PREFINAL DRAFT: Kirby, S. (2000). Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners. In Knight, C., editor, The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form, pages 303–323. Cambridge University Press.

Syntax without Natural Selection: How compositionality emerges from vocabulary in a population of learners

Simon Kirby
Language Evolution and Computation Research Unit
Department of Linguistics
University of Edinburgh
http://www.ling.ed.ac.uk/˜ simon

1 Introduction

How can we explain the origins of our uniquely human compositional system of communication?[1] Much of the recent work tackling this problem (e.g Bickerton 1990; Pinker & Bloom 1990; Newmeyer 1991; Hurford *et al*. 1998) explicitly attempts to relate models of our innate linguistic endowment with neo-Darwinian evolutionary theory. These are essentially functional stories, arguing that the central features of human language are genetically encoded and have emerged over evolutionary time in response to natural selection pressures.

In this paper I put forward a new approach to understanding the origins of some of the key ingredients in a syntactic system. I show, using a computational model, that compositional syntax is an inevitable outcome of the dynamics of observationally learned communication systems. In a simulated population of individuals, language develops from a simple idiosyncratic vocabulary with little expressive power, to a compositional system with high expressivity, nouns and verbs, and word order expressing meaning distinctions.[2] This happens without natural selection of learners — indeed, without any biological change at all — or any notion of function being built into the system.

This approach does not deny the possibility that much of our linguistic ability may be explained in terms of natural selection, but it does highlight the fact that biological evolution is by no means the only powerful adaptive system at work in the origins of human language.

2 The origins of syntax

Pinker & Bloom (1990) argue that an analysis of the design features of human language, and of syntax in particular, leads to the conclusion that the best way of understanding their origins is as biological adaptations. The central questions that should be asked in their view are:

[1]This research was carried out at the Collegium Budapest Institute for Advanced Study, and the Language Evolution and Computation Research Unit in the Department of Linguistics at Edinburgh, funded by ESRC grant R000236551. Much of the work described was carried out in collaboration with Mike Oliphant and Jim Hurford. Some sections of this paper are available in an earlier technical report (Kirby 1998b) which also has an appendix dealing with the model more formally.

[2]For parallel, and in certain ways contrasting, work, see Hurford (this volume).

"Do the cognitive mechanisms underlying language show signs of design for some function in the same way the anatomical structures of the eye show signs of design for the purpose of vision? What are the engineering demands on a system that must carry out such a function? And are the mechanisms of language tailored to meet those demands?" (Pinker & Bloom 1990:712)

Pinker and Bloom claim that the features of grammar which they are interested in form part of the innate endowment of humans and work together to make "communication of propositional structures" possible. For example, the existence of linear order, phrase structure and major lexical categories together will allow a language user to "distinguish among the argument positions that an entity assumes with respect to a predicate" (p. 713), suggesting that their presence in human languages requires a biological/adaptationist explanation.

There have been many authors (see, e.g. Hurford 1998 for a recent review) who have argued that it is useful to look at syntax as a product of natural selection — Newmeyer (1991, 1992), for example, looks in detail at the features of the "Principles and Parameters" model of syntax and gives them an evolutionary explanation. The reasons for this are clear, as Pinker & Bloom (1990:707) point out: "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."

I will show in this paper that, for at least some features of syntax, there are in fact "alternative processes capable of explaining such complexity", and that some of the qualitative evolution of human language proceeded without natural selection. The kind of evolution we will be looking at is not biological, but relies on a notion that *languages* themselves act as complex adaptive systems (Hurford this volume; Worden this volume; Briscoe this volume; Kirby 1998a; Kirby 1997b; Christiansen 1994; Deacon 1997; Kirby 1997a; Briscoe 1997; Gell-Mann 1992).

The particular feature of syntax that will be explored in this light — and one which subsumes many of Pinker and Bloom's list — is *compositionality*. Cann (1993:4) gives the following definition of the principle of compositionality, a universal of human language:

"The meaning of an expression is a monotonic function of the meaning of its parts and the way they are put together."

This definition makes it clear that, although compositionality is often taken to be a property of semantics, it is actually a property of the system that links forms and meanings.

3 A computational approach

If we are to fully understand the ways in which a learned, culturally transmitted, system such as language can evolve we need some sophisticated population models of learners. Simple theorising about the likely behaviour of complex adaptive systems is not good enough. As Niyogi & Berwick (1997) point out, our intuitions about the evolution of even simple dynamical systems are often wrong. Recently, many researchers have responded to this problem by taking a *computational* perspective (for example, Hurford 1989; Hurford 1991; MacLennan 1991; Batali 1994; Oliphant 1996; Cangelosi & Parisi 1996; Steels 1996; Kirby & Hurford 1997; Briscoe 1997; Briscoe this volume, Batali this volume, Hurford this volume).

This paper follows on from this line of work, and also borrows from language learning algorithms developed in computational linguistics (namely, Stolcke 1994) in order to see if a significant portion of the evolution of syntax can proceed without biological change. In many ways, this work is a logical extension of the work of Batali (1997) who simulates a population of recurrent neural networks.

3.1 Features of a desirable model

In order for it to be a successful model of the cultural adaptation of language, the computational simulation has to have a set of key features. These set out our minimum requirements. In general, we wish to make the model as simple as possible initially, and see if the complex behaviour that we are looking for emerges without extra assumptions. The basic requirements are:

1. Individuals that *learn observationally*. In other words, all the knowledge in the population is learned by individuals observing other's behaviour. Following Oliphant (1997), I use this term to contrast the model with ones which assume that learning proceeds through explicit reinforcement.

2. A gradual turnover of members of the population over time. By ensuring that members of the population are not "immortal" we can see that there is true historical/cultural transmission of knowledge through the system.

3. No selection of individuals. In order to show that biological evolution is not a factor in the results of the simulation, the "death" of members of the population should be completely random and not related in any way to their success at communication.

4. Initial non-linguistic population. Those individuals that make up the initial population should have no communication system at all. This means that any biases that emerge in later states of the simulation are purely a product of the learners and the population model.

The basic structure of the model is similar to that used by Oliphant (1997). Figure 1 shows Oliphant's diagram of how we can model populations of observational learners. The simulation maintains a population of individual learners which produce observable behaviour. Occasionally, individuals will die and be removed from the population. These individuals will be replaced with new individuals which learn from the body of observable behaviour that the population has produced.

There is actually not much more than this to the computational model. All that remains is to define what is meant by "observable behaviour", and expand on how we model individuals that can produce and learn this behaviour.

3.2 Utterances

For a model of a population of communicating individuals, we clearly need something for our individuals to talk about — in other words, we must provide the simulation with a set of *possible meanings*. For the purposes of demonstrating emergent compositionality, it is important that this set of meanings be structured in some way. If meanings were not decomposable

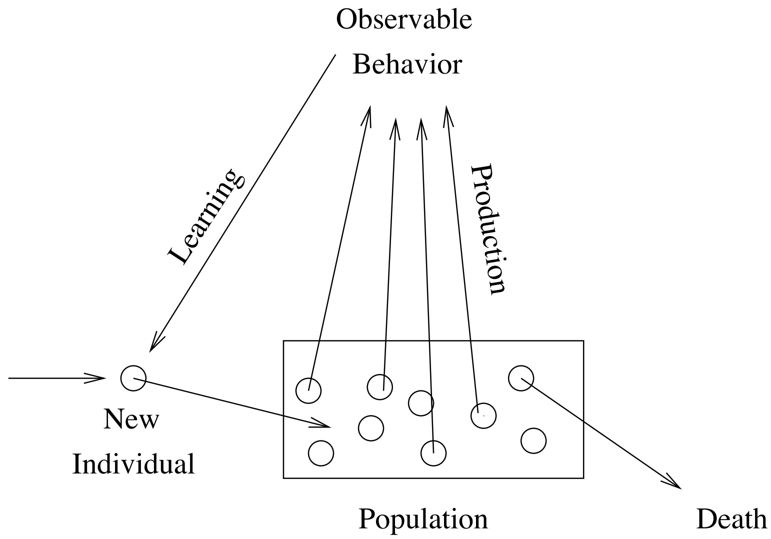

Figure 1: A framework for modelling populations of observational learners (from Oliphant 1997).

then it would be impossible for there to be a compositional system for communicating those meanings.

Each meaning in the simulation is a triple of attribute-value pairs. The three attributes can be glossed as: *Agent*, *Patient*, and *Predicate*. The set of possible values is divided into two classes, which can be glossed as: *Objects* and *Actions*. The *Agent* and *Patient* attributes can be paired with only *Objects*, whilst the *Predicate* attribute can be paired with only *Actions*. The *Object* class contains the values *Mike, John, Mary, Tünde* and *Zoltan*. The *Action* class contains the values *Loves, Knows, Hates, Likes* and *Finds*. An example meaning in this scheme could be: $< Agent = Zoltan, Patient = Mary, Predicate = Knows >$, which we can think of as being equivalent to the English sentence "Zoltan knows Mary". Essentially, the individuals' meanings are all about who did what to whom.[3] For purely implementational reasons, meanings with the same value for *Agent* as *Patient* are disallowed. This leads to a complete semantic space made up of 100 possible meanings.

The individuals in the simulation communicate through a serial channel with discrete symbols concatenated into a string. They have five of these basic symbols: *a*, *b*, *c*, *d* and *e*, which can be thought of as phonetic gestures. In principle, there is no limit on the length of an utterance, and the shortest possible utterance is one symbol long.

The observable behaviour in the model (which corresponds to the top part of Oliphant's diagram) is made up of pairs of meanings and symbol strings. This builds in an assumption that the intended meanings of utterances are, at least some of the time, accessible to learners.

[3] I hope it will be clear that in choosing these particular attributes and values I am not making any claims about what sort of things real individuals want to talk about. The names 'Agent', 'Patient' and 'Predicate' are purely devices to help us think about these triples as meanings. They could equally well have been given numbers (as indeed they are in the computational implementation of the model). The important feature of this semantics is that it has inherent structure, albeit a very simple one.

3.3 Individuals

In order to be able to produce utterances, the individuals in the model must have some way of representing a communication system internally, and a way of inducing such a representation from experience. There are many ways in which we might implement this. In Batali (1997), for example, the communication system is represented as a set of connection weights in an artificial neural network, and these weights are learned using a standard algorithm. The techniques used in the simulations described in this paper are described in detail in Kirby (1998b), but a flavour of them will be given here.

3.3.1 Internal representation

Each individual represents its communication system as a context-free grammar.[4] Importantly, the space of possible grammars is huge, and almost all of them are very un-language-like. In other words, by choosing a grammatical framework like this, we are not building in any unwanted inherent biases towards a compositionally structured system. Context free grammars allow us to express a range of systems from completely non-compositional to highly compositional.

Although, in this chapter I do not intend to go into much of the purely technical detail behind the simulations, it is worth illustrating what a context-free grammar looks like, and how it can be compositional *or* non-compositional. The two examples below should make this clearer.

The first is a (very simple!) non-compositional grammar[5] that produces the string *zoltanknowsmary* meaning $< Agent = Zoltan, Patient = Mary, Predicate = Knows >$:

$$S/ < Agent = Zoltan, Patient = Mary, Predicate = Knows > \rightarrow \text{ zoltanknowsmary}$$

This grammar has only one rewrite rule for the category S (i.e. sentence). The grammar can be interpreted as stating: "A sentence that means $< Agent = Zoltan, Patient = Mary, Predicate = Knows >$ can be expressed as the string of symbols *zoltanknowsmary*". This grammar is obviously non-compositional since the meaning of the sentence is not built up from the meaning of parts of that sentence.

The second example grammar also produces the same string/meaning pairing:

$$\begin{gathered} S/ < Agent = \mathbf{x}, Patient = \mathbf{y}, Predicate = \mathbf{p} > \rightarrow \; N/\mathbf{x} \;\; V/\mathbf{p} \;\; N/\mathbf{y} \\ V/ < Knows > \rightarrow \text{ knows} \\ N/ < Zoltan > \rightarrow \text{ zoltan} \\ N/ < Mary > \rightarrow \text{ mary} \end{gathered}$$

This grammar can be interpreted as stating: "A sentence can be made up of something of category N followed by something of category V followed by something of category N, if the meaning of that sentence is constructed by assigning the meaning of the first N to *Agent*, the second N to *Patient* and the V to *Predicate*. In turn something of category V that means

[4]Actually, the grammars are *probabilistic attribute grammars* (Stolcke 1994). These are context-free grammars which are enriched with statistical information and a simple semantics.

[5]The illustrative formalism used here is essentially identical to the one used internally in the simulation. It is just like a traditional phrase-structure grammar with semantics attached to category labels (after a slash) and variables indicated in bold. In the simulation, these rules also have frequency counts attached to them. Furthermore, the category labels (such as N and V in our example) are assigned arbitrary numerals in the simulation. This means that the learner does not have a limit on the number of categories that might be postulated.

$< Knows >$ can be expressed as a string of symbols *knows*, something of category N that means $< Zoltan >$ can be expressed as *zoltan*, and something of category N that means $< Mary >$ can be expressed as *mary*". This grammar contrasts with the previous one in being compositional, in that the meaning of the whole is built up from the meanings of its parts.

3.3.2 Invention

The initial individuals in the population have no linguistic knowledge — at the start of the simulation runs no-one is able to say anything. For anything to get off the ground there must be a way for novel forms to be produced. It is assumed that occasionally individuals, even though they have no normal way in which to express a certain meaning, will nonetheless produce some invented string of symbols.

There are different ways in which this might be done. The simplest approach is to produce a completely random string of symbols. Another possibility, used by Hurford (this volume), is to break down the meaning that is to be expressed into its atomic components, and then try to "synthesise" a symbolic representation of the sum of those components, perhaps by checking a lexicon for any matches to these atomic meanings. So, for example, if an individual was trying to express $< Agent = Zoltan, Patient = Mike, Predicate = Knows >$, then Hurford's technique would check to see if there was a way to say "Zoltan", "Mike" and "Knows" in isolation, and put together an utterance by combining these parts.

However, Hurford's (this volume) goal is not to model the emergence of compositionality, so his approach may not be the best one to use in this simulation. Indeed, a *synthetic* approach to some extent is bound to build-in the central feature of compositionality — that the meaning of the whole is composed of the meanings of its parts. Moreover, Wray (1998, this volume) suggests that language evolution did not proceed through the synthesis of small components into larger syntactic units, but rather that protolanguage consisted of holistic (i.e. non-compositional) utterances for complex meanings.

Given this, it would seem sensible to opt for a random invention technique. However, this is rather unrealistic for some cases. For example, imagine that you, as an English speaker, do not know the word for a new object that you have never seen before. It seems implausible that, if you needed to express a meaning that mentioned this object somewhere in it, you would utter a completely random string of phonetic gestures *for the whole sentence*.

Instead, whenever individuals invent a new form for a particular meaning, they do not introduce new structure, but equally, they do not throw away structure that is already part of the language they have acquired. The computational implementation of this invention strategy is described in detail in Kirby (1998b). Briefly, the invention algorithm used by the simulation generates random strings where the speaker has no grammatical structure, but for meanings that can be partially expressed with a particular grammar will only randomise those parts of the string that are *known by the speaker* not to correspond to expressible meaning.[6]

[6] In the simulation results reported here, the completely novel utterances invented by the speakers were set to vary randomly in length between 6 and 10 symbols. On top of this, during partial invention of utterances, the lengths of the invented strings could increase or decrease by one symbol with a small probability. The purpose of these arbitrary variables was to allow the string length to be potentially infinite, but likely to remain within a workable range. In fact, as the results described in a later section show, the languages that emerge seem to favour shorter utterances that can still express the meaning space.

3.3.3 Induction

Each individual in the simulation acquires a grammar based on experience of meaning-form pairs produced by the rest of the population. The simulation uses a simplified version of an algorithm developed by Stolcke (1994) for induction of context free grammars with semantics. Full details of the methodology are given in Kirby (1998b). Essentially, the learning process involves two steps:

1. **Incorporation** On receiving a meaning-form pair, the algorithm immediately builds a grammatical model for that pair which makes no generalising assumptions about it. In other words, the inducer will simply add a (completely non-compositional) rule to the grammar that states directly that a legal sentence in the language has the given form and corresponds to the given meaning. (For example, an incorporated rule for the English sentence "Zoltan knows Mary" would thus look very like the first simple grammar rule given in section 3.3.1.)

2. **Merging** Having built a grammatical model of a single utterance, the algorithm seeks to merge this model with the existing model for any previous utterances. Merging involves making changes to the rules in the grammar in such a way that two or more rules in the grammar become more similar to each other. The rational behind this is that learning can be viewed as compression of training into a compact hypothesis (Osborne & Briscoe 1997). If two rules in the grammar become identical, then one is redundant and is deleted. The merging algorithm thus tends to produce "minimal length" grammars for the observed utterances.

In practical terms, the way in which the induction algorithm seeks to merge the grammar will introduce constraints on the space of possible grammars that the learners can acquire. For example, the learners described in this paper cannot acquire recursive grammars (although, see Kirby (1999) for a simulation in which recursion is possible). This is not a serious concern, however, since the simple "who did what to whom" meanings that they have to convey are not recursive anyway.

3.4 The population dynamic

Given a computational model of an individual we need to set out the ways in which a population of individuals interacts. The population in the simulations reported here is made up of ten individuals at any one point in time, organised in a ring. In other words, each member of the population has two neighbours. Figure 2 shows how this population is updated over time.

Each cycle through the inner loop of figure 2, the speaker is "instructed" to produce a randomly chosen meaning. Especially at the start of the simulation, the speaker may well not be able to produce a string that corresponds to that meaning with the grammar that it has internalised. At this point, one of two things may happen; either the speaker says nothing, or the speaker may try and invent a new string (as described earlier). The rate at which inventions are introduced can be easily controlled in the simulation. For the results reported here, speakers produce inventions on average one time out of every fifty. If, on the other hand, the speaker *can* produce a string which corresponds to the meaning, then it does so, although noise is simulated in the model by replacing this string with a random one one time out of a thousand. The key points here are:

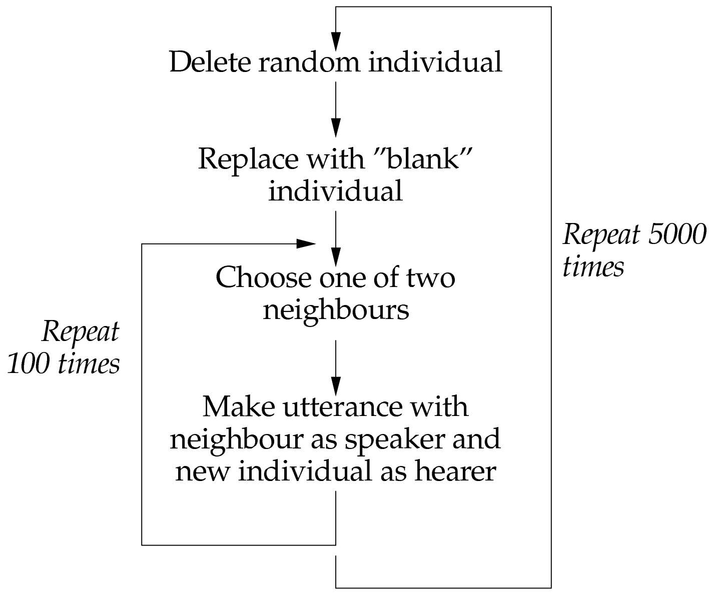

Figure 2: The main loop used in the simulations.

- Each individual learns only from utterances (form-meaning pairs) produced by its neighbours.
- The makeup of the population changes over time.
- Individuals are replaced entirely at random.
- The probability that one individual will hear all forms for all the possible meanings is vanishingly small.[7]

4 Results

This section looks in some detail at one particular run of the simulation described above. The behaviour of the simulation is consistent from run to run, so a careful analysis of one case is worthwhile.

The initial population is made up of ten individuals, all of which have no knowledge of language — that is, they have empty grammars. The simulation loop described in figure 2 is then initialised and left to run until the behaviour of the population stabilises (after several thousand generations). Periodically, various measures of the population's behaviour and knowledge are taken:

1. **Meanings** The number of meanings that an individual can express (without invention).

[7]There are 100 different possible meanings, and a maximum of 100 utterances heard by each individual. Even if an individual is lucky enough to hear 100 utterances in its lifetime, the chances that these will cover the entire meaning space are $\frac{100!}{100^{100}}$.

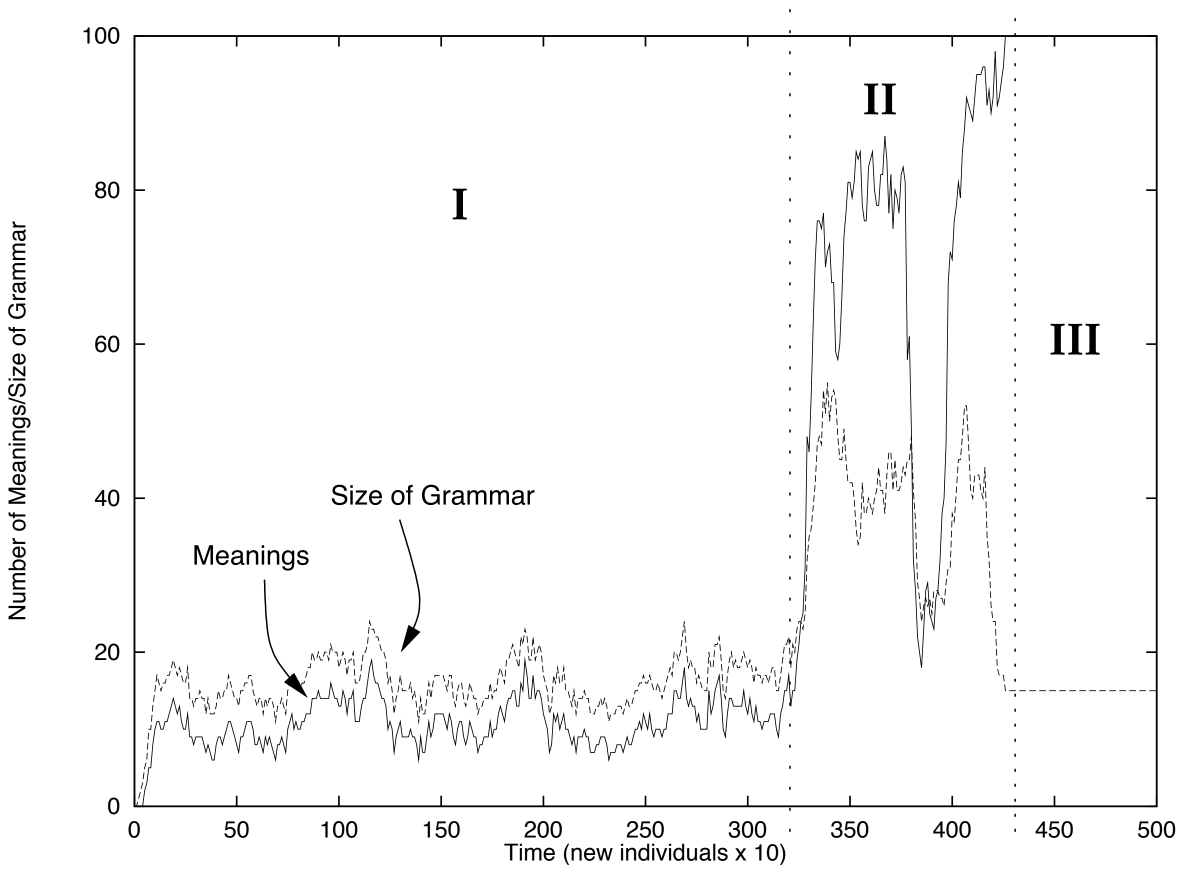

Figure 3: The population average of size, meanings and coverage over 500,000 sentences, where a "sentence" is an instruction to a speaker to produce a random meaning. The graph is divided into three stages signifying major "phase changes" in the grammars of the population.

2. **Size** The number of rules in an individual's grammar.[8]

3. **Grammars** The actual grammars of the individuals in the simulation can be directly inspected, so that we can analyse any internal structure to the language that evolves in the community.

A graph of the population average of meanings and size over a run of 5000 cycles through the simulation is given in figure 3.

The graph has been partitioned into three stages between which the population appears to make "phase transitions" into radically different types of behaviour. In particular, the

[8]The size is calculated by inspecting each individuals context free grammar and counting the number of rewrite rules. Notice that there are more rules than meanings initially in the graph, this is because, for purely technical reasons (discussed more fully in Kirby 1998b) each letter in an individual's language has an associated rewrite rule. In other words, the grammar contains an intermediate "pre-terminal" layer between sentences/vocabulary and strings of symbols. This fact does not affect the results of the simulation in any interesting way, but it does mean that the measure of grammar size is slightly higher than might otherwise be predicted. In fact, usually each language will use all 5 terminal symbols a, b, c, d and e, so there will be five extra rewrite rules in the grammar.

relationships between the two measures graphed and also the structure of the grammars changes radically at these points. These stages are present in every run of the simulation, although the timing of the transitions is variable.

4.1 Stage I

In the first few cycles of the simulation run nothing much happens. No individual in the population has any grammar, so they have no way of producing utterances. Each time an individual is asked to produce a string for a particular randomly chosen meaning, they consult their grammar and discover they have no way of producing a string so they say nothing. Consequently the new individuals have no exemplars for acquisition and also end up with empty grammars. Recall, however, that there are occasional random *invention* and *noise* events. Whenever one of these occurs, the new individual has something to internalise: a pairing of a randomly constructed string of symbols, and a randomly chosen meaning. Then, if this individual is later called upon to produce an utterance with that meaning, that same string of symbols will again appear in the input of a new learner.

This process of random invention and re-use leads to the situation that is stable throughout the first emergent stage in the simulation. The population can express only a small percentage of the meanings, using a small grammar. In fact, the grammars in this stage are basically vocabulary lists, with each complex meaning being expressed as an arbitrary unanalysed string of symbols. One such vocabulary list for a random individual picked out of the population at this stage is shown below:

Meaning (glossed in English)	String
John finds Mary	aceabbceeeabeea
John finds Zoltan	ceadaeeabbe
John hates Zoltan	ecdceaabdda
Mary finds Zoltan	adabeeb
Mary hates John	ddadbbbbabeedaeee
Mary hates Tünde	adababcccecadcbce
Mary hates Zoltan	ceeaebeebcecabdee
Mary loves Tünde	abacdddbe
Mike hates Mary	adddbdcceaa
Zoltan hates John	d
Zoltan hates Mike	e

Notice that only 11 out of the full 100 meanings can be expressed by this individual, and there is no consistent way in which the meanings are related to the strings. For example, *John hates Zoltan* is expressed as *ecdceaabdda* while *Zoltan hates John* is expressed as the completely unrelated string *d*. This complete lack of structure is confirmed when we look at a tree diagram produced by using the grammar of this individual to parse the string *aceabbceeeabeea* (figure 4).

4.2 Stage II

The second stage in the simulation results is marked by a sudden change in the population measures. The number of meanings covered increases dramatically as does the size of the grammar. More importantly, the number of meanings becomes greater than the number of

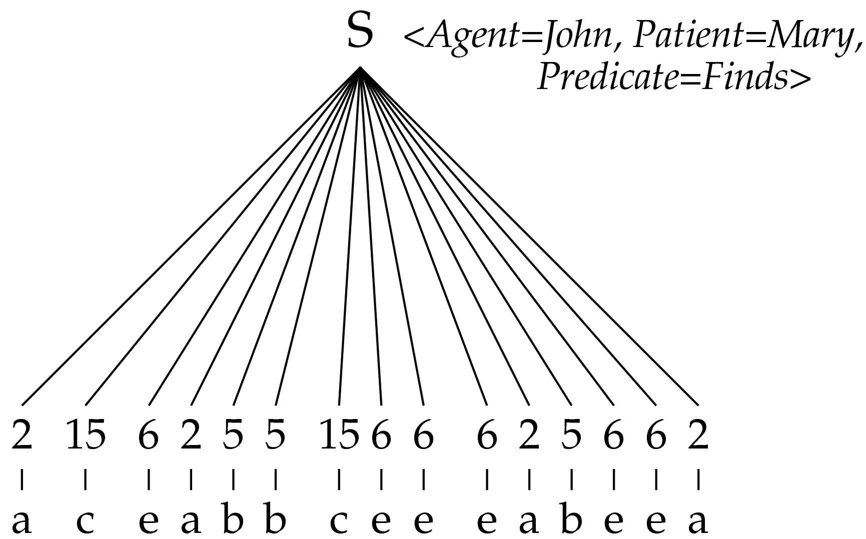

Figure 4: A **stage I** phrase structure tree showing the utterance *aceabbceeeabeea* meaning *John finds Mary*. Note the complete flatness of the structure. The numbers attached to the nodes are the actual arbitrary category labels assigned by the learning algorithm — notice (as discussed previously) that the each terminal symbol has an associated preterminal category. So, the symbol *a* has bneen arbitrarily assigned the category label 2 by the learner.

rules in the grammar. It is clear from this that the language is no longer simply behaving as a list of unanalysed vocabulary items for complex meanings as it was in **stage I**.

In fact, the grammars at this stage are far more complex and byzantine than the earlier ones. The details of what is going on in the language of the population at this stage are hard to figure out. There are, however, a few points that should be noted. Firstly, there are now syntactic categories that are intermediate between the sentence level and the level of individual symbols. Importantly, some of these intermediate categories, or *words*, have a semantics of their own. We can see this from the example tree in figure 5. Here, as we can see from this parse tree, the substring *ce* means *John* in the context of the string *dceddd*. This utterance is therefore partly compositional.

4.3 Stage III

After a second abrupt change, the population switches into a third and final stage. This stage appears to be completely stable, and in all runs no significant changes occur after this point. The transition is marked by a sudden increase of the number of meanings that can be produced to the maximum value and a drop in the size of the grammars.

A look at the behaviour of an individual in this stage reveals a marked contrast with the typical behaviour earlier in the simulation. Some of the utterances of a typical individual are shown below:

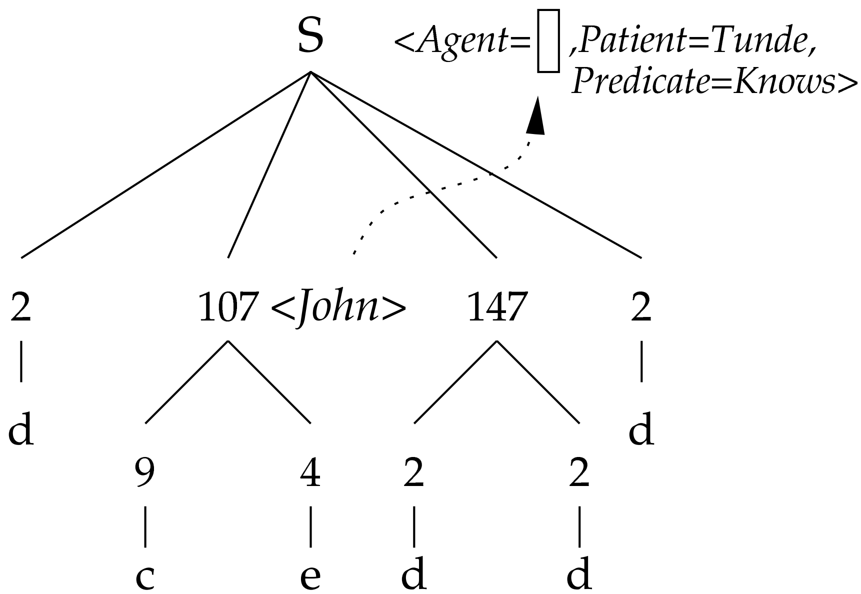

Figure 5: A **stage II** tree showing *dceddd* meaning *John knows Tünde*. The arrow shows how the meaning of the whole is partially composed from the meaning of one of its lower constituents. Notice, again that arbitrary numerical category labels have been assigned by the inducer. There are two types of label here, however. We have preterminal labels as before such as 2, 4 and 9, which stand in for the terminal symbols, but there are also intermediate categories such as 107 and 147 which begin to look more like the standard lexical categories we find in real language which rewrite to strings of preterminals.

Meaning (glossed in English)	String
John finds Mary	daecde
John finds Mike	daadde
John finds Tünde	daccde
...	...
John hates Mary	cdecde
John hates Mike	cdadde
...	...
Mary finds John	dadeec
...	...
Zoltan loves Tünde	ceccca

This individual is able to express all 100 possible meanings because there is a regular correspondence between meanings and forms. Each string is composed of three substrings which correspond to the predicate, the patient, and the agent, in that order. The table below and the example tree in figure 6 make this clearer.

Meaning	String
John	de
Mary	ec
Mike	ad
Tünde	cc
Zoltan	ca
Finds	da
Hates	cd
Knows	ee
Likes	ae
Loves	ce

Not only is this language completely compositional but, by directly inspecting the grammars of the individuals, it can be shown that the language also groups all the objects (*Mary*, *Zoltan*, *Mike*, *Tünde* and *John*) under one syntactic category (*62*) and all the actions (*Likes*, *Loves*, *Knows*, *Finds* and *Hates*) under a second category (*66*). In other words, this language encodes a classic noun/verb distinction syntactically.

The language is a VOS language in that the verb is the first word in the sentence, and the semantic roles of the two following nouns is determined by word order such that the first noun is the patient and the second is the agent. The emergent ordering differs from run to run, but the general pattern is the same: a noun/verb distinction encoded in the lexicon with the agent/patient distinction encoded by word order.[9] The eventual grammar size in this run is 15 rules. This works out as one top-level sentence rule, 10 lexical rules (one for each noun and verb), and 4 preterminal rules (one each for the symbols *a*, *c*, *d* and *e* as the symbol *b* happens not to be used in this language).

[9]Although the result of this run is full compositionality, in that the sentence rule does not add any atomic semantic content, this is not always the case. Occasionally, one of the atomic meanings does not become *lexicalised* as a noun or a verb, and an idiosyncratic sentence rule is used to express meanings that include the missing word.

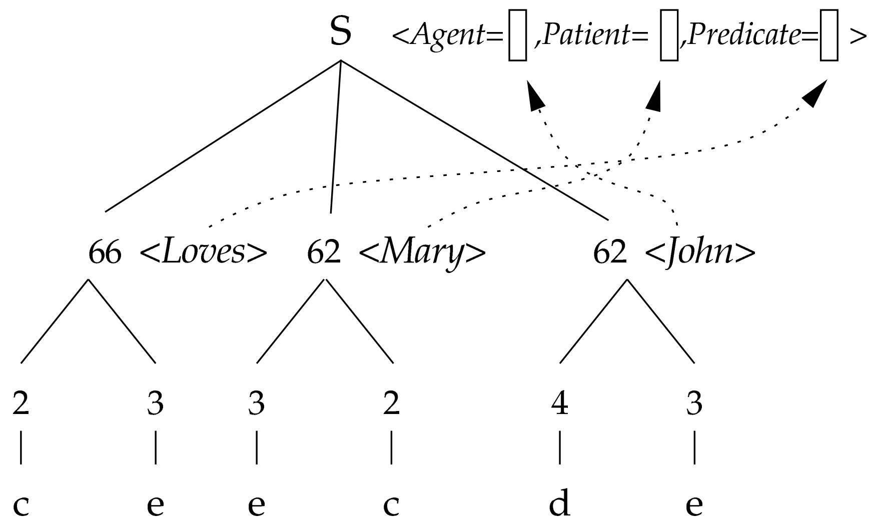

Figure 6: A **stage III** tree showing *ceecde* meaning *John loves Mary*. Here, the arbitrarily assigned category label 66 we could gloss as "verb" and the label 62 as "noun".

4.4 Summary of the results

What we have seen in this run, and in every run of the simulation that has been attempted, is the emergence of simple, yet language-like, syntax from randomness in a population that is *not* constrained to learn *only* a compositional language.

The communication system of the population that quickly emerges from nothing is an impoverished, idiosyncratic vocabulary of one-word utterances — in fact, nothing more than an inventory of calls expressing unanalysed meanings. This system is passed on only "culturally" through observational learning by new individuals, and there is nothing else inherited by later generations from earlier ones.

After many generations, the system that is used to express meanings balloons in complexity. Utterances are no longer unanalysed strings of symbols. They are made up of common chunks of several symbols. Some of these chunks even have meanings of their own, although they are not regularly used to signify these meanings in a larger context. The language of the population now goes through radical and unpredictable changes over time as the range of meanings that are readily expressible changes wildly. The language appears to be brittle in some way and liable to break and lose its expressive power suddenly.

At some point, all this changes, and the population converges on a simple system, a syntactic system. Now, every sentence is made up of nouns and verbs (drawn from a concise lexicon lacking synonymy and homonymy) in a fixed order which encodes meaning distinctions compositionally, and every possible meaning can be expressed.

5 Why does this model work?

The individuals in the simulation simply observe each others' behaviour and learn from it, occasionally inventing, at random, new behaviours of their own. From this apparent randomness, organisation emerges. Given that so little is built into the simulation, why is a

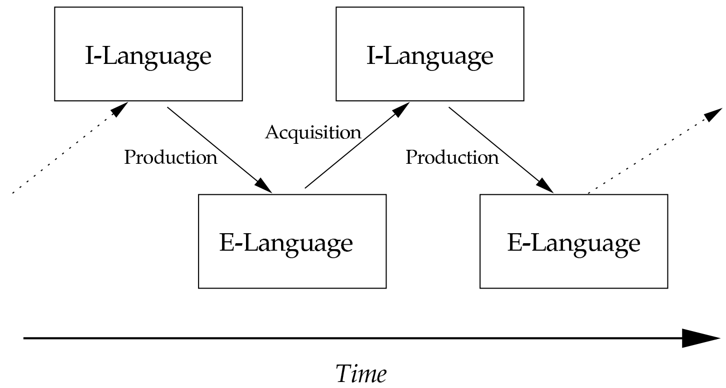

Figure 7: The cycle of language acquisition and use, which maps I-language objects to E-language objects and vice versa. These transformations act as bottlenecks for the information flowing through the system. For a particular feature of language to survive over time, it must be faithfully preserved through these mappings.

compositional syntax inevitable?

To answer this question, we need to look at how languages persist over time in the population. Language exists is two forms, both in reality and in the simulation (Chomsky 1986; Hurford 1987; Kirby 1998a):

I-language This is (internal) language as represented in the brains of the population. It is the language user's knowledge of language. In the simulation, the I-language of an individual is completely described by its grammar.

E-language This is the (external) language that exists as utterances in the arena of use (Hurford 1987). In the simulation, we can describe E-language by listing the form-meaning pairs of an individual.

These two types of language influence each other in profound ways. E-language is a product of the I-language of speakers. However, the I-language of language learners is a product of the E-language that they have access to (see figure 7). A particular I-language or E-language can fail to persist over time because the processes that map from one to the other and back again are not necessarily preservative.

We can divide up I-language into units — *replicators* — that may or may not persist through time. The persistence of an I-language over time is related to the success of the replicators that make up that language. In other words, the languages which are more easily transmitted from generation to generation will persist.

Within a population, certain replicators actually compete for survival. That is, the success of one must be measured relative to the success of others in the population at that time. These competing replicators are those rules which potentially express the same meaning. If there are two ways of saying *John loves Mary*, then on a particular exposure to this meaning, the learner can obviously only hear one of them. Therefore, on one exposure, only one of the rules (or, more properly, set of rules) that can be used to express *John loves Mary* has a chance of being induced by the learner.

At face value, it would seem that the two competing rules (or rule-sets) will have an equal chance of being the one chosen for producing the meaning, so the replicative success of all rules in a language should be equal. This would be true *if each rule only ever expressed one meaning*. However, if one rule can be used to express more meanings than another, then, all other things being equal, that rule will have a greater chance of being expressed in the E-language input to the learner. In this case, the more general rule is the better replicator.

For a more concrete example, consider a situation where, in the population of I-languages, there are two competing rules. One is a rule that expresses *John loves Mary* as an unanalysed string of symbols — essentially as one word. The other rule expresses *John loves Mary* as a string of symbols, but can also be used to express any meaning where someone *loves Mary*. So, the latter rule can also be used to express *Zoltan loves Mary* and so on. Further imagine that both rules have an equal chance of being used to express *John loves Mary*. The more general rule is still a better replicator, because for any randomly chosen set of meanings, we can expect it to be used more often than the idiosyncratic rule. Its chances of survival to the next generation are far more secure than the idiosyncratic rule.

Of course, the more general rule will not be learned as easily as the idiosyncratic rule. In the simulations described above, an idiosyncratic pairing of one meaning to one form takes only one exposure to learn, but the most general rule takes several. However, the idiosyncratic rule only covers one meaning, whereas the most general rule covers 100. It is clear, therefore, that the probability of a acquiring a particular rule given a random sample of meanings increases with the generality of that rule. The success of I-languages which contain general rules seems secure.

The picture that emerges, then, is of the language of the population acting as an adaptive system in its own right. Initially, the rules are minimally general, each pairing one string with one meaning. At some point, a chance invention or random noise will lead a learner to "go beyond the data" in making a generalisation that the previous generation had not made. This generalisation will then compete with the idiosyncratic rule(s) for the same meaning(s). Given that generalisations are better replicators, the idiosyncratic rules will be pushed out over time. The competition will then be replayed amongst generalisations, always with the more general rules surviving.

The inevitable end state of this process is a language with a syntax that supports compositionally derived semantics in a highly regular fashion. The grammar for such a language appears to be the shortest (in terms of numbers of rules) that can express the entire meaning space. The shorter the grammar, the higher the generality of each of the rules — the shortest grammar that can still do the job of expressing meanings is therefore the one made up of optimal replicators.

There is an interesting way in which this replicator-based theory can be tested using the simulation. If the emergence of compositionality is due to the differential success of competing replicators, then there should be effects introduced by changing the frequency of particular meanings. For example, if one meaning is expressed particularly frequently by speakers, any rule that contributes to the production of a string for that meaning will be a good replicator. In the simulation results presented so far, idiosyncratic rules have died out because they contribute to a relatively small portion of E-language. However, if one meaning is particularly frequent, then we should find that an idiosyncratic form for that meaning will survive longer.

To test this, the simulation was run again, but the maximum number of utterances was doubled to 200. The meaning, *John loves Mary*, was made far more frequent so that it made up approximately half of the utterances. The results of such runs are consistent with the idea that replicator dynamics are driving the evolution of language in the simulation. The pattern of change in the simulation is similar to the one described earlier, with three stages showing evolution towards compositional syntax. Even in the final stage, however, an idiosyncratic, non-compositional way of saying *John loves Mary* survived.

This mechanism — whereby frequent meanings can withstand the pressure to become compositionally expressed — may explain some features of human languages. For example, in morphology, suppletive forms tend to correlate with highly frequent meanings. The past tense form of the frequent verb, *go* is the non-compositional *went* not *goed*. The ordinal versions of the English numbers after *three* are compositional — *third, fourth, fifth* etc. — but the more frequent *first* and *second* are not.

6 Conclusion

In this paper I have argued the case for an appreciation of the role of truly linguistic evolution (as opposed to biological evolution) in the emergence of syntax. Human language is unique amongst communication systems in being compositional. It is also unique in the natural world in being a phenomena that persists over time through observational learning. These two facts are clearly connected. Once an observationally learned communication system is off the ground (see Oliphant (this volume) for discussion of why this is not a trivial problem for evolution), the dynamics introduced make the emergence of compositionality inevitable without further biological change.

References

BATALI, JOHN. 1994. Innate biases and critical periods: Combining evolution and learning in the acquisition of syntax. In *Artificial Life IV*, ed. by Rodney Brooks & Pattie Maes, 160–171. MIT Press.

——. 1997. Computational simulations of the emergence of grammar. In *Approaches to the Evolution of Language: Social and Cognitive Bases*, ed. by James Hurford, Chris Knight, & Michael Studdert-Kennedy. Cambridge University Press. In press.

BICKERTON, DEREK. 1990. *Language and Species*. University of Chicago Press.

BRISCOE, TED, 1997. Language acquisition: the bioprogram hypothesis and the Baldwin Effect. MS, Computer Laboratory, University of Cambridge.

CANGELOSI, ANGELO, & DOMENICO PARISI. 1996. The emergence of a language in an evolving population of neural networks. Technical Report NSAL–96004, National Research Council, Rome.

CANN, RONNIE. 1993. *Formal Semantics: An introduction*. Cambridge: Cambridge University Press.

CHOMSKY, NOAM. 1986. *Knowledge of Language*. Praeger.

CHRISTIANSEN, MORTEN, 1994. *Infinite Languages, Finite Minds: Connectionism, Learning and Linguistic Structure*. University of Edinburgh dissertation.

DEACON, TERRENCE W. 1997. *The Symbolic Species: The co-evolution of language and the brain*. New York: Norton.

GELL-MANN, MURRAY. 1992. Complexity and complex adaptive systems. In *The Evolution of Human Languages*, ed. by J.A. Hawkins & M. Gell-Mann. Addison-Wesley.

HURFORD, JAMES. 1987. *Language and Number: the Emergence of a Cognitive System*. Cambridge, MA: Basil Blackwell.

——. 1989. Biological evolution of the Saussurean sign as a component of the language acquisition device. *Lingua* 77.187–222.

——. 1991. The evolution of the critical period for language acquisition. *Cognition* 40.159–201.

——. 1998. The evolution of language and languages. In *The Evolution of Culture*, ed. by Chris Knight Robin Dunbar & Camilla Power. Edinburgh University Press. In press.

——, CHRIS KNIGHT, & MICHAEL STUDDERT-KENNEDY (eds.) 1998. *Approaches to the Evolution of Language: Social and Cognitive Bases*, Cambridge. Cambridge University Press.

KIRBY, SIMON. 1997a. Competing motivations and emergence: explaining implicational hierarchies. *Language Typology* 1.5–32.

——. 1997b. Fitness and the selective adaptation of language. In *Approaches to the Evolution of Language: Social and cognitive bases*, ed. by J. Hurford, C. Knight, & M. Studdert-Kennedy. Cambridge University Press. In press.

——. 1998a. *Function, Selection and Innateness: The Emergence of Language Universals*. Oxford: Oxford University Press. In press.

——. 1998b. Language evolution without natural selection: From vocabulary to syntax in a population of learners. Technical Report EOPL–98–1, Department of Linguistics, University of Edinburgh.

——. 1999. Learning, bottlenecks, and the evolution of recursive syntax. In *Linguistic evolution through language acquisition: formal and computational models*, ed. by Ted Briscoe. Cambridge: Cambridge University Press. Forthcoming.

——, & JAMES HURFORD. 1997. Learning, culture and evolution in the origin of linguistic constraints. In *Fourth European Conference on Artificial Life*, 493–502. MIT Press.

MACLENNAN, BRUCE. 1991. Synthetic ethology: an approach to the study of communication. In *Artificial Life II*, ed. by C.G. Langton, C. Taylor, J.D. Farmer, & S. Ramussen, 631–657. Addison-Wesley.

NEWMEYER, FREDERICK J. 1991. Functional explanation in linguistics and the origins of language. *Language and Communication* 11.3–28.

—— 1992. Iconicity and generative grammar. *Language* 68.756–796.

NIYOGI, PARTHA, & ROBERT BERWICK, 1997. Populations of learners: the case of Portugese. Unpublished manuscript, MIT.

OLIPHANT, MICHAEL. 1996. The dilemma of saussurean communication. *BioSystems* 37.31–38.

——, 1997. *Formal Approaches to Innate and Learned Communication: Laying the Foundation for Language*. UCSD dissertation.

OSBORNE, MILES, & TED BRISCOE. 1997. Learning stochastic categorial grammars. Technical report, Computer Laboratory, Cambridge University.

PINKER, STEVEN, & PAUL BLOOM. 1990. Natural language and natural selection. *Behavioral and Brain Sciences* 13.707–784.

STEELS, LUC. 1996. Emergent adaptive lexicons. In *Proceedings of the Simulation of Adaptive Behavior Conference*, ed. by P. Maes. Cambridge, Ma: The MIT Press.

STOLCKE, ANDREAS, 1994. *Bayesian Learning of Probabilistic Language Models*. University of California at Berkeley dissertation.

WRAY, ALISON. 1998. Protolanguage as a holistic system for social interaction. *Language and Communication* 18.47–67.