-(1-1/m)^B$ respectively. Thus, in the simplified case where observation of just one example is sufficient to induce a generalization of any degree of coverage in a learner, a stronger generalization is more likely than a weaker generalization to be exemplified for the next generation, and thus transmitted

[8]assuming, for simplicity, no homonymy and no synonymy.

onward in the history of the language. If we further take into account the fact that the number of examples actually required to induce a generalization of coverage n is likely to increase at most as some exponential root of n increases, the enhanced survival potential of stronger generalizations becomes even clearer. The favouring of linguistic generalizations by the very fact of social transmission (iterated learning) is explored in more detail in Hurford (2000); the mathematics of the required relationships between sample size and coverage of the generalization are discussed in more detail in (Brighton and Kirby 2001b), and Brighton and Kirby (2001a).

The historical persistence of patterns of generalization in a language can be observed in its E-language, in a corpus of utterances produced by members of the population. In addition, native speaker investigators of a language can examine their own intuitions about made-up examples to test the generality of hypothesized rules. In real empirical linguistics, the actual representations of general rules in speakers' heads are inaccessible; they are hypothesized from E-language performance and grammatical intuitions. Generative linguistics has typically emphasized the role of child language acquirers as powerful generalizing machines. This infant drive for internal generalization has been taken to be the prime mover in causing the regularities observable in languages. And this conclusion is partially reinforced by the computational work reported here. Creatures with no drive at all to make mental generalizations from their observations would produce no historical E-languages with persistent regular patterns. But, as in evolution generally, one must distinguish between the evolutionary *source* of an attested phenomenon and the *reason for its persistence*. The evolutionary source of generalizations in languages is the child's innate capacity to generalize. The reason for the historical persistence of generalizations is the inherent advantage that general patterns have over idiosyncratic facts in being propagated across generations in the repeated spiral of acquisition and production (iterated learning). This inherent advantage of generalizations has nothing to do with the fitness of individual agents, nor with selection for ease of processing or usefulness in the community. Given the natural assumptions of the iterated learning model, it is a mathematical truism that languages characterized by very general patterns will emerge over time.

Linguists have tended to relegate any capacity for memorizing arbitrary knowledge items to facts which do not fall under generalizations, such as idioms and specific lexical items. Perhaps because the inherent advantage of generalizations in transmission across generations has not hitherto been recognized, it has been assumed that learners store economical, non-redundant grammars. If the organization of linguistic knowledge inside peoples' heads were messy, not capitalizing maximally on available generalizations, then, so the implicit argument goes, we should expect languages themselves to be messy. But languages are very regular (though admittedly not *completely* regular, and so the internal representation of grammars should be as regular and general as is compatible with the E-language data.

Computer simulations allow us an insight into the possible relationships between the linguistic knowledge stored in speakers' heads and the patterns of use in the community's E-language which exemplify generalizations. As in the examples of earlier sections, we can actually inspect the internal grammars of simulated agents. An experiment reported in Hurford (2000) shows that even in a population of agents which are significantly handicapped in their capacity to generalize over observed data, a regular, general E-language will emerge in the course of glossogeny. Agents in this experiment, like humans and all the artificial agents in the other simulations reported here, had both a capacity to represent lists of arbitrary unrelated facts and a capacity to make generalizations over facts perceived as related. But in this experiment, the agents' capacity for generalization was deliberately switched off for

a random 75% of the training examples. In this model, agents either internalized general rules where possible, (25% of the time), or simply rote-memorized the current example as a fact unrelated to any of the rest of their stored knowledge (75% of the time). The resultant internalized grammars of all the agents, were, as expected, very redundant, containing both general rules and lists of one-off records of form-meaning pairings. But in fact all of the one-off records were completely consistent with the general rules which the agents had also internalized. The general rules had become fixed in the language even though three-quarters of every instance of language-learning had been biased against the internalization of generalizations. An observer who could see only the actual utterances of the agents would have no reason to suspect that their internal representations redundantly duplicated facts by storing both general rules and many individual examples entirely consistent with them.

6 Extending the scope and context of ILM models

We have presented the central features of the iterated learning model, and outlined the main results obtained so far. We believe the model illustrates a selective force inherent in the fact of the social transmission of languages, through their I-language and E-language phases across generations. As with any progressive research paradigm, new questions are raised and new avenues for investigation are immediately suggested. We review some of these related issues here.

6.1 The treatment of meanings

The implementations of the iterated learning model described here all make the simplifying assumption that children are presented with signals paired with whole meanings. Obviously, if the whole meaning of an utterance were actually observable in this way, there would be no need for the utterance in the first place. Language is used to convey meanings which are *not* wholly obvious from the context. However, in order for learning of a mapping between meanings and forms to be possible, at least some parts of intended meanings must be evident to the child. Several works have explored and implemented algorithms for learning meaning-form mappings from data in which the meanings are partially masked (Siskind 1996, Hurford 1999). It seems clear that the simplifying assumption of the child being given whole meanings does not make possible what would otherwise be impossible. Its principal effect is to simplify the implementations and considerably to shorten the experimental time in which results can be obtained. The masking of meanings in actual language acquisition can be safely idealized away from, as far as the implications of the ILM discussed here are concerned.

A related, and deeper, objection to the treatment of meanings in these ILM models is that they all assume meanings to be entirely inside the agents' heads, and moreover the 'same' meanings are identically represented in all agents. The meanings used here are not related in any way to aspects of an external environment in which the agents must survive, and about which they communicate. In ILM models it is taken for granted that agents learn meaning-form mappings as children and as adults produce utterances when prompted by particular meanings. The idealization away from considerations of fitness and survival was useful insofar as it revealed a mechanism by which languages adapt themselves to transmission across the generations of their host populations. But consideration of the wider context invites us to ask whether the treatment of meanings can be justified or explained.

Ongoing research at LEC (Smith 2001a), following and extending the work of Steels (e.g., Steels 1999), attempts to model the growth of conceptual representations inside agents' heads in response to a 'discrimination task' acted out in a (simulated) world of object external to the agents. In these simulations, objects in the world presented agents with gradient information about objects along several channels. Agents develop internal representations implemented as sets of trees, with one tree for each perceptual channel. The trees start as single nodes and get arborized or ramified more or less densely at various regions in the continuous ranges of the channels, to the point where every object in the environment can be distinguished by a combination of one or more node-addresses on the trees. This begins to approach the classical semantic distinction between sense and reference ((Frege 1892)). The external objects correspond to the classical referents, while the developed internal trees correspond to agents' senses. Interestingly, under certain conditions, especially where there is no innate bias toward developing one channel/tree in preference to any other, populations emerge in which all agents are capable of making the same distinctions between objects, but may have distinct internal representations. Further work adds communication between agents to this system, and investigates whether the task of communicating about classes of objects in the environment results in more uniform internal representations across the population.

If it should turn out that agents do not necessarily share common internal semantic representations, this does not necessarily invalidate the basic results of ILM simulations, which so far have assumed common meaning representations. Undoubtedly, real humans share a common external environment, perceived through essentially similar sense organs, and there is surely some substantial overlap between representations in the minds of different individuals, even if there are also differences. The current ILM simulations can be taken as modelling the evolution of meaning-form mappings in the areas of overlap, which presumably are substantial. For areas in which different people conceptualize the world differently, we could conclude, with Wittgenstein, that "whereof one cannot speak, thereupon one must be silent".

The evolutionary story need not end on this Wittgensteinian note, envisaging a stable fixed boundary between the effable and the ineffable. Presumably, a population which manages to reduce the domain of "whereof one cannot speak" will reap certain benefits from being able to communicate about a wider set of experiences. Indeed, it is becoming clear that ILM models can shed light on how agents' conceptual representations adapt to the problem of fitting into a system of meaning-form mappings which can be transmitted across generations. The work surveyed above emphasizes that, for large or infinite meaning spaces, and with the natural severe bottleneck at the point of learning, only languages characterizable by general, compositional rules are stably transmitted across generations. Work by Brighton and Kirby (Brighton and Kirby 2001b, Brighton and Kirby 2001a) relates the likelihood of achieving such stable compositional systems to a range of possible organizations of the space of meanings. They conclude

> "If the perceptual space of the agent is not broken into multiple features or multiple values, then compositionality is not possible. However, even when the conceptual space is cut up into a few features or values, compositionality is still unlikely. The principal reason for this is that a simple meaning space structure results in the rate of observation of feature values to be near the rate of observation of whole meanings. This situation would result in less of a stability advantage for compositionality. ... The more complex the meaning space, the more payoff in stability

> compositional language offers. However, too much complexity leads to a decrease in payoff. ... with a highly complex meaning space structure the meanings corresponding to the objects are scattered over a vast space, and as a result, regularity in the correspondence between signals and meanings is weakened." (Brighton and Kirby 2001b)

Several of the simulations mentioned above have taken predicate logic representations 'off the shelf' as convenient ways of representing propositional meanings. Hurford, in recent work, has argued that neural correlates exist for a basic component of such logical formulae, *PREDICATE(x)*, where x is an individual variable and *PREDICATE* is any 1-place predicate constant. Such a formula represents the brain's integration of the two processes of sensory 'reference' to the location of an object, mapped in parietal cortex, and analysis of the object's properties by perceptual subsystems. The brain computes actions with a few 'deictic' variables pointing to objects, linking them with 'semantic' information about the objects, corresponding to logical predicates. Mental scene-descriptions pre-existed language phylogenetically, constituting a preadaptive platform for human language. (Hurford 2001b)

6.2 Which learning algorithms work?

The versions of the ILM reported here are the successful ones, where 'successful' means leading to the emergence of a stable communication system with at least some of the characteristics of a human language. We have concentrated here on models in which simple, but general and compositional, syntactic systems arise in the community. Earlier work naturally started with the emergence of simple vocabularies (e.g., Werner and Dyer (1991), Ackley and Littman (1994), Maclennan and Burghardt (1994), Levin (1995), Cangelosi and Parisi (1996), Oliphant (1996), Bullock (1997), de Bourcier and Wheeler (1997), Di Paolo (1997), Werner and Todd (1997), Noble (1998)). In both the lexical and the syntactic work, it transpires that a stable system only arises in the community if the learning algorithm has a certain property, which we have called 'obverter'. As we mention is section 2.1, an obverter learner is one whose acquired production behaviour is guided by a consideration of how the learner itself would interpret a signal. The learner thinks something like this: "To express meaning M, I will use signal S, because I would understand S as meaning M." An obverter strategy for language learning anticipates (or constructs) the essential bi-directional nature of the linguistic sign, allowing use of the same body of knowledge both in speaking and in hearing.

Smith (2001b) has systematically investigated a class of simple neural nets which take representations of signals as input and give representations of meanings as output, and whose connection weights can be adjusted in response to training on input-output pairs. Smith classified the different net-types into a hierarchy of 'Constructors', 'Maintainers', 'Learners' and 'Nonlearners'. Constructors are nets embodying a learning algorithm which, if embedded in an ILM starting from no language at all, will result after many generations in an emergent efficient vocabulary code shared by the whole population. Maintainers, if used in an ILM, can manage to maintain an initially given code against a certain amount of noise in transmission across generations, but cannot contribute to the glossogenetic evolution of a code emerging from an initial 'zero' situation. Learners are able to acquire, and subsequently faithfully transmit a code to a succeeding generation, but only in the complete absence of noise. Nonlearners cannot even faithfully acquire a code to which they are exposed. Only the constructor nets have the obverter property.

Because acquisition of syntax is much more complex than acquisition of vocabulary, it is hardly possible to conduct such a systematic comparison in the space of possible syntax acquisition algorithms. But it seems very likely that the only learning algorithms which will enable an ILM to evolve a stable syntactic system will also have the obverter property.

6.3 Coevolution

ILM simulations model a form of cultural evolution of language, which we call 'glossogeny'. As our introduction mentioned, the glossogenetic process is just one of several interacting evolutionary systems. ILM simulations work with specific learning algorithms. It is of interest to investigate how the two evolutionary processes, the phylogenetic and the glossogenetic, interact with each other. Complex simulations can be set up in which an ILM is embedded into a larger genetic algorithm which breeds language learners selected for good adult communication abilities.

Almost all simulations of the phylogeny of language associate fitness with the capacity to communicate. Combine this idea with the essential arbitrariness of the meaning-to-signal mappings that constitute languages, and a highly significant peculiarity of language evolution becomes apparent. A would-be communicating agent is born into a community with a pre-existing code of arbitrary meaning-form mappings. If fitness is correlated with the power to communicate, the agent's first priority is to become a fully participating member of the communicating population into which it was born. The language environment to which a human baby must adapt is not universal across the species, but is local and was constructed by previous generations of the local group solving the same survival problem. Now, if the language of the community is actually not an efficient code (say by having too much homonymy, or by being noncompositional), it is nevertheless more important for the newcomer to conform to that code than to attempt to improve it in any way. This implies that glossogenetic evolution can put a significant drag on phylogenetic adaptation. There are parallels with cultural evolution generally, where a society can perpetuate dysfunctional traits, such as approval of extreme forms of self-harm.

Kirby and Hurford (1997) describes a simulation in which an ILM is embedded into a genetic algorithm selecting over variants of an idealized learning algorithm, modelling the 'Principles and Parameters' version of generative linguistic language acquisition theory. A language learner can be innately endowed with more or less plasticity. In this theory, grammatical Principles correspond to genetically fixed aspects of language; Parameters represent aspects of language where the learner can adapt its acquired grammar to the ambient language of the community. The presence of mutations and crossover in the genetic algorithm allow innate language learning Principles to be replaced by Parameters, and *vice versa*. In this simulation, certain aspects of grammar were artificially and arbitrarily assigned a functional advantage, so that language learners could be influenced by a preference for languages with certain characteristics in preference to others. As expected, the emergent languages of the populations in the ILM embedded in this model tended over time to veer towards these preferred characteristics. After some time, the languages remaining in the simulated world were significantly skewed towards incorporating the designated positive functional characteristics. This represents a gradual skewing of all the language environments in the world into which a child can be born. Languages evolve glossogenetically in directions in which they are steered by functional pressures. Bearing in mind that it is always better for a child to learn the language of its community than to branch out on its own, this glossogenetic response

to functional pressure is slow. But once such functional pressure has had some appreciable effect on the distribution of language universals, all human children will be born into linguistic environments biased in a certain direction. At this point it it can be expected that phylogenetic evolution will begin to bite, and that there may be some evolution of the innate language learning device toward a preference for universals which originated as outcomes of the glossogenetic process (see also, Yamauchi, 1999).

7 Summary

In this chapter we have argued that there are three complex adaptive systems at work in the evolution of language: ontogeny, glossogeny and phylogeny. Our understanding of the interaction of these aspects of language evolution can be enhanced through a computational modelling approach. If, as we argue, much of the structure of language is *emergent*, then a modelling methodology is appropriate, since it is notoriously difficult to come up with reliable intuitions about emergent behaviours. To help approach the problem of modelling how language acquisition influences glossogenetic evolution, we have put forward the iterated learning model.

Using the ILM we have argued that some of the most fundamental features of human language can be best explained in terms of the pressures on language transmission. As Deacon suggests in the quotation that begins this chapter, the primary pressure on the evolution of languages (as opposed to language users) is the need to be learnt. If a language, or some part of a linguistic system, is not learnable then it will not persist. In a very real sense, the data a learner is exposed to is a bottleneck in the transmission of the knowledge of language. In one of our simulations, the meaning-signal mapping has a potentially infinite extent, but must nevertheless be recovered every generation from a finite sample of randomly chosen utterances. This provides a severe selection pressure (though a *linguistic* rather than *natural* one) on the behaviour that is being replicated from generation to generation.

Fundamentally, the ILM demonstrates that we cannot draw a direct parallel between the innate properties of the language user and the structure of language. The early languages in the simulations are learnt and used by the agents — in other words they are not *ruled out* innately. However, they are not stable; only compositional languages have that property. From this we can conclude two things:

1. Even if we could examine the innate language faculty in minute detail, we could not directly read-off the structure of human language from it.

2. Conversely, given some universal property of language that we wish to explain, we should not necessarily place the whole burden of that explanation on an innate biological property.

Finally, if we consider what recursive compositionality gives us as a species it is tempting to conclude that communicative function must have a central role in its explanation. With it we can express an infinite range of ideas. We are not trapped in the prison of prior experience, forever constrained to convey only those things that have previously been conveyed to us. We do not deny here the obvious communicative advantage we have as a species that syntax buys us. However, counter to intuition, our explanation for syntactic structure does not make any reference to communication. It is the advantage that compositionality gives to language itself that drives its evolution, the benefits this has to us are merely a fortunate side-effect.

References

Ackley, D. and M. Littman (1994). Altruism in the evolution of communication. In R. Brooks and P. Maes (Eds.), *Artificial Life 4: Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems*, pp. 40–48. Redwood City, CA: Addison-Wesley.

Batali, J. (1998). Computational simulations of the emergence of grammar. In J. R. Hurford, M. Studdert-Kennedy, and C. Knight (Eds.), *Approaches to the Evolution of Language: social and cognitive bases*, pp. 405–426. Cambridge: Cambridge University Press.

Batali, J. (2001). The negotiation and acquisition of recursive grammars as a result of competition among exemplars. In E. Briscoe (Ed.), *Linguistic Evolution through Language Acquisition: Formal and Computational Models*. Cambridge: Cambridge University Press.

Belew, R. (1990). Evolution, learning, and culture: computational metaphors for adaptive algorithms. *Complex Systems 4*, 11–49.

Bickerton, D. (1990). *Language and Species*. University of Chicago Press.

Boyd, R. and P. J. Richerson (1985). *Culture and the Evolutionary Process*. University of Chicago Press.

Brighton, H. and S. Kirby (2001a). Meaning space structure determines the stability of culturally evolved compositional language. Submitted to the Cognitive Science Society conference 2001.

Brighton, H. and S. Kirby (2001b). The survival of the smallest: stability conditions for the cultural evolution of compositional language. Submitted to the European Conference on Artificial Life 2001.

Briscoe, T. (2000). Evolutionary perspectives on diachronic syntax. In S. Pintzuk, G. Tsoulas, and A. Warner (Eds.), *Diachronic Syntax: Models and Mechanisms*. Oxford: Oxford University Press.

Bullock, S. (1997). An exploration of signalling behaviour by both analytic and simulation means for both discrete and continuous models. In P. Husbands and I. Harvey (Eds.), *Fourth European Conference on Artificial Life*, pp. 454–463. Cambridge, MA: MIT Press.

Cangelosi, A. and D. Parisi (1996). The emergence of a 'language' in an evolving population of neural networks. Technical Report NSAL-96-004, Institute of Psychology, National Research Council, Rome.

Chomsky, N. (1986). *Knowledge of Language*. Praeger.

de Bourcier, P. and M. Wheeler (1997). The truth is out there: The evolution of reliability in aggresive communication systems. In P. Husbands and I. Harvey (Eds.), *Fourth European Conference on Artificial Life*, pp. 444–453. Cambridge, MA: MIT Press.

Deacon, T. (1997). *The Symbolic Species*. Penguin.

Di Paolo, E. (1997). An investigation into the evolution of communication. *Adaptive Behaviour 6*, 285–324.

Francis, N. and H. Kucera (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin.

Frege, G. (1892). über sinn und bedeutung. *Zeitschrift für Philosophie und philosophische Kritik 100*, 25–50.

Goroll, N. (1999). (the deep blue) nile: Neuronal influences on language evolution. Master's thesis, University of Edinburgh.

Hurford, J. R. (1999). Language learning from fragmentary input. In K. Dautenhahn and C. Nehaniv (Eds.), *Proceedings of the AISB'99 Symposium on Imitation in Animals and Artifacts*, pp. 121–129. Society for the Study of Artificial Intelligence and the Simulation of Behaviour.

Hurford, J. R. (2000). Social transmission favours linguistic generalization. In C. Knight, M. Studdert-Kennedy, and J. Hurford (Eds.), *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, pp. 324–352. Cambridge: Cambridge University Press.

Hurford, J. R. (2001a). Expression/induction models of language evolution: dimensions and issues. In T. Briscoe (Ed.), *Linguistic Evolution Through Language Acquisition: Formal and Computational Models*. Cambridge: Cambridge University Press.

Hurford, J. R. (2001b). The neural basis of predicate argument structure. Submitted to Behavioral and Brain Sciences.

Kirby, S. (1999a). *Function, selection and innateness: the emergence of language universals*. Oxford: Oxford University Press.

Kirby, S. (1999b). Learning, bottlenecks and infinity: a working model of the evolution of syntactic communication. In K. Dautenhahn and C. Nehaniv (Eds.), *Proceedings of the aisb'99 symposium on imitation in animals and artifacts*. Society for the Study of Artificial Intelligence and the Simulation of Behaviour.

Kirby, S. (1999c). Syntax out of learning: the cultural evolution of structured communication in a population of induction algorithms. In D. Floreano, J. D. Nicoud, and F. Mondada (Eds.), *Advances in Artificial Life*, Number 1674 in Lecture notes in computer science. Springer.

Kirby, S. (2000). Syntax without natural selection: how compositionality emerges from vocabulary in a population of learners. In C. Knight, M. Studdert-Kennedy, and J. R. Hurford (Eds.), *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, pp. 303–323. Cambridge: Cambridge University Press.

Kirby, S. (2001). Learning, bottlenecks and the evolution of recursive syntax. In T. Briscoe (Ed.), *Linguistic Evolution through Language Acquisition: Formal and Computational Models*. Cambridge: Cambridge University Press.

Kirby, S. (in press). Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity. *IEEE Journal of Evolutionary Computation*.

Kirby, S. and J. R. Hurford (1997). Learning, culture and evolution in the origin of linguistic constraints. In P. Husbands and I. Harvey (Eds.), *Fourth European Conference on Artificial Life*, pp. 493–502. Cambridge, MA: MIT Press.

Levin, M. (1995). The evolution of understanding: a genetic algorithm model of the evolution of communication. *Biosystems 36*, 167–178.

Maclennan, B. and G. Burghardt (1994). Synthetic ethology and the evolution of cooperative communication. *Adaptive Behaviour 2*, 161–187.

Newmeyer, F. (1991). Functional explanation if linguistics and the origins of language. *Language and Communication 11*, 3–28.

Niyogi, P. and R. Berwick (1995). The logical problem of language change. Technical Report AIM-1516, MIT AI Lab.

Noble, J. (1998). Evolved signals: Expensive hype vs. conspiratorial whispers. In C. Adami, R. Belew, H. Kitano, and C. Taylor (Eds.), *Artificial Life 6: Proceedings of the Sixth International Conference on Artificial Life*. Cambridge, MA: MIT Press.

Nowak, M., N. Komarova, and P. Niyogi (2001). Evolution of universal grammar. *Science 291*, 114–118.

Nowak, M. A., J. B. Plotkin, and V. A. A. Jansen (2000). The evolution of syntactic communication. *Nature 404*, 495–498.

Oliphant, M. (1996). The dilemma of saussurean communication. *BioSystems 37*, 31–38.

Oliphant, M. and J. Batali (1997). Learning and the emergence of coordinated communication. *Center for Research on Language Newsletter 11*(1).

Pinker, S. (1994). *The Language Instinct*. Penguin.

Pinker, S. (1999). *Words and Rules*. Weidenfeld & Nicolson.

Pinker, S. and P. Bloom (1990). Natural language and natural selection. *Behavioral and Brain Sciences 13*, 707–784.

Siskind, J. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. In M. R. Brent (Ed.), *Computational Approaches to Language Acquisition*, pp. 39–91. Cambridge, MA: MIT Press.

Smith, A. (2001a). Establishing communication systems without explicit meaning transmission. Submitted to the European Conference on Artificial Life 2001.

Smith, K. (2001b). The evolution of learning mechanisms supporting symbolic communication. Submitted to CogSci2001, the 23rd Annual Conference of the Cognitive Science Society.

Smith, K. (in press). Learners are losers: Natural selection and learning in the evolution of communication. *Adaptive Behaviour*.

Steels, L. (1997). The synthetic modeling of language origins. *Evolution of Communication 1*, 1–34.

Steels, L. (1999). *The Talking Heads Experiment*, Volume I. Words and Meanings. Antwerpen: Laboratorium. Special pre-edition.

Tonkes, B. and J. Wiles (in prep). Methodological issues in simulating the emergence of language. Submitted to the volume arising out of the Third Conference on the Evolution of Language, Paris 2000.

Werner, G. and M. Dyer (1991). Evolution of communication in artificial organisms. In C. Langton, C. Taylor, J. Farmer, and S. Rasmussen (Eds.), *Artificial Life 2*, pp. 659–687. Redwood City, CA: Addison-Wesley.

Werner, G. and P. Todd (1997). Too many love songs: Sexual selection and the evolution of communication. In P. Husbands and I. Harvey (Eds.), *Fourth European Conference on Artificial Life*, pp. 434–443. Cambridge, MA: MIT Press.

Wray, A. (1998). Protolanguage as a holistic system for social interaction. *Language and Communication 18*, 47–67.

Yamauchi, H. (1999). Evolution of the LAD & the baldwin effect. Master's thesis, University of Edinburgh.

Zipf, G. K. (1936). *The Psycho-Biology of Language*. London: Routledge.