From UG to Universals

Linguistic adaptation through iterated learning

Simon Kirby, Kenny Smith and Henry Brighton
University of Edinburgh

What constitutes linguistic evidence for Universal Grammar (UG)? The principal approach to this question equates UG on the one hand with language universals on the other. Parsimonious and general characterizations of linguistic variation are assumed to uncover features of UG. This paper reviews a recently developed evolutionary approach to language that casts doubt on this assumption: the Iterated Learning Model (ILM). We treat UG as a model of our prior learning bias, and consider how languages may adapt in response to this bias. By dealing directly with populations of linguistic agents, the ILM allows us to study the adaptive landscape that particular learning biases result in. The key result from this work is that the relationship between UG and language structure is non-trivial.

1. Introduction

A fundamental goal for linguistics is to understand why languages are the way they are and not some other way. In other words, we seek to explain the particular universal properties of human language. This requires both a characterisation of what these universals are, and an account of what determines the specific nature of these universals. In this paper we examine a particular strategy for linguistic explanation, one which makes a direct link between language universals and an innate Universal Grammar (UG). It seems reasonable to assume that, if UG determines language universals, then language universals can be used as evidence for the structure of UG. However, we will argue that this assumption is potentially dangerous. Our central message is that we can seek linguistic evidence for UG only if we have a clear understanding of the mechanisms that link properties of language acquisition on the one hand and language universals on the other.

In the following section we will discuss what is actually meant by the term UG. There are a number of differing senses of the term, but a neutral definition can be given in terms of *prior learning bias.* We will then sketch an account of the universal properties of language in terms of this bias.

Studies in Language 28:3 (2004), 587–607.
ISSN 0378–4177 / E-ISSN 1569–9978 © John Benjamins Publishing Company

In Section 3, we will compare this kind of explanation to an alternative approach, linguistic functionalism, which focuses on the use of language. A well-recognised difficulty with this approach is the *problem of linkage*: what is the mechanism that links universals to linguistic functions? We claim that not only does the UG-approach suffer exactly the same problem, but the solution is the same in both cases.

Section 4 sets out this solution in terms of *Iterated Learning*, an idealised model of the process of linguistic transmission. We survey some of the results of modelling iterated learning to show how it can help solve the problem of linkage.

Finally, in the closing sections of the paper we argue that language universals, and linguistic structure more generally, should be viewed as adaptations that arise from the fundamentally *evolutionary* nature of linguistic transmission.

2. What is Universal Grammar?

Before we discuss the role of UG in explaining language universals, we need to be clear what we mean. Unfortunately, there is some variation in how the term is used (see Jackendoff 2002 for an excellent review of the literature):

i. *UG as the features that all languages have in common.*
Clearly, this equates UG exactly with universals. This is not the sense of UG that we will be concerning ourselves with in this paper. Initially, it may seem absurd to imply that a characterisation of UG in this sense could possibly be an *explanation* of the universal characteristics of human language. Rather, it may appear only to be a description of the properties of language. However, we should be careful about dismissing the explanatory significance of a theory of UG that 'merely' sets out the constraints on cross-linguistic variation.

In fact, it is conceivable that a truly explanatory theory of language *could* consist of an account of UG in this sense. Chomsky (2002) gives an illuminating analogy that makes clear there is more than one way to explanatory adequacy in science. Consider, he suggests, the case of the discovery of the Periodic Table in late 19th century chemistry. To simplify somewhat, chemists, through careful experimental observations of the elements, were able to uncover a range of regularities that made sense of the behaviour of those elements. Repeating, periodic patterns could be seen if the elements were arranged in a particular way — approximately, as a table made up of rows of a fixed length.

In one sense we could see the periodic table as being merely a description of the behaviour of matter. We could claim that the discovery of the periodic table does nothing to explain the mass of experimental data that chemists have collected. This seems wrong. Surely such a concise and elegant generalisation is, in some sense, explanatory. See Eckman (this special issue) for an extended discussion of the relationship between generalisation and explanation. The periodic table itself can now be explained by physicists with reference to more fundamental constituents of matter, but this does not alter the status of the table in chemistry itself.

Are linguists in the process of discovering an equivalent of the periodic table? Is there a model of UG 'out there' that has the same combination of formal simplicity and predictive power? It is a worthy research goal, and one that is being pursued by many, but we may be chasing phantoms. As we will argue in this paper, UG should be considered as only part of an explanatory framework for language.

ii. *UG as the initial state of the language learning child.*
This sense of UG is very closely related to the previous sense. Jackendoff (2002) notes that Chomsky (1972) uses the term UG to denote the configuration of a language-ready child's brain that sets the stage for language acquisition. This 'state-zero' can, in fact, be thought of as specifying the complete range of possible grammars from which a maturation process 'picks' a target grammar in response to linguistic data. It is natural to equate the space of languages specified in state-zero with the range of possible languages characterised by language universals. The main difference between this sense of UG and the previous one is that it gives UG an explicit psychological reality.

iii. *UG as initial state* **and** *Language Acquisition Device.*
Jackendoff (2002) points out that in its most common usage, UG is taken to correspond to the knowledge of language that the child is born with. This consists not only of the initial state, but also the machinery to move from this state to the final target grammar. Chomsky refers to this machinery as the *Language Acquisition Device* or LAD. For convenience, we will consider this device to encapsulate the initial state as well as the machinery of acquisition. This means that we will treat this sense of UG as simply a description of the LAD.

In summary, there are a number of different ways we can think about what Universal Grammar actually is. This may seem like terminological confusion, but really all these senses have something fundamental in common: they all appear to relate UG directly with universals. The different senses we have surveyed differ primarily with respect to how UG is situated in a wider theory of cognition. The picture is something like the one shown in Figure 1. The broadly Chomskyan program for linguistics is to uncover the properties of UG. Since UG and language universals are coextensive, then the evidence for UG can be derived directly from a careful characterisation of the (universal) properties of linguistic structure.

A sensible question is how we can characterise UG/LAD in such a way that there is a clear relationship between the theory and constraints on linguistic variation (i.e., universals). Various approaches are possible. For example, in Principles and Parameters theory (Chomsky 1981) there is a direct relationship between cross-linguistic parametric variation and the elements of the model, parameters, that are set in response to input data.[1] Similarly, in Optimality Theory

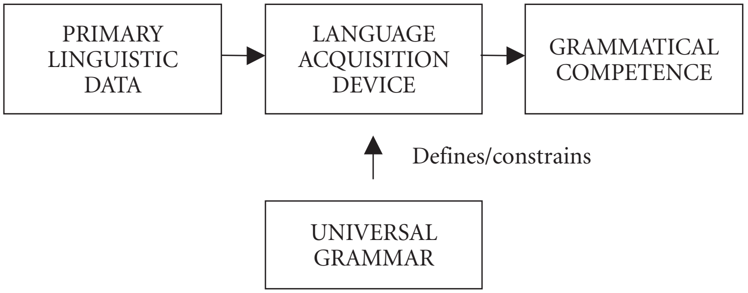

Figure 1. The language acquisition device (LAD) takes primary linguistic data and generates the adult grammatical competence of a language. Universal grammar defines or constrains the operation of the LAD.

(Grimshaw 1997) variation arises from the constraint ranking that is arrived at through the acquisition process.

The literature on machine learning (e.g., Mitchell 1997) suggests a general way of characterising the relationship between language learning and linguistic variation. We can think of the learning task for language to be the identification of the most probable grammar that generates the data observed. More formally, given a set of data D and a space of hypotheses about the target grammar H, we wish to pick the hypothesis $h \in H$ that maximises the probability $\Pr(h \mid D)$, in other words, the probability of h given D. From Bayes law, we have:

$$\Pr(h \mid D) = \frac{\Pr(D \mid h)\Pr(h)}{\Pr(D)}$$

The task of the learner is to find:

$$\arg\max_{h \in H} \Pr(h \mid D) = \arg\max_{h \in H} \Pr(D \mid h)\Pr(h)$$

(We can ignore the term $\Pr(D)$ since this is constant for all hypotheses).

What is the contribution of UG/LAD in this framework? It is simply the *prior bias* of the learner. This bias is everything[2] that the learner brings to the task *independent of the data.* In other words, it is the probability $\Pr(h)$ assigned to each hypothesis $h \in H$.

One of the interesting things about this Bayesian formulation is that it allows us to see the classic problem of induction in a new light (Li & Vitanyi 1993). Consider what a completely 'general purpose' learner would look like. Such a learner would not be biased *a priori* in favour of any one hypothesis over another. In other words, $\Pr(h)$ would be equal for all hypotheses. Such a learner would then simply pick the hypothesis that maximised $\Pr(D \mid h)$. In other words, the best a learner can do is pick the hypothesis that recreates the data exactly. Such a learner cannot, therefore, generalise. Since language learning involves generalisation, then *any* theory of language learning must have a model of prior bias. Where does this prior bias come from? An obvious answer is that it is innate.

Note, however, that we have said nothing about *domain specificity*. It is crucial that the issues of innateness and domain specificity are kept separate. It is a fascinating but difficult challenge to discover which features of the child's prior bias (if any) are there *for* language. We note here only that an approach to this problem must be based on a theory of the relationship between the structure of innate mechanisms and the functions to which they are put (e.g., language learning). In other words, answers to questions about domain-specificity will come from a better understanding of the biological evolution of the human brain.

To summarise, a central goal for linguistics is to discover the properties of UG. We argue that, in general, this amounts to a characterisation of the prior learning bias that children bring to bear on the task of language acquisition. Since it is Universal Grammar that leads to universal properties of human languages, a sensible strategy seems to be to use observable properties of languages to infer the content of UG.

In the next section, we will show that this argument suffers from a problem that has been identified with a quite different approach to linguistic explanation: functionalism.

3. The problem of linkage

The functionalist approach to explaining language universals (see e.g., Hawkins 1988) seems at first blush to be incompatible with explanations that appeal to UG. A functionalist explanation for some aspect of language structure will relate it to some feature of language use. This runs completely counter to the generativist program, which focuses on explaining linguistic structure on its own terms, explicitly denying a place for language use 'inside' a theory of UG. If chemistry is a good analogy for the generativist enterprise, then perhaps biology is the equivalent for functionalists. The central idea is that we can only make sense of structure in light of an understanding of what it is used for. (See Newmeyer (1998) and Kirby (1999) for further discussion of functionalism and the generativist tradition.)

A particularly ambitious attempt to explain a wide range of data in terms of language use is Hawkins' (1994) processing theory. Hawkins' main target is an explanation of the universal patterns of word-order variation. For example, he notes that there is a constraint on possible ordering in noun-phrases — a universal he calls the *prepositional noun-modifier hierarchy*: In prepositional languages, within the noun-phrase, if the noun precedes the adjective, then the noun precedes the genitive. Furthermore, if the noun precedes the genitive, then the noun precedes the relative clause.

This hierarchy predicts that, if a language has structure *n* in the following list, then it will have all structures less than *n*:

1. $_{PP}[P\ _{NP}[N\ S']]$
2. $_{PP}[P\ _{NP}[N\ NP]]$
3. $_{PP}[P\ _{NP}[N\ Adj]]$

Hawkins' explanation rests on the idea that when processing such structures, stress on our working memory increases as the distance between the preposition and the noun increases. He argues that the NP node in the parse-tree is only constructed once the head noun is processed. This means that the immediate daughters of the PP are only available for attachment to the PP node when both the preposition and noun have been heard. Since relative clauses are typically longer than noun-phrases, which are usually longer than adjectives, the difficulty in processing each of these structures increases down the list.

Assuming this account is correct, does the relative processing difficulty of each structure actually explain the language universal? Kirby (1999) points out that the identification of a processing asymmetry that corresponds to an asymmetry in the distribution of languages is not quite enough to count as an explanation. What is missing is something to connect working-memory on the one hand with numbers of languages in the world on the other.

> *The problem of linkage:* Given a set of observed constraints on cross-linguistic variation, and a corresponding pattern of functional preference, an explanation of this fit will solve the problem: how does the latter give rise to the former? (Kirby 1999: 20)

Kirby (1999) sets out an agent-based model of linguistic transmission to tackle this problem. Agent-based modelling is a computational simulation technique used extensively in the field of artificial life (see Kirby 2002b for a review of the way this field has approached language evolution). 'Agents' in these simulations are simple, idealised models of individuals, in this case language users. The details of the simulation are not important, but the basic idea is that variant word-orders are transmitted over time from agent to agent through a cycle of production, parsing, and acquisition. In the simulations, different word-order variants appear to compete for survival, with universal patterns of cross-linguistic variation emerging out of this competition.

These models show that for some functional explanations, processing asymmetries do indeed result in equivalent language universals. However, this is not always the case. In general, hierarchical universals cannot be explained using only one set of functional asymmetries. The particular details are not relevant here,[3] but the moral should be clear: without an explicit mechanism linking *explanans* and *explanandum* we cannot be sure that the explanation really works.

At this point we might ask what relevance this discussion has for the generative type of explanation, which treats language universals as being encoded in Universal Grammar. In fact, we would argue that there is very little difference between these two modes of explanation, and as such the same problem of linkage applies.

In Hawkins' approach to functional explanation, a direct link is made between a feature of the language user's psychology (such as working memory) and the universal properties of language. Similarly, the generative approach makes a direct link between another feature of the language user's psychology (this time, learning bias) and language universals.

The problem of linkage holds for both functionalist and generative explanations for language universals. In the next section, we look at a development of the model put forward in Kirby (1999) that demonstrates the rather subtle connections between language learning and language structure arising out of the process of linguistic transmission.

4. Iterated learning

Over the last few years there has been a growing interest in modelling a type of cultural information transmission we call *Iterated Learning* (Kirby & Hurford 2002). The central idea underlying the iterated learning framework is that behaviour can be transmitted culturally by agents learning from other agents' behaviour *which was itself the result of the same learning process.* Human language is an obvious example of a behaviour that is transmitted through iterated learning.[4] The linguistic behaviour that an individual exhibits is both a result of exposure to the behaviour of others *and* a source of data that other learners may be exposed to.

The Iterated Learning Model (ILM) gives us a tool with which we can explore the properties of systems that are transmitted in this way. In this section we will briefly review some of the ways the ILM has been used to look at systems for mapping meanings to signals that are transmitted through repeated learning and use. The main message we hope to convey is that the relationship between learning and the structure of what is being learned is non-trivial. Hence, when we look at the 'real' system of human language, we should expect the relationship between UG and universals to be similarly complex.

A simple ILM

Consider a system where there are a number of meanings that agents want to express. They are able to do this by drawing on a set of possible signals. The way in which they relate signals and meanings is by using some internal grammar. The means by which they arrive at this grammar is through observation of particular instances of other agents' expression of meanings.

We can imagine systems like this with large populations of agents interacting and learning from each other, with the possibility for various kinds of population turnover (i.e., how the population changes over time). The simplest possible population model is shown in Figure 2. Here there are only two agents at any one time: an adult and a learner. The adult will be prompted with a randomly chosen meaning and, using its grammar, will generate a signal. This signal-meaning pair will then form part of the input data to the learner. From a set of the adult's signal-meaning pairs (the size of the set being a parameter in the simulation) the learner will try and induce the adult's grammar.

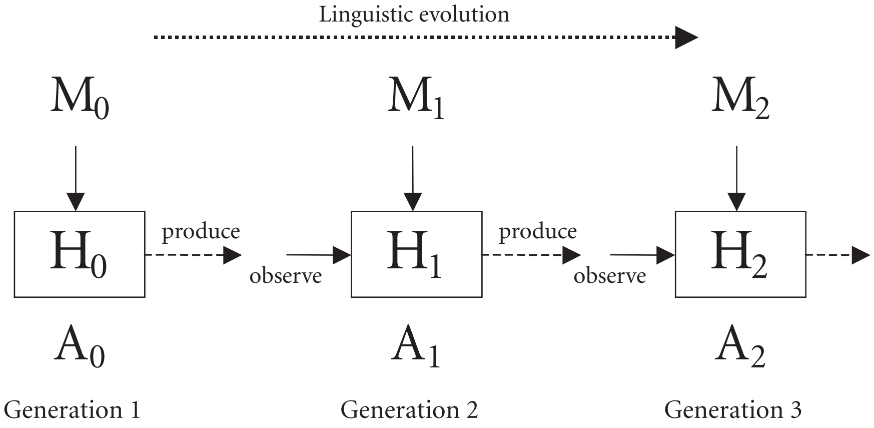

Figure 2. A simple population model for iterated learning. Each generation has only one agent, *A*. This agent observes utterances produced by the previous generation's agent. The learner forms a hypothesis, *H*, based on these utterances. In other words, the agent aims to acquire the same language as that of the previous generation. Prompted by a random set of meanings, *M*, this agent goes on to produce new utterances for the learner in the next generation. Note that, crucially, these utterances will not simply be a reiteration of those the agent has heard because the particular meanings chosen will not be the same.

We are interested in what happens when a language (conceived of as a mapping between meanings and signals) is transmitted in this way. Will the language change? If so, are there any stable states? What do stable languages look like and what determines their stability?

Ultimately, we can only begin to find answers to these questions by actually implementing the ILM in simulation. To do this we need to implement a model agent, decide what the set of meanings and signals will look like, and also the structure and dynamics of the population. In an ILM, the particular learning algorithm used determines the prior bias of the agents. We can think of the learning algorithm as essentially a model of UG. A wide range of designs of ILM simulations have been employed in the literature. The following is a partial list (there has also been further work looking at models of language that do not treat it as a mapping from meanings to signals, such as Jäger 2003, Teal & Taylor 1999, and Zuidema 2001):

- (Batali 1998). Models a population of simple recurrent networks (Elman 1990). Meanings are bit-vectors with some internal structure. There is no population turnover in this simulation.
- (Kirby 2000). Agents learn using a heuristically-driven grammar inducer. Meanings are simple feature-structures, and the population has gradual turnover.
- (Kirby 2002a). Similar learning algorithms, but with recursively structured meaning representation. Described in more detail below.

- (Kirby 2001). Same learning algorithm. Meanings are coordinates in a two-dimensional space, with a non-uniform frequency distribution.
- (Batali 2002). Population of agents using instance-based learning techniques. Meanings are flat lists of predicates with argument variables.
- (Brighton & Kirby 2001). Agents acquire a form of finite-state transducer using Minimum Description Length learning. Many runs of the simulation are carried out with different meaning-spaces.
- (Tonkes 2002). Along with a number of other models, Tonkes implements an ILM with a population of simple recurrent networks with a continuous meaning space (each meaning is a number between 0.0 and 1.0).
- (Smith, Brighton & Kirby, forthcoming). Uses associative networks to map between strings and feature-vectors.
- (Vogt 2003). Implements a simulation of a robotics experiment — the 'Talking Heads' model of Steels (1999). The agents communicate about objects of various shapes, colours and locations. This is part of a broader research effort to get round the problem that the ILM requires a pre-existing model of meanings. By grounding the ILM in a real environment, both signals *and* meanings can be seen to emerge.

These simulations are typically seeded with an initial population that behaves randomly — in other words, agents simply invent random signals (usually strings of characters) for each meaning that they wish to produce. This idiosyncratic, unstructured language is learned by other agents as they are exposed to these utterances, and in turn these learners go on to produce utterances based on their own experience. The remarkable thing is that, despite their very different approaches to modelling learning (i.e., models of UG) the same kind of behaviour is seen in all these models. The initial random language is highly unstable and changes rapidly, but over time stability begins to increase and some structure in the mapping between meanings and signals emerges. Eventually, a stable language evolves in which something like syntactic structure is apparent.

For example, Kirby (2002a) uses the ILM to explore how recursive compositionality could have evolved. In this model, the population structure is as in Figure 2. The agents' model of language is represented as a form of context-free grammar, and a heuristic-based induction algorithm is used to acquire a grammar from a set of example utterances. The signals are simply strings of characters, and the meanings take the form of simple predicate logic expressions. (This is not the place to go into the technical details of the model — these are given in the original article.)

Here are a few of the sentences produced by an agent early on in the simulation run. The meaning of each sentence is glossed in English. (Note that the letters that make up these strings are chosen at random — there is no role for phonetics or phonology in this simulation):

(1) *ldg*
'Mary admires John'

(2) *xkq*
'Mary loves John'

(3) *gj*
'Mary admires Gavin'

(4) *axk*
'John admires Gavin'

(5) *gb*
'John knows that Mary knows that John admires Gavin'

In this early stage, the language of the population is unstructured. Each meaning is simply given a completely idiosyncratic, unstructured string of symbols. There is no compositionality or recursion here, and it is better to think of the language as a vocabulary where a word for every possible meaning has to be individually listed.

This type of syntax-free language, which Wray (1998) refers to as a *holistic protolanguage*, may have been a very early stage in the evolution of human language. It can be compared with animal communication systems inasmuch as they typically exhibit no compositional structure.[5] Wray suggests that *living fossils* of this protolanguage still exist today in our use of formulaic utterances and holistic processing.

The hallmark of these early languages in the ILM is instability. The pairing of meanings and strings changes rapidly and as a result the communicative ability of the agents is poor. It is easy to see why this is. The learners are only exposed to a subset of the range of possible meanings (which, strictly speaking, are infinite in this model because the meanings are defined recursively). This means each learner can only accurately reproduce the language of the adult for meanings that it has seen. Given the five sentences listed above, how would you generalise to another meaning, say 'Mary loves Gavin'? The best you could do would be to either say nothing, or produce a string of random syllables of approximately the same length as the ones you have seen. This is precisely the challenge agents early in the simulation are faced with (although the number of sentences they are exposed to is much higher).

Thousands of generations later, however, and the language looks very different (note that the speakers do not actually generate spaces within the signals — these are included here for clarity only):

(6) *gj h f tej m*
John Mary admires
'Mary admires John'

(7) *gj h f tej wp*
John Mary loves
'Mary loves John'

(8) *gj qp f tej m*
Gavin Mary admires
'Mary admires Gavin'

(9) *gj qp f h m*
Gavin John admires
'John admires Gavin'

(10) *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'

This is clearly a compositional language. The meaning of the whole string is a function of the meanings of parts of the string. The compositional structure is also recursive as can be seen in the last example. What is interesting is that this language is completely stable. It is successfully learned by generation after generation of agents. The grammar of this language is also completely expressive. There is perfect communication between agents.

Again, it is easy to see why this is so. If you were asked the same question as before — how to express the meaning 'Mary loves Gavin' — you would probably give the answer *gjqpftejwp*. What is happening here is that you, just like the agents, are able to generalise successfully from this small sample of sentences, by uncovering substrings that refer to individual meanings, and ways to put these substrings together. There is no need for recourse to random invention. Because of this, the language is stable. All agents will (barring some unfortunate training set) converge on the same set of generalisations. They will all be able to communicate successfully about the full range of meanings (which are infinite in this case).

To summarise: in the ILM, not all languages are equally stable. A language's stability is directly related to its *generalisability*. If the language is such that generalisation to unseen meanings is difficult, then noise will be introduced to the transmission process. A crucial feature of the process of iterated learning is that if a learner makes a generalisation, even if that is an over-generalisation, the utterances that the learner produces will themselves be evidence for that generalisation. In other words, generalisations propagate. As the language comes to exhibit more and more generalisability, the level of noise in the transmission process declines, leading finally to a completely stable and highly regular linguistic system.[6] A similar process is seen in every simulation run although the particular words used, and their word-order is different each time.

It is important to realise that this is not an idiosyncratic feature of this particular model. For example, with a quite different learning model (simple recurrent networks), meaning space (bit-vectors), and population model, Batali (1998) also observed a similar movement from unstructured holism to regular compositionality. There seems to be a universal principle at work here. As Hurford (2000) puts it, social transmission favours linguistic generalisation.

There appear to be two crucial parameters in these models that determine the types of language that are stable through iterated learning: the size of the training set the learners are exposed to, and the structure of the space of possible meanings.

Hurford (2002) refers to the size of the training data as the 'bottleneck' on linguistic transmission. The bottleneck is the expected proportion of the space of possible meanings that the learners will be exposed to. When the bottleneck is too tight, no language is stable — the learners do not have enough data to reconstruct even a perfectly compositional system. If, on the other hand, the bottleneck is very wide then unstructured, holistic languages are as stable as compositional ones. This is because there is no pressure to generalise.

It is possible in these models to vary the frequency with which different meanings are expressed. This means that the bottleneck will not be equal for all meanings. In this case, we should expect frequent meanings to tend to exhibit less regularity than infrequent ones — a result that matches what we find in the morphology of many languages. This result is exactly what we find in simulation (Kirby 2001) which confirms the central role of the bottleneck in driving the evolution of linguistic structure.

The result also demonstrates that the particular choice of meanings that the agents are to communicate about is important.[7] Brighton (2002) examines the relationship between stability in the ILM and the particular structure of each meaning. In this study, meanings are treated as feature vectors. Different results are obtained depending on the number of features and the number of values each feature can take. Using both simulation and mathematical models of the iterated learning process, the relationship between feature structure and the relative stability of compositional languages can be determined. This approach is extended by Smith (2003) in a set of simulations where only some meanings are actually used by the agents. In both cases it can be shown that there is a complex relationship between meanings and the types of language that will emerge. The broad conclusion that can be drawn is that compositional structure evolves when the environment is richly structured and the meanings that the agents communicate about reflect this structure.

This work on iterated learning is at a very early stage. There is a huge gulf between the elements of these models and their real counterparts. Obviously, neither feature vectors nor simple predicate logic formulae are particularly realistic models of how we see the world. The learning algorithms the agents use and their internal representations of linguistic knowledge are not adequate for capturing the rich structure of real human languages. Does this render the results of the modelling work irrelevant?

Unsurprisingly, we would argue to the contrary. Just as simulation modelling has proved invaluable in psycholinguistics and cognitive science more generally (Elman et al. 1996), we feel that it can be used as a way of testing hypotheses about the relationship between individuals, the environment, and language universals. We know that language is transmitted over time through a process of iterated learning, but as yet we do not have a complete understanding of what this implies. We gain insights from idealised models which can be brought to bear on fundamental questions in linguistics.

In this section, we have put forward a general solution to the problem of linkage. UG, instantiated in individuals as prior learning bias, impacts on the transmission of language through iterated learning. This results in a dynamical system — some languages are inherently unstable and communicatively dysfunctional. These could never be viable human languages. Nevertheless, this fact may not be recoverable purely through examination of the biases of the learner. In other words, universals (such as compositionality) are derived in part by prior learning biases, but are not built into the learner directly. Through the iterated learning process, these languages evolve towards regions of relative stability in this dynamic landscape. The implication is clear: UG and universals cannot be directly equated. Rather, the connection is mediated by the dynamics of iterated learning. From this we can conclude that we must be very cautious in setting out a theory of UG on the basis of the observed structure of human languages — we may unwittingly be setting up a situation that results in a hidden prediction of other universals. In general, the languages that are stable through iterated learning will be a subset of those that appear to be predicted by the model of learning used.

5. Universals as emergent adaptations

The models we described in the previous section looked at how recursive compositionality, perhaps the most fundamental property of human language, can evolve through the iterated learning process. Why does this happen, and can this result help us understand language universals more generally? Earlier, we discussed what happens in the ILM from the point of view of the learner, but to answer these questions it helps to take a quite different perspective.

We are used to thinking about language from the individual's point of view. For example, we are keen to understand what the structure of the language acquisition mechanism needs to be in order for children to acquire language. Similarly, we think about language processing in terms of a challenge posed to the user of language. For many linguists, implicit in this thinking is the view that humans are adapted to the task of acquiring and using language. If language is the problem, individual human psychology is the solution.

What if we turn this round? In the context of iterated learning, it is *languages* not language users that are adapting.

Let us imagine some linguistic rule, or set of rules, that mediates the mapping between a set of meanings and their corresponding signals. For that rule to survive through iterated learning it must be repeatedly used and acquired. Consider first the case of an early-stage holistic language. Here, each rule in the language covers only a single meaning. In the example given in the last section, there was a rule that maps the meaning for 'Mary loves John' onto the string *xkq*. That is all the rule does, it is not involved in any other points in the meaning-space. For this rule to survive into the next generation, a learner must hear it being used to express 'Mary loves John'.

Now we consider the case of the perfectly compositional language. Here things are more complex because there are a number of rules used to map the meaning 'Mary loves John' onto the string *gjhftejwp*. However, the important point is that all of these rules are used in the expression of many more than this single meaning. These rules therefore produce more evidence for themselves than the idiosyncratic rule in the previous example.

The challenge for rules or regularities in a language is to survive being repeatedly squeezed through the transmission bottleneck. As Deacon (1997) puts it, “language structures that are poorly adapted to this niche simply will not persist for long” (p. 110). To put it simply, sets of rules that have general, rather than specific, application are better adapted to this challenge. In this case, recursive compositionality is a linguistic adaptation to iterated learning.

In this view, language universals can be seen as adaptations that emerge from the process of linguistic transmission. They are adaptive with respect to the primary pressure on language itself — its successful social transmission from individual to individual. Taking this perspective on the structure of language shows how compatible the generativist and functionalist approaches actually are. Figure 3 shows how adapting to innate learning bias is only one of the many problems language faces. Every step in the chain that links the speaker’s knowledge of language to the hearer’s knowledge of language will impact on the set of viable, stable human languages (see, for example, Kirby & Hurford 1997 for a model that combines processing pressures and a parameter-setting learner).

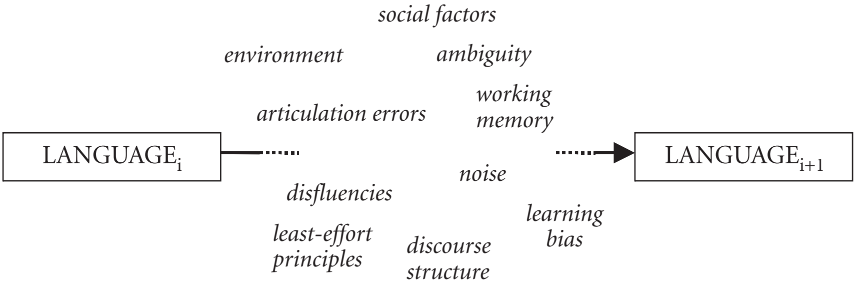

Figure 3. Many factors impinge on linguistic transmission. Language adapts in response to these pressures.

Indeed, there may be cases where the boundary between explanations based on acquisition, and explanations based on processing is very hard to draw. We mentioned Hawkins’ (1994) approach to word-order universals in Section 2. This has been applied to the general universal tendency for languages to order their heads consistently at either left-edge or right-edge of phrases throughout their grammars. This is argued to reflect a preference of the parser to keep the overall distance between heads as short as possible to reduce working-memory load. Kirby (1999) implements this preference in a very simple ILM of the transmission of word-order variants to show how the head-ordering universal emerges.

It seems clear in this case that we are talking about a quintessentially *functionalist* explanation — an explanation couched in terms of the use of language. However, Christiansen & Devlin (1997) explain the same facts in terms of language *learning*, using a general model of sequential learning: the Simple Recurrent Network (Elman 1990). The networks exhibit errors in learning in precisely those languages that are rare cross-linguistically. This seems a completely different explanation to Hawkins'. But do we really know what it is that causes the network errors? To test how well these networks have learned a language, the experimenter must give them example sentences to process. As a result, we do not know if the problem with the languages exhibiting unusual word-order arises from processing or acquisition. Perhaps we should collapse this distinction entirely. In some sense, when we acquire language we are acquiring an ability to use that language.[8]

The purpose of this discussion is to show that the distinction between functionalist approaches to typology and generativist explanations of language structure is not as clear as it might appear. UG and language function both play a role rather like the environment of adaptation does in evolutionary biology. Natural selection predicts that organisms will be *fit*. They will show the appearance of being designed for successful survival and replication. Similarly, linguistic structure will reflect properties of the bottleneck in linguistic transmission.

Once this analogy is made, it is tempting to try and apply it further. Could we explain the emergence of linguistic structure in terms of a kind of natural selection applied to cultural evolution? There have been many attempts to do just this both in general (Blackmore 1999) and in the case of language (e.g., Croft 2000; and Kirby 1999). We would like to sound a note of caution, however. There are important differences between iterated learning and biological replication (see Figure 4). In biology, there is direct copying of genetic material during reproduction. The central dogma of molecular biology (Crick 1970) states that transformation from DNA to organism is one-way only. In iterated learning, however, there is repeated transformation from internal representation to external behaviour *and back again*. The function of learning is to try and reconstruct the other agent's internal representation on the basis of their behaviour. This disanalogy with the process of selective replication in biology must be taken into account in any theory of linguistic transmission based on selection. This is not to say that an explanatory model that utilises selection is impossible. Much depends on exactly how the model is formulated. For example, Croft's (2000) model focuses on the replication of *constructions* (as opposed to induced grammatical competence). By identifying the construction as the locus of replication, Croft's model has a more natural selectionist interpretation.

A final comment should be made about the notion of adaptation we are appealing to. The simulations discussed in the previous section exhibited a universal tendency for a movement from inexpressive holistic languages to maximally expressive compositional ones. It is obvious that agents at the end of the simulation are capable of far more successful communication than those early on. In some models they are capable of *infinite* expressivity that can be reliably acquired from sparse evidence — a defining hallmark of human language.

These late-stage agents are using a far more communicatively functional language than those earlier in the simulation run. However strange it sounds, this

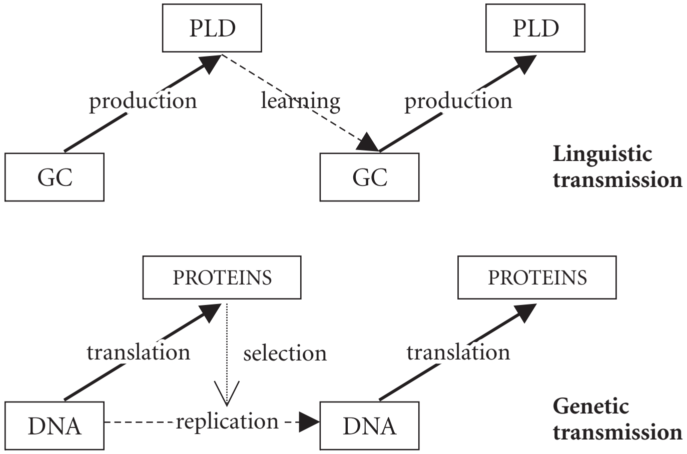

Figure 4. Similarities and differences between linguistic and genetic transmission. The central dogma of molecular biology states that there is no reverse translation from phenotype (i.e., proteins) to genotype (i.e., DNA). Genetic information persists by direct copying of the DNA. The only influence of the phenotype is in determining whether or not the organism has a chance of replication (hence, selection). In linguistic transmission, there is a far more complex mechanism — learning — that attempts to reconstruct grammatical competence (GC) by "reverse engineering" the primary linguistic data (PLD).

is merely a happy bi-product of the adaptive mechanism at work. Languages are not adapting to be more useful for the agents (at least not directly). Rather, they are simply adapting to aid their own transmission fidelity. In practice, this will usually be the same thing.

If this idea is correct, then it would be interesting to try and find examples where the needs of language (to survive from generation to generation) and the needs of its users (to communicate easily and successfully) diverge. In other words, can we find apparently disfunctional aspects of language that are nevertheless stable, and furthermore can we give these a natural explanation in terms of iterated learning? This is a challenging research goal, but there may be places we can start to look. For example, there are constructions that are notoriously hard to parse, such as centre-embedded relative clauses that are nevertheless clearly part of everyone's linguistic competence. Why are we burdened with these apparently suboptimal aspects of grammar? Perhaps the answer will lie in understanding the relative stability through iterated learning of a language with centre-embedding and a minimally different one that ruled-out the difficult constructions.

6. Conclusion

In this paper, we have explored the relationship between Universal Grammar and universal properties of language structure in the light of recent computational models of linguistic transmission. In summary:

- We treat Universal Grammar as a theory of what the language learner brings to the task of language acquisition that is independent of the linguistic data. In other words, UG is determined by the initial state of the child in addition to the Language Acquisition Device.
- UG in this sense can be equated to prior learning bias in a general Bayesian approach to learning. This prior bias is innately coded.
- It is fruitless to search for a bias-free model of language acquisition. In other words, there will always be a role for innateness in understanding language acquisition.
- The degree to which our innate bias is language specific is an open question — one that will probably require an evolutionary approach to answer.
- Both functionalist explanations for language universals and explanations in terms of UG suffer from the problem of linking an individual-level phenomenon (e.g., learning bias, processing pressures, social factors, etc.) with a global property of linguistic distribution.
- Language is a particular kind of cultural adaptive system that arises from information being transmitted by iterated learning.
- Computational models have been employed to uncover properties of iterated learning. For example, where the language model is a mapping between structured meanings and structured signals, compositionality emerges.
- One way of understanding language universals in the light of iterated learning is as adaptive solutions to the problem language faces of being successfully transmitted.

Because the connection between UG and universal properties of linguistic structure is not direct, we need to be cautious about how we use linguistic evidence. As Niyogi & Berwick (1997) show in their work on the link between acquisition and language change, a theory of acquisition that is explicitly designed to account for syntactic variation may actually make the wrong predictions once linguistic transmission is taken into account.

On the other hand, iterated learning can lift some of the burden of explanation from our theories of universal grammar. Jäger (2003) examines a model of variation in case-systems based on functional Optimality Theory. To account for the known facts a rather unsatisfying extra piece of theoretical machinery — the case hierarchy of Aissen (2003) — has been proposed. Using simulations of iterated learning, in combination with a model of the linguistic environment based on corpora, Jäger demonstrates that this hierarchy emerges 'for free' from the iterated learning process.

We hope that future research will continue to discover general, universal properties of iterated learning as well as relating these to questions of genuine interest to linguistics. In some ways these goals are orthogonal. The most idealised models of linguistic transmission tend to have questionable relevance to linguistics. For example, the 'language dynamical equation' developed by Nowak, Komarova & Niyogi (2001) treats language acquisition simply as a matrix of transition probabilities, and combines this with a model of reproductive fitness of speakers in a population. This leads to mathematically tractable solutions for a very limited subset of possible models of acquisition, but it is far from clear that these results correspond to anything in the real world (for example, it seems implausible that language change is driven primarily by the number of offspring a particular speaker has).

Nevertheless, we *do* need idealised models such as those we have presented; but crucially, models that can help us to understand how the real linguistic system adapts. Getting the balance right between tractable idealisation, and relevant realism is likely to be the biggest challenge facing future research.

Notes

[1] Note however that Newmeyer (this special issue) argues that, in practice, parametric theories of UG are poor explanations for implicational and statistical universals.

[2] This is actually a slight simplification. For a given hypothesis, h, that is not learnable, we can treat this as being excluded from the set H (giving us a second type of information the learner brings to the learning task), or by including it in the set and assigning it a prior probability of zero.

[3] A key component of explanations for universals that license different types in an asymmetrical markedness relationship is the existence of 'competing motivations' that create complex dynamics — see Kirby (1999) for details.

[4] Music might be another example. Miranda, Kirby & Todd (in press) use simulations of iterated learning to explore new compositional techniques which reflect the cultural evolution of musical form.

[5] We should be a little cautious of this comparison, however. The holistic protolanguage in the simulation is learned, whereas most animal communication systems are innately coded — although there appear to be some exceptions to this generalisation.

[6] This process bears some similarity to an optimisation technique in computer science called 'simulated annealing' (Kirkpatrick & Vecchi 1983). The search-space is explored over a wide area initially, but as the solution is approached, the search focuses in more closely on the region of the relevant region of space. It is interesting that this kind of optimisation arises naturally out of iterated learning without it being explicitly coded anywhere in the model.

[7] The frequency of meaning expression is presumably driven largely by the environment (although Tullo & Hurford 2003 look at a model where ongoing dialog determines meaning-choice in deriving the Zipfian distribution). Grounded models from robotics give us increasingly sophisticated ways of relating meanings and environment (e.g., Vogt 2002).

[8] There is another possible way of explaining why languages typically exhibit these word-order patterns. Dryer (1992) and Christiansen & Devlin (1997) refer to consistent branching

direction rather than head-ordering, although these are nearly equivalent. Consistently left- or right-branching languages are more common than mixed types. Brighton (2003) shows that a general property of stable languages in the ILM is the simplicity of their grammatical representation, where simplicity is defined in terms of the number of bits the learners use for storage. A topic for ongoing research is whether the commonly occurring word-order patterns are those that result in maximally compressible representations.

References

Aissen, J. 2003. “Differential object marking: iconicity vs. economy”. *Natural Language and Linguistic Theory* 21: 435–483.

Batali, J. 1998. “Computational simulations of the emergence of grammar”. In: Hurford, J.R.; Studdert-Kennedy, M.; and Knight, C. (eds), *Approaches to the evolution of language: social and cognitive bases* 405–426. Cambridge: CUP.

Batali, J. 2002. “The negotiation and acquisition of recursive grammars as a result of competition among exemplars”. In: Briscoe, E.J. (ed.), *Linguistic evolution through language acquisition* 111–172. Cambridge: CUP.

Blackmore, S. 1999. *The meme machine.* Oxford: OUP.

Brighton, H. 2002. “Compositional syntax from cultural transmission”. *Artificial Life* 8: 25–54.

Brighton, H. 2003. Simplicity as a driving force in linguistic evolution. Unpublished PhD Thesis, University of Edinburgh.

Brighton, H.; and Kirby, S. 2001. “The survival of the smallest: stability conditions for the cultural evolution of compositional language”. In: Kelemen, J.; and Sosik, P. (eds), *Advances in artificial life* (vol. 2159) 592–601. Berlin: Springer.

Chomsky, N. 1972. *Language and mind* (2nd ed.). New York: Harcourt, Brace & World.

Chomsky, N. 1981. *Lectures on government and binding.* Dordrecht: Foris.

Chomsky, N. 2002. *On nature and language.* Cambridge: CUP.

Christiansen, M.; and Devlin, J. 1997. “Recursive inconsistencies are hard to learn: a connectionist perspective on universal word-order correlations”. In: Shafto, M; and Langley, P. (eds), *Proceedings of the 19th annual Cognitive Science Society Conference* 113–118. Mahwah, NJ: Erlbaum.

Crick, F. 1970. “Central dogma of molecular biology”. *Nature* 227: 561–563.

Croft, W. 2000. *Explaining language change.* London: Longman.

Deacon, T. 1997. *The symbolic species.* New York: Norton.

Dryer, M. 1992. “The Greenbergian word-order correlations”. *Language* 68: 81–138.

Elman, J. 1990. “Finding structure in time”. *Cognitive Science* 14(2): 179–211.

Elman, J.; Bates, E.A.; Johnson, M.H.; Karmiloff-Smith, A.; Parisi, D.; and Plunkett, K. 1996. *Rethinking innateness.* Cambridge, MA: MIT Press.

Grimshaw, J. 1997. “Projection, heads and optimality”. *Linguistic Inquiry* 28: 373–422.

Hawkins, J.A. 1994. *A performance theory of order and constituency.* Cambridge: CUP.

Hawkins, J.A. (ed.). 1988. *Explaining language universals.* Oxford: Basil Blackwell.

Hurford, J.R. 2000. “Social transmission favours linguistic generalisation”. In: Knight, C.; Studdert-Kennedy, M.; and Hurford, J.R. (eds), *The evolutionary emergence of language: social function and the origins of linguistic form* 324–352. Cambridge: CUP.

Hurford, J.R. 2002. “Expression/induction models of language evolution: dimensions and issues”. In: Briscoe, E.J. (ed.), *Linguistic evolution through language acquisition* 301–344. Cambridge: CUP.

Jackendoff, R. 2002. *Foundations of language: brain, meaning, grammar, evolution.* Oxford: OUP.

Jäger, G. 2003. "Simulating language evolution with functional OT". In: Kirby, S. (ed.), *Language evolution and computation: ESSLLI workshop proceedings* 52–61.

Kirby, S. 1999. *Function, selection and innateness: the emergence of language universals.* Oxford: OUP.

Kirby, S. 2000. "Syntax without natural selection: how compositionality emerges from vocabulary in a population of learners". In: Knight, C.; Studdert-Kennedy, M.; and Hurford, J.R. (eds), *The evolutionary emergence of language: social function and the origins of linguistic form* 303–323. Cambridge: CUP.

Kirby, S. 2001. "Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity". *IEEE Journal of Evolutionary Computation* 5(2): 102–110.

Kirby, S. 2002a. "Learning, bottlenecks and the evolution of recursive syntax". In: Briscoe, E.J. (ed.), *Linguistic evolution through language acquisition.* Cambridge: CUP.

Kirby, S. 2002b. "Natural language from artificial life". *Artificial Life* 8: 185–215.

Kirby, S.; and Hurford, J.R. 1997. "Learning, culture and evolution in the origin of linguistic constraints". In: Husbands, P.; and Harvey, I. (eds), *Proeceedings of the 4th European Conference on Artificial Life* 493–502. Cambridge, MA: MIT Press.

Kirby, S.; and Hurford, J.R. 2002. "The emergence of linguistic structure: an overview of the iterated learning model". In: Cangelosi, A.; and Parisi, D. (eds), *Simulating the evolution of language* 121–148. Berlin: Springer.

Kirkpatrick, S.; Gelatt, C.D. Jr.; and Vecchi, M.P. 1983. "Optimization by simulated annealing". *Science* 220 (4598): 671–680.

Li, M.; and Vitanyi, P. 1993. *Introduction to Kolmogorov complexity.* London: Springer.

Miranda, E.; Kirby, S.; and Todd, P. In press. "On computational models of the evolution of music: from the origins of musical taste to the emergence of grammars". *Contemporary Music Review.*

Mitchell, T. 1997. *Machine learning.* New York: McGraw Hill.

Newmeyer, F.J. 1998. *Language form and language function.* Cambridge, MA: MIT Press.

Niyogi, P.; and Berwick, R. 1997. "A dynamical systems model of language change". *Complex Systems* 11: 161–204.

Nowak, M.; Komarova, N.; and Niyogi, P. 2001. "Evolution of Universal Grammar". *Science* 291: 114–118.

Smith, K. 2003. The transmission of language: models of biological and cultural evolution. Unpublished PhD Thesis, University of Edinburgh.

Smith, K.; Brighton, H.; and Kirby, S. Forthcoming. "Complex systems in the language evolution: the cultural emergence of compositional structure". *Advances in Complex Systems.*

Steels, L. 1999. *The talking heads experiment, vol. 1: words and meanings.* Antwerpen: LABORATORIUM.

Teal, T.; and Taylor, C. 1999. "Compression and adaptation". In: Floreano, D.; Nicoud, J.D.; and Mondada, F. (eds), *Advances in artificial life* (vol. 1674) 709–719. Berlin: Springer.

Tonkes, B. 2002. On the origins of linguistic structure: computational models of the evolution of language. Unpublished PhD Thesis, University of Queensland, Australia.

Tullo, C.; and Hurford, J.R. 2003. "Modelling Zipfian distributions in language". In: Kirby, S. (ed.), *Language evolution and computation: ESSLLI workshop proceedings* 62–75.

Vogt, P. 2002. "The physical symbol grounding problem". *Cognitive Systems Research* 3(3): 429–457.

Vogt, P. 2003. "Iterated learning and grounding: from holistic to compositional languages". In: Kirby, S. (ed.), *Language evolution and computation: ESSLLI workshop proceedings* 76–86.

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

Zuidema, W. 2001. “Emergent syntax: the unremitting value of computational modelling for understanding the origins of complex language”. In: Kelemen, J.; and Sosik, P. (eds), *Advances in artificial life* (vol. 2159) 641–644. Berlin: Springer.

Authors' addresses

Simon Kirby
Language Evolution and Computation
Research Unit,
School of Philosophy, Psychology and
Language Sciences,
University of Edinburgh,
40, George Square,
Edinburgh, EH8 9LL
UK

simon@ling.ed.ac.uk

Kenny Smith
Language Evolution and Computation
Research Unit,
School of Philosophy, Psychology and
Language Sciences,
University of Edinburgh,
40, George Square,
Edinburgh, EH8 9LL
UK

kenny@ling.ed.ac.uk

Henry Brighton
Language Evolution and Computation
Research Unit,
School of Philosophy, Psychology and
Language Sciences,
University of Edinburgh,
40, George Square,
Edinburgh, EH8 9LL
UK

henryb@ling.ed.ac.uk