Language as a Complex Adaptive System: Coevolution of Language and of the Language Acquisition Device

Ted Briscoe
ejb@cl.cam.ac.uk
Computer Laboratory
University of Cambridge
Pembroke Street
Cambridge CB2 3QG, UK

July 1, 1998

Abstract

An account of parameter setting during grammatical acquisition is presented in terms of Generalized Categorial Grammar embedded in a multiple default inheritance hierarchy, providing a natural partial ordering on the setting of parameters (Briscoe, 1997a). Experiments reported show that several experimentally effective learners can be defined in this framework capable of reliably acquiring a grammar from a sequence of triggers drawn from one of 70 full languages (or the 200+ more restricted subset languages of these full languages).

Evolutionary computational simulations of evolving populations of such language learners/users suggest that: 1) languages evolve towards greater learnability, interpretability and/or expressivity; 2) learning procedures evolve towards more efficient variants depending on the linguistic environment of adaptation. The reciprocal evolution of language learning procedures and of language creates a genuinely coevolutionary dynamic, despite the relative speed of linguistic selection for language variants compared to natural selection for variant language learning procedures.

1 Theoretical Background

It is now widely accepted that language acquisition is guided by an innate language learning procedure and a partial innate specification of the form of language. Language acquisition by children is a near-universal feat, where (partial) failure appears to correlate more with genetic deficits (e.g. Gopnik, 1994) or with an almost complete lack of linguistic input during the critical period (e.g. Curtiss, 1988), than with measures of general intelligence (e.g. Smith and Tsimpli, 1991) or the quality of the learning enviroment (e.g. Ochs, 1982). Pinker and Bloom (1990) have argued for an adaptationist account of the evolution of the language acquisition device (LAD) suggesting that the domain-specific linguistic (broadly grammatical) knowledge required to support reliable language learning was genetically assimilated via natural selection for more successful language learners since the emergence of language. Recently, Deacon (1997) has argued that there has been no genetic assimilation of language-specific knowledge, rather languages themselves rapidly evolved to be easily learnable given general domain-independent learning biases such as working memory limitations.

Genetic assimilation is a neo-Darwinian mechanism supporting apparent 'inheritance of acquired characteristics' (e.g. Waddington, 1975). The fundamental insights are that: 1) plasticity in the relationship between phenotype and genotype is under genetic control, 2) novel environments create selection pressures which favour organisms with the plasticity to allow within-lifetime developmental adaptations to the new environment, 3) natural selection will function to 'canalize' these developmental adaptations by favouring genotypic variants in which the appropriate trait develops reliably on the basis of minimal environmental stimulus, providing that the environment, and consequent selection pressure, remains constant over enough generations.[1] Durham (1991) discusses in detail the case of widespread, though by no means universal, lactose tolerance in adult humans. Many of us, uniquely amongst mammals, continue to be able to easily digest milk after weaning. In many parts of the world the growth of animal husbandry created a new and reliable source of nutrition – milk. Thus, individuals more able to exploit this resource for longer periods of their lifetime were selected for. Lactose tolerance has been genetically assimilated by the great majority in populations where milk has been reliably available over many generations. Although it is not possible to relate lactose tolerance directly to specific genetic differences (yet), Durham demonstrates convincingly that the incidence of intolerance correlates, in a manner compatible with a genetic explanation, with a fairly recent introduction of diary products and with warm climates, where lack of Vitamin D is less potentially problematic.

Waddington, himself, suggested that genetic assimilation provided a possible mechanism for the gradual evolution of a LAD: 'If there were selection for the ability to use language, then there would be selection for the capacity to acquire the use of language, in an interaction with a language-using environment; and the result of selection for epigenetic responses can be, as we have seen, a gradual accumulation of so many genes with effects tending in this direction that the character gradually becomes genetically assimilated.' (1975:305f). Pinker and Bloom (1990:479-80) make the same suggestion, citing Hinton and Nowlan's (1987) computational simulation showing genetic assimilation of initial node settings facilitating learning in a population of neural networks.[2] Deacon (1997:102f,327f) rejects this account though,

[1]Waddington's work on genetic assimilation is a neo-Darwinian refinement of an idea independently discovered by Baldwin, Lloyd Morgan and Osborne in 1896, and often referred to as the Baldwin Effect (see Richards, 1987 for a detailed history). Waddington refined the idea by emphasizing the role of canalization and the importance of genetic control of ontogenetic development – his 'epigenetic theory of evolution'. He also undertook experiments with flies which directly demonstrated genetic assimilation for artificial environmental changes. Evolutionary biologists accept the possibility of genetic assimilation (e.g. Maynard Smith, 1993:319f; Rose, 1997:217f), however, many (e.g. Dawkins, 1982:284) regard it as a 'hypothetical' mechanism because, though it has been demonstrated experimentally, it has not been conclusively demonstrated to occur in the field. It is extremely difficult to conclusively prove a case of adaptive genetic assimilation. Nevertheless, the developmental view of evolution, which Waddington pioneered, is gaining ground as more is understood about the relationship between genes and environment in morphogenesis (e.g. Jablonka and Lamb, 1995).

[2]A complication for this account of language learning is that it does not explain why genetic assimilation should not have continued until the point where a fully-specified language (or at least grammar) had been assimilated, and learning became redundant. Waddington (1975:307) remarks: 'Evolution is quite capable of performing such a feat... But in the case of language, there is certainly little reason to see why it would have been advantageous to press the matter further. If a child which had never met a language-user developed the ability to talk, who after all would it talk to?' Nevertheless, the propensity to use a fully-specified grammar given minimal triggering input would simplify the language learning problem to one of vocabulary acquisition. Pinker and Bloom (1990:480), following Hinton and Nowlan (1987), argue that selection pressure to set the remaining initial nodes in the neural networks is weak once networks have evolved to learn reliably. However, Harvey (1993) demonstrates that this is an artifact of Hinton and Nowlan's simulation design – later more effective networks almost invariably evolve from a single ancestor, causing 'premature' fixation of some unset nodes, and thus preventing the population from evolving further. As long as there is selection pressure for a fully-developed capacity and no premature fixation, we expect no learning, and thus no delay in acquisition of the trait, to be the optimal solution. Ackley and Littman (1991) and Cecconi *et al.* (1995) describe quite unrelated simulations which, unlike Hinton and Nowlan, distinguish phenotype and genotype, do not make use of a fixed externally-defined fitness function, and do model learning cost – in these simulations

on the basis that genetic assimilation requires an unchanging environment to create the sustained selection pressure over the many generations required for genotypic adaptation. Pinker and Bloom (1990) simply assume that linguistic universals are evidence of enough constancy in the environment to allow genetic assimilation. However, once we view language itself as an adaptive system, this assumption, that universals are unambiguous evidence of genetic assimilation of linguistic knowledge into a LAD, is no longer valid.

Hurford (1987) and Kirby (1996, 1997) argue that many linguistic universals, and especially typological, implicational or statistical universals, are the result of historical adaptations by languages to the capacities and limitations of language users. Drawing on Hawkins' (1994) work on constituent order universals and his metric of parsability of different ordering configurations, Kirby demonstrates that, on the assumption that differential parsability translates into differential learnability, more learnable order variants will emerge and persist through the language learning 'bottleneck' across generations of language users. Kirby's simulations tied to Hawkins' theory of parsability and constituent order provide one of the most detailed demonstrations of the possibility of genuinely linguistic evolution, where variant construction types compete for language users and more learnable variants are selected.[3]

In Hurford (1987) and Kirby's work the term evolution is used in its technical 'universal Darwinist' sense of (random) variation, adaptive selection and differential inheritance applied to any dynamic system (e.g. Dawkins, 1983; Cziko, 1995). Keller (1994) provides an extensive argument for the view that language is a 'phenomenon of the third kind'; that is, a human artefact which naturally evolves – a type of (complex) adaptive system.[4] We will refer to linguistic selection, as opposed to natural selection, to emphasize the claim that evolution is operating here on linguistic constructions, rather than on their users. Linguistic variants compete for host minds/brains, as Dawkins (1989:192f) and Dennett (1991:341f) have argued that ideational units or 'memes' do. Given this perspective, linguistic universals could have arisen via historical linguistic evolution in response to similar pressures for learnability, interpretability and/or expressivity. Therefore, Pinker and Bloom (1990) are not justified in assuming that the existence of universals, whether absolute or statistical, is incontrovertible evidence that genetic assimilation of such universals into a LAD has occurred.

Deacon (1997:116f) independently argues for the even stronger position that all linguistic 'universal[s]... emerged spontaneously and independently in each evolving language, in response to universal biases in the selection processes affecting language transmission. They are *convergent* features of language evolution in the same ways as dorsal fins of sharks, ichthyosaurs, and dolphins are independent convergent adaptations of acquatic species.' He suggests, in particular, that languages have evolved to be easily learnable by a learning procedure which 'starts small' (Elman, 1993) with a limited working memory only capable of 'seeing' local grammatical dependencies. It is known that working memory grows through childhood (e.g.

learning is eventually entirely displaced, given a constant environment.

[3]Though the idea is not new: Müller, Schleicher and other 19th century linguists speculated that languages evolved according to Darwinian theory, and Darwin (1871) endorsed the idea, quoting with approval from Müller: 'A struggle for life is constantly going on amongst the words and grammatical forms in each language. The better, the shorter, the easier forms are constantly gaining the upper hand, and they owe their success to their own inherent virtue.' See Harris and Taylor (1997:ch14) and McMahon (1994:ch12) for more discussion of the relationship between Darwinian and linguistic theory, and Keller (1994:46f) for a critical discussion of Müller and Schleicher's theory of language.

[4]Complexity in dynamic (adaptive) systems has many sources (see Casti, 1994 for an overview). One source of complexity in natural language arises from the often conflicting selective pressures of learnability, expressivity and interpretability. Different languages represent different and unpredictable responses to such pressures.

Baddeley, 1992), and this may assist learning by ensuring that trigger sentences gradually increase in complexity through the acquisition period by obliging the learner to ignore more complex potential triggers that occur early in the learning process. However, this working memory matures towards the end of the critical period into a system with adult working memory performance capable of parsing long-distance grammatical dependencies too. Thus, maturing working memory acts as a filter on linguistic input. Crucially though, there is nothing language-specific in this learning bias. Furthermore, Deacon (1997:328f) argues that the surface grammatical organization of languages changes with such speed relative to genetic evolution that there could not have been consistent enough selection pressure for genetic assimilation. He, therefore, rejects Pinker and Bloom's (1990) argument for the evolution of a LAD, arguing that languages are learnable because they have evolved to be so, rather than because we have evolved machinery to make them learnable.

Deacon's strong position can be criticized on two levels. Firstly, it is unclear that he recognizes the import of linguistic learnability arguments and the relevance of abstract universals (without clear 'surface' effects). For example, the language learning procedure presented below can parse and learn grammatical constructions involving cross-serial grammatical dependencies, such as those exemplified in the formal language $a^n b^n c^n$, Swiss German syntax and Bambara morphology (e.g. Shieber, 1985; Gazdar 1988), but not constructions involving the MIX or Bach language variant in which any ordering of equal numbers of the as, bs and cs is grammatical, creating arbitrarily intersecting dependencies. Furthermore, no formal or computational demonstration of learnability for such 'mildly context-sensitive', indexed languages has been presented (Joshi *et al.*, 1991) – Elman's (1993) neural network experiments, cited extensively by Deacon, demonstrate, at best, an ability to learn an approximate context-free language recognizer. Whether a language exhibits cross-serial or arbitrarily intersecting dependencies is an apparently rather abstract feature which does not fit well into traditional more 'surfacy' characterizations of languages as, say, inflecting, agglutinating or isolating. Nevertheless, it has profound consequences for the kind of rule system capable of expressing the mapping from surface syntax to meaning / logical form. The genetic assimilation of a language-specific rule *system* (the UG component of a LAD) remains a theoretical possibility even if the emergence of such abstract universals can be traced to non-domain-specific factors, such as working memory limitations.

Secondly, Waddington's work on genetic assimilation is not the only approach relevant to evolutionary interactions between distinct evolving systems. In recent years, the increased use of mathematical tools and computational simulation has demonstrated the probability of extensive coevolutionary interactions between species, such as predator-prey interactions, competitive and benign host-parasite interactions, plant-insect interactions, and so forth (e.g. Futuyma and Slatkin, 1983; Kauffman, 1993). Most of these interactions involve species evolving at different rates, as the lifespan of the parasite is usually far shorter than that of the host. Though Waddington's neo-Darwinian mechanism of genetic assimilation remains the basis for (co)evolution in response to environmental change, this work suggests that relative speed alone cannot conclusively be used to reject the possibility of genetic assimilation in response to pressure from an evolving linguistic environment. Interestingly, though Deacon (1997:112-13) draws the analogy between language and symbiotic bacteria (for example, those found in the human gut which aid digestion) and subtitles his book 'co-evolution of language and brain', he does not explicitly discuss the recent literature on coevolution, or whether this might warrant reconsideration of how environmental changes affect genetic assimilation.

In the simulation model described below, the fitness functions, when utilized, assume a benign, symbiotic relationship between languages and their potential users in which the ability to communicate via language confers selective advantage, but additionally the ability to communicate using a more learnable, expressive or interpretable variant language can confer greater relative advantage. Roughgarden (1983) argues that mutualistic coevolution between 'host' (language users) and 'guest' (language idiolects) organisms will only occur when the host benefits. Linguistic variants compete for language users on the basis of their relative learnability, interpretability and/or expressivity. In this sense, a language is a parasitic coevolving species. However, there is a critical difference: language is a human artefact, and not self-replicating. Language variants may compete for brains, but they cannot have a fitness, in the technical sense of expected number or proportion of offspring. Rather the primary mechanism of linguistic inheritance is through a child language learner actively learning her idiolect (rather than the gene actively promoting its replication). Idiolects also change through an adult's within-lifetime responses to the changing linguistic environment, and the needs of interpretability and expressivity. Thus, the speed at which linguistic changes can diffuse through a population will be potentially far faster than that at which genetic change can do so. However, there is clearly a speed limit to linguistic change within a successfully communicating population and that speed limit means that only a small part of the space of possible grammars may be sampled over the period required for biological evolution. This may lead to a constant selection pressure capable of supporting genetic assimilation of a LAD.

Experiments with the (co)evolutionary simulation model and model of a language learner/user, described in section 2 and section 3 respectively, demonstrate that: 1) language can be insightfully modelled as a complex adaptive system, constantly evolving under often conflicting linguistic selection pressures of learnability, interpretability and expressivity; 2) some universals may be a consequence of such linguistic selection; 3) specifically, constituent order universals of the type studied by Hawkins (1994) largely follow from the model of differential parsability based on working memory limitations; 4) Elman's (1993) 'starting small' hypothesis, modelled as differential parsability and combined with a learning procedure incorporating grammatical knowledge, creates linguistic selection for more memory efficient ordering configurations and languages; 5) nevertheless, a language-specific innate learning system (i.e. a LAD) can emerge through a coevolutionary process, despite the potentially rapid pace of linguistic evolution relative to biological evolution.

2 The Language Acquisition Device

In this section, a parameter setting model of the language learning procedure is described. Grammatical acquisition proceeds on the basis of a partial genotypic specification of (universal) grammar (UG) complemented with a genetically-specified learning *procedure* enabling the child to acquire a target grammar given a sequence of triggering sentence types. In the parameter setting framework of Chomsky (1981) learning involves fixing the values of a finite set of finite-valued parameters to select a single fully-specified grammar from within the space defined by UG. Triggers are defined as presentations of particular sentence types in a context which makes their meaning clear. Thus, the task of the learner is defined as that of recovering the mapping between surface form (SF) and logical form (LF) given a particular presentation sequence of such SF-LF pairings (e.g. Wexler and Culicover, 1980). Formal accounts of parameter setting have been developed for small fragments but even given this idealized model of language learning, search spaces contain local maxima and subset-superset relations which may cause a learner to converge to an incorrect grammar given a particular presentation sequence of triggers (Clark, 1992; Frank and Kapur, 1996; Gibson and Wexler, 1994; Niyogi and Berwick, 1995). One possible solution to these problems involves defining default, unmarked initial values for (some) parameters and/or partially ordering the setting of parameters during learning (Briscoe, 1997a). Another possibility is that filtering learning data, say, according to parsability, in accord with the 'starting small' hypothesis may avoid problematic presentation sequences.

Bickerton (1984) argues for the Bioprogram Hypothesis as an explanation for universal similarities between historically-unrelated creoles, and for the rapid increase in grammatical complexity accompanying the transition from pidgin to creole languages. From the perspective of the parameters framework, this hypothesis claims that children are endowed genetically with a UG which, by default, specifies the stereotypical core creole grammar, with right-branching syntax and subject-verb-object order, as in Saramaccan. Others working within the parameters framework have proposed unmarked, default parameters (e.g. Lightfoot, 1991), but the Bioprogram Hypothesis can be interpreted as towards one end of a continuum of proposals ranging from all parameters initially unset to all parameters set to default values. In work on formal learnability of parameters, it is usually assumed that the 'starting point' for learning is arbitrarily set parameters (i.e. any grammar in the set defined by the model), leading to the negative results reported above. Though the model we develop below makes no specific assumptions about the starting point for learning, the evolutionary experiments reported in section 5 suggest that genetic assimilation will result in a 'bioprogram-style' learner with a strong learning bias in favour of specific default, unmarked parameter settings.

The model of the language acquisition device incorporates a UG with associated parameters, a parser, and an algorithm for updating initial parameter settings on parse failure during learning. The following subsections define such a model.

2.1 The Grammar (set)

Classical (AB) categorial grammar uses one rule of application which combines a functor category, containing a slash, with an argument category to form a derived category with one less slashed argument category (e.g. Wood, 1993). Grammatical constraints of order and agreement are captured by only allowing directed application to adjacent matching categories. Generalized Categorial Grammar (GCG) extends CG with further rule schemata. The rules of forward application (FA), backward application (BA), generalized weak permutation (P) and forward and backward composition (FC, BC) are given in Figure 1 (where X, Y and Z are category variables, | is a variable over slash and backslash, and ... denotes zero or more further functor arguments). Permutation enables cyclical permutation of argument categories, but not modification of their directionality. Once permutation is included, several semantically equivalent derivations for *Kim loves Sandy* become available, Figure 2 shows the non-conventional left-branching one. Composition also allows alternative non-conventional semantically equivalent (left-branching) derivations (e.g. Steedman, 1988, 1996).

GCG as presented is inadequate as an account of UG or of any individual grammar. In particular, the definition of atomic categories needs extending to deal with featural variation, and the rule schemata, especially composition and weak permutation, must be restricted in various parametric ways so that overgeneration is prevented for specific languages. Nevertheless, GCG does represent a plausible kernel of UG; Hoffman (1995, 1996) explores the descriptive power of a very similar system, in which generalized weak permutation is not required because functor arguments are interpreted as multisets. She demonstrates that this system can handle (long-distance) scrambling elegantly, and can generate mildly context-sensitive languages, though not some MIX languages.

Forward Application:	
X/Y Y $\Rightarrow$ X	λ y [X(y)] (y) $\Rightarrow$ X(y)
Backward Application:	
Y X\Y $\Rightarrow$ X	λ y [X(y)] (y) $\Rightarrow$ X(y)
Forward Composition:	
X/Y Y/Z $\Rightarrow$ X/Z	λ y [X(y)] λ z [Y(z)] $\Rightarrow$ λ z [X(Y(z))]
Backward Composition:	
Y\Z X\Y $\Rightarrow$ X\Z	λ z [Y(z)] λ y [X(y)] $\Rightarrow$ λ z [X(Y(z))]
(Generalized Weak) Permutation:	
$(X\|Y_1)\ldots\|Y_n \Rightarrow (X\|Y_n)\|Y_1 \ldots$	$\lambda\ y_n \ldots, y_1\ [X(y_1 \ldots, y_n)] \Rightarrow \lambda \ldots y_1, y_n\ [X(y_1 \ldots, y_n)]$

Figure 1: GCG Rule Schemata

Kim	loves	Sandy
NP	(S\NP)/NP	NP
kim'	λ y,x [love$'$(x y)]	sandy$'$
	————P	
	(S/NP)\NP	
	λ x,y [love$'$(x y)]	
————————BA		
S/NP		
λ y [love$'$(kim$'$ y)]		
————————————————FA		
S		
love$'$(kim$'$ sandy$'$)		

Figure 2: GCG Derivation for *Kim loves Sandy*

NP	**N**	**S**	**gen-dir**	**subj-dir**	**applic**
A T	A T	A T	D R	D L	D T
NP	**gendir**	**applic**	**S**	**N**	**subj-dir**
A T	D R	D T	A T	A T	D L
applic	**NP**	**N**	**gen-dir**	**subj-dir**	**S**
D T	A T	A T	D R	D L	A T
...					

Figure 3: Sequential encodings of the grammar fragment

The relationship between GCG as a theory of UG (GCUG) and as a specification of a particular grammar is captured by embedding the theory in a default inheritance network. This is represented as a semi-lattice of typed default feature structures (TDFSs) representing subsumption and default inheritance relationships (Lascarides *et al*, 1996; Lascarides and Copestake, 1996, in press). The network defines intensionally the set of possible categories and rule schemata via type declarations on nodes. For example, an intransitive verb might be treated as a subtype of verb, inheriting subject directionality by default from a type **gendir** (for general direction). For English, **gendir** is default **right** but the node of the (intransitive) functor category, where the directionality of subject arguments is specified, overrides this to **left**, reflecting the fact that English is predominantly right-branching, though subjects appear to the left of the verb. A transitive verb would inherit structure from the type for intransitive verbs and an extra NP argument with default directionality specified by **gendir**, and so forth.

For the purposes of the evolutionary simulation described in section 3, GC(U)Gs are represented as a sequence of *p-settings* (where *p* denotes principles or parameters) based on a flat (ternary) sequential encoding of such default inheritance networks. The inheritance hierarchy provides a partial ordering on parameters, which is exploited in the learning procedure. For example, the atomic categories, **N**, **NP** and **S** are each represented by a parameter encoding the presence/absence or lack of specification (T/F/?) of the category in the (U)G. Since they are unordered in the semi-lattice, their ordering in the sequential coding is arbitrary. However, the ordering of the directional types **gendir** and **subjdir** (with values L/R) is significant as the latter is a more specific type. The distinctions between absolute, default or unset specifications also form part of the encoding (A/D/?). Figure 3 shows several equivalent and equally correct sequential encodings of the fragment of the English type system outlined above.

A set of grammars based on typological distinctions defined by basic constituent order (e.g. Greenberg, 1966; Hawkins, 1994) was constructed as a (partial) GCUG with independently varying binary-valued parameters. The eight basic language families are defined in terms of the unmarked, canonical order of verb (V), subject (S) and objects (O). Languages within such families further specify the order of modifiers and specifiers in phrases, the order of adpositions and further phrasal-level ordering parameters. Figure 4 lists the language-specific ordering parameters used to define the full set of grammars in (partial) order of generality, and gives examples of settings based on familiar languages such as "English", "German" and "Japanese".[5] "English" defines an SVO language, with prepositions (adpos R) in which spec(ifiers), complementizers (cpl) and some mod(ifiers) precede heads

[5]Throughout double quotes around language names are used as convenient mnemonics for familiar combinations of parameters. Since not all aspects of these actual languages are represented in the grammars, conclusions about actual languages must be made with care.

	gen	v1	n	subj	obj	v2	mod	spec	rcl	adpos	cpl
Engl	R	F	R	L	R	F	R	R	R	R	R
Ger	R	F	R	L	L	T	R	R	R	R	R
Jap	L	F	L	L	L	F	L	L	L	L	?

Figure 4: The Grammar Set Constituent Ordering Parameters

of phrases. There are other grammars in the SVO family in which all modifiers follow heads, there are postpositions, and so forth. Not all combinations of parameter settings correspond to attested languages and one entire language family (OSV) is arguably unattested (though see Pullum 1981). "Japanese" is an SOV language with postpositions in which specifiers and modifiers follow heads. There are other languages in the SOV family with less consistent left-branching syntax in which specifiers and/or modifiers precede phrasal heads, some of which are attested. "German" is a more complex SOV language in which the parameter verb-second (v2) ensures that the surface order in main clauses is usually SVO.[6]

There are 20 p-settings which determine the rule schemata available, the atomic category set, and so forth. In all, this CGUG defines just under 300 grammars – including 70 distinct full languages. Not all of the resulting languages are (stringset) distinct and some are proper subsets of others. "English" without the rule of permutation results in a stringset-identical language, but the grammar assigns different derivations to some strings, though the associated logical forms are identical. "English" without composition results in a subset language. Some combinations of p-settings result in 'impossible' grammars (or UGs). Others yield equivalent grammars, for example, different combinations of default settings (for types and their subtypes) can define an identical category set.

The grammars defined generate (usually infinite) stringsets of lexical syntactic categories. These strings are sentence types since each is equivalent to a finite set of grammatical sentences, formed by selecting a lexical item consistent with each lexical syntactic category. Such sequences of lexical syntactic categories can be viewed as triggers (determinate SF-LF pairings) because in this framework knowing the lexical syntactic category of each word in a sentence is enough to deterministically recover an unscoped LF. Languages are represented as a finite subset of sentence types generated by the associated grammar. These represent a sample of degree-0 or degree-1 learning triggers for the language (e.g. Lightfoot, 1991).[7] Subset languages are exemplified by between 3 and 9 such sentence types and full languages by 12 sentence types. The constructions exemplified by each sentence type and their length are equivalent across all the languages defined by the grammar set, but the sequences of lexical categories can differ. For example, two SOV language renditions of a sentence type / degree-1 trigger corresponding to *The man who Bill likes gave Fred a present*, one with premodifying and the other postmodifying relative clauses, both with a relative pronoun at the right boundary of the relative clause, are shown below with the differing category highlighted.

[6] Representation of the v1/v2 parameter(s) in terms of type constraints determining allowable verbal functor categories is discussed in more detail in Briscoe (1998).

[7] The degree-1 triggers / sentence types for relative clauses are included because these sentence types are important for distinguishing the languages in terms of parsability. So no theoretical claim that degree-1 triggers are essential to learning is being made.

1. The Reduce Step: if the top 2 cells of the stack are occupied,
   then try
   a) Application, if match, then apply and goto 1), else b),
   b) Composition, if match then apply and goto 1), else c),
   c) Permutation, if match then apply and goto 1), else goto 2)
2. The Shift Step: if the first cell of the Input Buffer is occupied,
   then pop it and move it onto the Stack together with its associated lexical syntactic category and goto 1),
   else goto 3)
3. The Halt Step: if only the top cell of the Stack is occupied by a constituent of category S,
   then return Success,
   else return Fail

The Match and Apply operation: if a binary rule schema matches the categories of the top 2 cells of the Stack, then they are popped from the Stack and the new category formed by applying the rule schema is pushed onto the Stack.

The Permutation operation: each time step 1c) is visited during the Reduce step, permutation is applied to one of the categories in the top 2 cells of the Stack (until all possible permutations of the 2 categories have been tried in conjunction with the binary rules). The number of possible permutation operations is finite and bounded by the maximum number of arguments of any functor category in the grammar.

Figure 5: The Parsing Algorithm

Bill likes who the-man a-present Fred gave
NP_s $(\mathrm{S}\backslash\mathrm{NP}_s)\backslash\mathrm{NP}_o$ $\mathrm{Rc}\backslash(\mathrm{S}\backslash\mathrm{NP}_o)$ $\mathbf{NP}_s\backslash\mathbf{Rc}$ NP_{o2} NP_{o1} $((\mathrm{S}\backslash\mathrm{NP}_s)\backslash\mathrm{NP}_{o2})\backslash\mathrm{NP}_{o1}$

The-man Bill likes who a-present Fred gave
$\mathbf{NP}_s/\mathbf{Rc}$ NP_s $(\mathrm{S}\backslash\mathrm{NP}_s)\backslash\mathrm{NP}_o$ $\mathrm{Rc}\backslash(\mathrm{S}\backslash\mathrm{NP}_o)$ NP_{o2} NP_{o1} $((\mathrm{S}\backslash\mathrm{NP}_s)\backslash\mathrm{NP}_{o2})\backslash\mathrm{NP}_{o1}$

The expressivity of a grammar/language is modelled (crudely) in terms of the proportion of sentence types which can be generated and parsed from the finite subset for the associated full language.

2.2 The Parser

The parser is a deterministic, bounded-context stack-based shift-reduce algorithm (see Briscoe, 1987 for further details and justification). The parser operates with two data structures, an input buffer or queue, and a stack or push down store. The algorithm for the parser working with a GCG which includes application, composition and permutation is given in Figure 5. This algorithm finds the most left-branching derivation for a sentence type because Reduce is ordered before Shift. It also finds the derivation involving the least number of parsing operations because only one round of permutation occurs each time application and composition fail.[8] The category sequences representing the sentence types in the data for the entire language set are designed to be unambiguous relative to this 'greedy, least effort' algorithm, so it will always assign the appropriate LF to each sentence type. However, there

[8]The preference for left-branching derivations and those involving the least number of parsing operations can be seen as a precise instantiation of the economy principle in Minimalist Grammar (Chomsky, 1995) within this framework.

After each parse step (Shift, Reduce, Halt (see Fig 5):

1. Assign any new Stack entry in the top cell (introduced by Shift or Reduce) a WML value of 0
2. Increment every Stack cell's WML value by 1
3. Push the sum of the WML values of each Stack cell onto the WML-record

When the parser halts, return the sum of the WML-record which gives the total WML for a derivation.

Figure 6: The WML Algorithm

Stack	Input Buffer	Op.	Step	WML
	K. loves S.		0	0
K.:NP:kim′	loves Sandy	Shift	1	1
l.:(S\NP)/NP:λ y,x(love′ x, y) K.:NP:kim′	Sandy	Shift	2	3
S.:NP:sandy′ l.:(S\NP)/NP:λ y,x(love′ x, y) K.:NP:kim′		Shift	3	6
l. S.:S/NP:λ x(love′ x, sandy′) K.:NP:kim′		Red.(A)	4	5
K. l. S.:S:(love′ kim′, sandy′)		Red.(A)	5	1

Figure 7: WML for *Kim loves Sandy*

are frequently alternative less left-branching or more 'expensive' derivations of the same LF, and in some cases a distinct LF could be recovered by generating all permutations of functors before attempting application/composition.

The parser is augmented with an algorithm which computes working memory load during an analysis. Limitations of working memory (e.g. Baddeley, 1992) are modelled in the parser by associating a cost with each stack cell occupied during each step of a derivation, and recency and depth of processing effects are modelled by resetting this cost each time a reduction occurs (see Briscoe, 1987, 1998 for further discussion). The working memory load (WML) algorithm is given in Figure 6. Figure 7 gives the right-branching derivation for *Kim loves Sandy*, found by the parser utilizing a grammar without permutation. The WML at each step is shown for this derivation. The overall WML (16) is higher than for the left-branching derivation (9).

The WML algorithm is used to rank the parsability of sentence types, and thus indirectly languages, by parsing each sentence type from the exemplifying data with the associated grammar and then taking the mean of the WML obtained for these sentence types. "English" with Permutation has a lower mean WML than "English" without Permutation, though they are stringset-identical, whilst a hypothetical mixture of "Japanese" SOV clausal order with "English" phrasal syntax has a mean WML which is 25% worse than that for "English". The predictions of the WML algorithm are in broad accord with existing psycholinguistically and typologically motivated theories of parsing complexity (see. Briscoe, 1987,1998; Gibson, 1991; Hawkins, 1994; Rambow and Joshi, 1994).

2.3 The Parameter Setting Algorithm

The parameter setting algorithm is an extension of Gibson and Wexler's (1994) Trigger Learning Algorithm (TLA) to take account of the inheritance-based partial ordering and the role of memory in learning. The TLA is error-driven – parameter settings are altered in constrained ways when a learner cannot parse trigger input. Trigger input is defined as primary linguistic data which, because of its structure or context of use, is determinately unparsable with the correct interpretation (e.g. Lightfoot, 1991 and see section 2.2).[9]. The TLA is memoryless in the sense that a history of parameter (re)settings is not maintained, in principle, allowing the learner to revisit previous hypotheses. This is what allows Niyogi and Berwick (1995) to formalize parameter setting as a Markov process. However, as Brent (1996) argues, the psychological plausibility of this algorithm is doubtful – there is no evidence that children move between neighbouring grammars along paths that (blindly) revisit previous hypotheses. Therefore, in the modified algorithm each parameter can only be reset once during the learning process, resulting in a learning procedure with (limited) memory. The TLA is local in the sense that only one (random) parameter can be reset on parse failure. In the modified algorithm, we relax this requirement to n parameters per parse failure. Bertolo (1995) shows that this relaxation of the TLA does not alter fundamental results concerning local maxima and learnability.[10] The TLA is unordered in the sense that on parse failure a parameter is chosen at random to be reset. In the modified algorithm, parameters are (re)set starting with the most general in terms of the partial order defined by the inheritance hierarchy. Once (re)set they are not revisited because the procedure utilizes limited memory. Both the TLA and the modified algorithm are greedy in the sense that a parameter updated on parse failure is retained if that setting allows the current trigger to be reparsed successfully.

Each step for a learner can be defined in terms of three functions: P-SETTING, GRAMMAR and PARSER, as:

$$\text{PARSER}_i(\text{GRAMMAR}_i(\text{P-SETTING}_i))(\text{Sentence}_j)$$

A p-setting defines a grammar which in turn defines a parser (where the subscripts indicate the output of each function given the previous trigger). The modified algorithm is summarized in Figure 8. The core of the algorithm is the update rule, which is applied to the sequential p-setting encoding described in section 2.1; for instance, a default parameter can be reset to its opposite value and the 'D' encoding changed to a 'R" to record that this default parameter has been reset, and so forth.

The learning procedure can be made maturational and incorporate the 'starting small' hypothesis (see section 1). The WML of a sentence type can be used to determine whether it can function as a trigger at a particular stage in learning, thus filtering random presentation of triggers and ensuring that triggers are presented in (partial) order of decreasing parsability.

In summary, this account of the parameter setting procedure is error-driven, greedy, n-local, partially-ordered, utilizes limited memory, and can incorporate maturationally developing working memory limitations, effectively filtering trigger in-

[9]In the simulation, sentence types used as triggers are represented by p-setting schemata (see e.g. Clark, 1992) with associated memory loads to avoid the need for continuous on-line parsing of triggers

[10]The motivation for relaxing the single-value constraint and adopting a n-local variant of the TLA is twofold: firstly, the selection of a fair sample of triggers / sentence types with respect to WML creates unbalanced trigger paths with respect to the number of parameter resettings required to successfully learn a given language (see e.g. Frank and Kapur, 1996) for discussion of optimal sequences of triggers); secondly, the parameter n can be varied in the evolutionary simulation, creating a wider range of learning procedures to select from.

Data: $\{S_1, S_2, \ldots S_n\}$

unless
  $\textsc{parser}_i(\textsc{grammar}_i(\textsc{p-setting}_i))(S_j) = \text{Success}$
then
  p-setting$_j$ = $\textsc{update}$(p-setting$_i$)
  unless
  $\textsc{parser}_j(\textsc{grammar}_j(\textsc{p-setting}_j(S_j))) = \text{Success}$
  then
    RETURN p-setting$_i$
  else
    RETURN p-setting$_j$

UPDATE:
Reset the first n default parameter(s) or set the first n unset parameter(s) in a 'left-to-right' search of the p-settings (consistent with the partial order encoding their generality) according to the following table:

Input:	D 1	D 0	? ?
Output:	R 0	R 1	? 1/0 (random)

(where 1 = T/L and 0 = F/R – see figs. 3/4 above)

Figure 8: The Learning Algorithm

put. Finally, the initial configuration of the parameters in the TLA is usually taken to be any arbitrary grammar, though as Gibson and Wexler (1994) point out, assuming (some) specific default initial settings can remove local maxima. In this model, parameters can be initially unset (?) or have a default (D) value (see section 2.1). The precise choice of initial settings and of the n (re)settable parameters per trigger define a space of variant learning procedures for (biological) evolution to select from.

The learnability of languages in the model is ranked in terms of the number of parameters that must be updated to converge on the target grammar and also in terms of the maximum number of parameters which must be updated for a single trigger given an optimal presentation sequence of triggers. This ranking is calculated by assuming a learner with all parameters unset initially (see section 4.1 below). However, the ranking can also be made more dynamic by recalculating it for different potential initial p-settings.

3 The Evolutionary Simulation

The computational simulation supports the evolution of a population of Language Agents (LAgts), similar to Holland's (1993) Echo agents, but equipped with a LAD, as decribed in section 2, and a simple sentence generator based on random generation of a trigger / sentence type from the LAgt's current language (if any). LAgts generate and parse sentence types compatible with their current p-settings. They participate in linguistic interactions which are communicatively successful if their p-settings are compatible. Compatibility is defined in terms of the ability to map from a given SF to the same LF, rather than in terms of sharing of an identical grammar.[11] LAgts are either learning a grammar or have completed learning and fixed on the grammar and associated language acquired at that point.

In experiments which utilize natural (biological) selection for LAgts, the relative fitness of a LAgt is a function of the proportion of its linguistic interactions which have been successful, and optionally of the learnability, expressivity and/or interpretability of the grammar(s)/language(s) used by that LAgt during a cycle of interactions. Thus, fitness is dependent on an agents' linguistic compatibility with other agents, creating frequency-dependent selection, and also potentially on the complexity of the grammar(s)/language(s) they are using. The fitness functions and their components are given in Figure 10 and Figure 11 below. Learnability is modelled in terms of the number of parameters which need to be set to acquire a target grammar, the highest number which may be reset for a single trigger, and the agent's success rate at correctly setting parameters. Learning time, and thus the time taken to achieve maximal communicative performance, can also be increased by additional maturational memory limits during learning. However, this cost may also be offset by the tendency of such limits to create a more optimal presentation of triggers to the learning procedure. Interpretability is modelled by parsing cost (i.e. parsability) measured in terms of mean working memory load created during an interaction cycle, according to the WML model of section 2.2. Expressivity is modelled (crudely) in terms of an additional cost for using a proper subset language of one of the 70 full languages defined by the grammar space. This cost is graded on the basis of the number of sentence types / triggers associated with the language, and is necessary because otherwise agents will tend to converge on less expressive languages with lower average working memory load costs and fewer parameters to learn. In general, the pressures created for learnability, parsability and expressivity are conflicting, creating the potential for complex interactions and trade-offs in the search for (locally) optimal languages.

An interaction cycle consists of a prespecified proportion of individual random interactions between LAgts, with generating and parsing agents also selected randomly. LAgts which have a history of mutually successful interaction and higher than average fitness can 'reproduce'. A LAgt can 'live' for up to ten interaction cycles, but may 'die' earlier if its fitness is lower than average. It is possible for a population to become extinct (for example, if all the initial LAgts go through ten interaction cycles without any successful interaction occurring), and successful populations tend to grow at a modest rate (to ensure a reasonable proportion of adult speakers is always present). LAgts learn during a critical period from ages 1-4 and reproduce from 3-10, parsing and/or generating any language learnt throughout their life.

During learning a LAgt can reset genuine parameters which either were unset or had default settings 'at birth'. However, p-settings with an absolute value (principles) cannot be altered during the lifetime of a LAgt. Successful LAgts reproduce at the end of interaction cycles by one-point crossover of (and, optionally, single point mutation of) their *initial* p-settings – ensuring neo-Darwinian rather than Lamarckian inheritance. (The encoding of p-settings allows the deterministic recovery of the initial setting.) Fitness-based reproduction ensures that successful and somewhat compatible p-settings are preserved in the population and randomly sampled in the search for (locally) optimal versions of universal grammar, including initial settings of parameters. Thus, although the learning algorithm *per se* is fixed, a range of alternative learning procedures can be explored based on the definition of

[11] P-setting compatibility implements a weak notion of communicative success. Thus, there is no Gricean entailment of successful transmission of speaker intentions, or of a shared interpretation. Consequently, the model builds in no assumptions about the function(s) of language, whether this be to influence others, communicate (mis)information, or whatever (see e.g. Pinker and Bloom, 1990; Keller, 1994:84f for insightful discussion).

Variables	Typical Values	
Initial Population Size		32
Interaction Cycle	Av. Interactions/LAgt	65
Simulation Run	Int. Cycles	50-10k
Crossover Probability		0.9
Mutation Probability		0/0.05
Learning	memory limited	yes
	critical period	yes
	(re)settable n	3-5
Migrations	lg distance	3
	per cycle	2
	not genetic	T

Figure 9: Typical Simulation Options

1. Generate cost: 1 (GC)
2. Generate subset language cost: 1-3 (GSC)
3. Parse cost: 1 (PC)
4. Parse failure cost: 1 (PF)
5. Parse memory cost: WML(st)
6. Parse/Generate success benefit: 1 (SI)
7. Parameter (re)set cost: 1 (PS)
8. Parameter (re)set success benefit: 1 (SPS)
9. Maximum (re)settable parameters: n (MSP)

Figure 10: Cost/Benefits per Interaction

the inital set of parameters, their initial settings, and on the number of (re)settable parameters per trigger. Figure 9 summarizes crucial options in the simulation giving typical values used in the experiments reported in section 4, Figure 10 shows the potential costs and benefits to a LAgt of each interaction, and Figure 11 the components used to define the full or partial fitness functions. (For calculation of parsability only successfully parsed sentence types are utilized, hence parse failures (PF) are subtracted from the total number of parse interactions for a LAgt.)

4 Preliminary Experiments

The computational model must have several properties to qualify as a useful simulation of the potential coevolution of language and the language acquisition device. Firstly, it must be clear that for the chosen grammar set, at least some learning procedures in the space of possibilities definable in terms of LAgts' p-settings, are able to learn these grammars given finite and feasible (positive) input. Secondly, learning LAgts should converge reliably on the language of a population of homogeneous adult LAgts to ensure the continuity of speech communities. This latter property of stability of language acquisition through the generations is partly a property of the learning procedure and partly of other factors, such as the ratio of

Fitness Components

Communicative Performance:	$\frac{SI}{GC+PC}$
Expressivity:	$\frac{GC}{GC+GSC}$
Learnability:	$\frac{1}{MSP} \times \frac{SPS}{PS}$
Parsability:	$\frac{PC-PF}{WML(s_1 \ldots s_n)}$
Full Fitness Function:	$w1(CP) \times w2(Exp.) \times w3(Lrn.) \times w4(Pars.)$

(predefined weights balance relative strength of conflicting pressures)

Figure 11: LAgt Fitness

learners to adult LAgts in the population and the pattern of interactions between adults and learners.

4.1 The Effectiveness of Some Learning Procedures

Two learning procedures were predefined – a default learner and an unset learner. These LAgts were initialized with p-settings consistent with a minimal inherited CGUG consisting of application and the NP and S atomic categories already present. All the remaining p-settings were genuine parameters for both learners. The unset learner was initialized with all these unset (?), whilst the default learner had default settings for the parameters **gendir** and **subjdir** and **argorder** which specify a minimal SVO right-branching grammar. Both were able to update up to 5 parameters per trigger. The unset learner represents a 'pure' principles-and-parameters learner with innate knowledge of the noun-verb distinction and their (predicate-argument) mode of combination. This corresponds to what Deacon (1997) identifies as the minimal syntactic knowledge required to support his neo-Piercian concept of symbolic reference and, therefore, what he argues has been genetically assimilated. The default learner is modelled on Bickerton's (1984) Bioprogram Hypothesis, representing, additionally, a language learner with a preference for SVO predominantly right-branching syntax.

The two types of predefined learner LAgt were tested against an adult LAgt initialized to randomly generate triggers from one of seven full languages in the set which are close to an attested language; namely, "English" (SVO, predominantly right-branching), "Welsh" (SVOv1, mixed order), "Malagasy" (VOS, right-branching), "Tagalog" (VSO, right-branching), "Japanese" (SOV, left-branching), "German" (SOVv2, predominantly right-branching), "Hixkaryana" (OVS, mixed order), and one very rare, if not unattested, full OSV language with left-branching syntax. In these tests, a single learner interacted with a single adult. After every ten interactions, in which the adult randomly generated a trigger and the learner attempted to parse and/or learn from it, the state of the learner's p-settings was examined to determine whether the learner had converged on the same grammar as the adult. Table 1 shows the number of such interaction cycles (i.e. the number of input sentences to within ten) required by each type of learner to converge on grammars for each of the eight languages with probability ($p \geq 0.99$). These figures are each calculated from 1000 trials to a 1% error rate; they suggest that, in general, the default learner is more effective than the unset learner. However, for the OVS language (OVS languages represent 1.24% of the world's languages, Tomlin, 1986), and for the extremely rare OSV language, the default (SVO) learner is less effective. So, there are at least two learning procedures in the space defined by the model which can converge with high reliability on some of the full grammars in this

Learner	**Language**							
	SVO	SVOv1	VOS	VSO	SOV	SOVv2	OVS	OSV
Unset	60	80	70	80	70	70	70	70
Default	60	60	60	60	60	60	80	70

Table 1: Effectiveness of Two Learning Procedures

set. Stronger conclusions require either exhaustive experimentation or theoretical analysis of the model of the type undertaken by Gibson and Wexler (1994) and Niyogi and Berwick (1995).

4.2 Acquisition Stability

The simulation employs random interactions within a population, some of whom will be learners. Thus, a proportion will involve learning LAgts interacting with each other or generating input for adult LAgts, before they have converged on the target language. Even in a homogeneous adult LAgt environment with a critical period for learning, if the proportion of learners to adults in the population is too high, the learners will not converge to the target language as the distribution of sentence types becomes more skewed towards those of subset languages. Two series of 50 interaction cycle simulations were run each initialized with either 32 adult unset learners or 32 adult default learners all speaking one of the eight languages described above. LAgts reproduced (without mutation) and died as described in section 3. However, given the p-settings of the initial population, LAgts were only able to reproduce further unset or default learners or hybrid learners generated via crossover of these two learning strategies. Each condition was run ten times.

In all the runs, the population continued to speak the original language and learners reliably converged to that language by adulthood. Thus, any subset language speakers in the population at the end of an interaction cycle were without exception learners. In these runs, the proportion of adults never fell below 60%, the mean number of interactions per cycle for each agent was 65, and the levels of reproduction and death relative to population size were tuned to ensure stability. Similar tests were done using variant fitness functions not taking account of memory limitations in learning and/or parsing with identical results. Briscoe (1998) gives further details of experiments to test and tune the (potential) stability of the simulation model.

5 Linguistic Selection

To demonstrate linguistic selection for more (locally) optimal grammars/languages, a number of experiments were undertaken with genetically-invariant populations of LAgts operating in a continuously heterogeneous linguistic environment, providing the variation on which linguistic selection could work. In reality, variation is generated by language contact and borrowing, linguistic innovation, reanalysis during learning, and so forth (see e.g. Harris and Campbell, 1995). In the simulation, this was modelled by introducing additional adult LAgts with a different full language at regular intervals or by initializing the simulation with two genetically-identical adult groups speaking different full languages. That is, all variation is a consequence of 'population movement'. Language change occurs when learners converge preferentially on one or other language, or a mixture, or a subset, whilst exposed to data from more than one source grammar. There is also an increased possibility

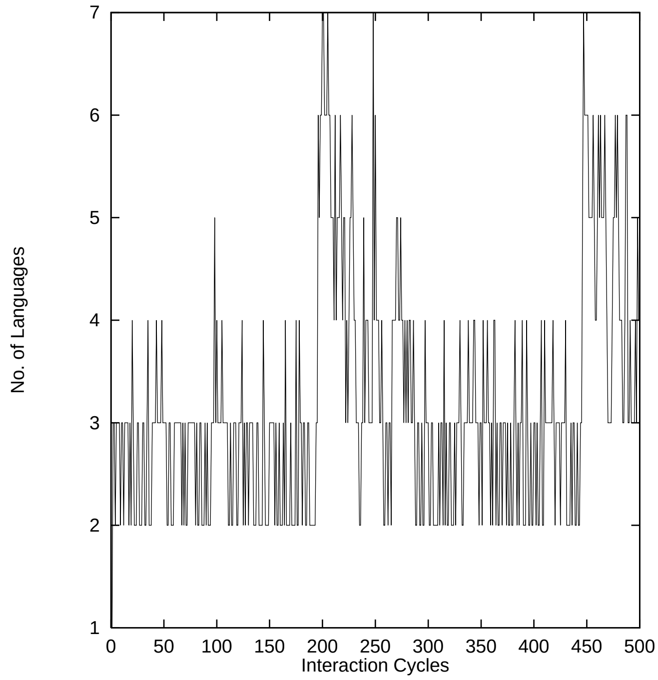

Figure 12: Number of languages in a typical run with migrations

of misconvergence to a grammar not exemplified in the data when the (uniform) distribution of triggers from a single source is skewed by the presence of several sources. This is particularly truw for parameters with default initial settings.

5.1 Linguistic Selection with Migrations

In this series of experiments, approximately one third additional adults were added to the population at regular intervals, all speaking the same new full language to ensure that the new language had a reasonable chance of surviving a number of cycles and thus influencing learners. LAgts added in this fashion had identical initial p-setting configurations as the existing population, so no genetic variation resulted. The maximal 'distance' between an existing dominant language and the new language was three parameters. 'Migrations' of this type occurred every other cycle provided that a clearly dominant language had emerged at the end of the previous cycle. Thus, migrations ensure a constant source of linguistic heterogeneity throughout a simulation run. The amount of variation introduced was tuned to the maximum consistent with the population maintaining a mean communicative performance level of 90% or better. After the first interaction cycle in all runs with migrations there are always two or more language variants present in the linguistic environment at any one time. Figure 12 plots the number of languages in the run with full LAgt selection discussed below.

In the first set of experiments, no fitness function was utilized, 500 cycle runs were used (approximately 125 generations of LAgts), and all LAgts were unset learners, as defined in section 4.1. LAgts reproduced randomly with no regard for communicative performance or the nature of the language they utilized. However, because all LAgts were using an effective learning procedure, because the simulation was initialized with a single full language, and because the amount of linguistic

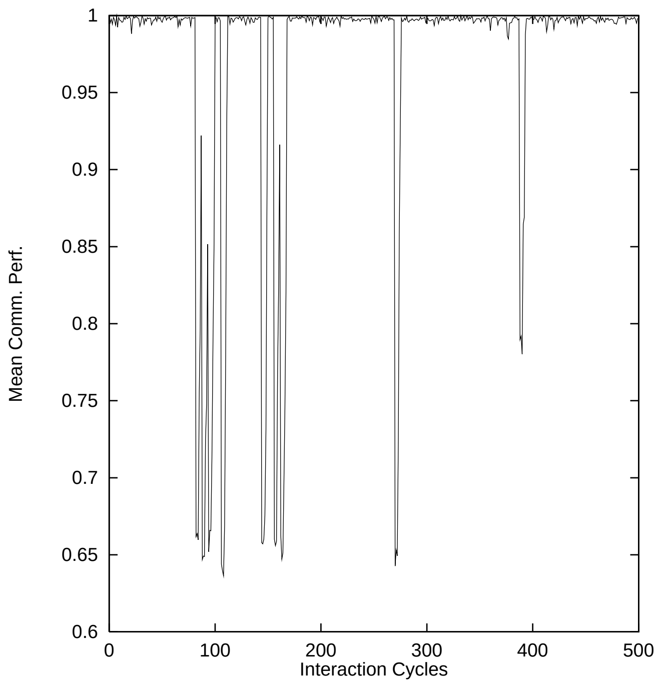

Figure 13: Communicative performances with random selection and migrations

variation was controlled, in all runs communicative performance averaged over 90%. This is plotted in Figure 13 for a typical run – dips correspond to points where migrations occurred. The overall mean costs of the languages adopted by the population were reduced during the course of this and other runs via linguistic selection for learnability, as illustrated in Figure 14. The figure plots an integrated measure for the mean learnability, parsability and expressivity of the languages present in each interaction cycle, and also breaks this down into the three components of memory load, generate subset cost and learning cost, so it can be seen clearly that the population is optimizing leanability at the expense of expressivity. In this and other runs with random LAgt selection, the population selected subset languages, which are less expressive but more easily learnable as they require fewer parameters set. As memory load plays a role in learnability via the filtering of triggers, often, but not in every case, parsability was also selected for. Similar results were obtained from all ten runs.

These results show that linguistic selection can occur without any *natural* selection for LAgts whatsoever. The bias of the learning procedure which the LAgts use is enough to create a process of selection for the most learnable languages. Kirby (1996) explores in detail this form of linguistic selection as languages, or more accurately triggers, pass repeatedly through the 'bottleneck' of language acquisition. Essentially, triggers compete for learners and those which are more able to pass through the filter of the learning procedure will set more parameters in more learners. In this way languages will over time adapt to the language learning procedure. Kirby argues that, on the assumption that parsability is *identical* to learnability, languages will, therefore, evolve to be optimally parsable, and demonstrates that this form of linguistic selection can predict statistical constituent order universals without the need for any natural selection for LAgts.[12] One weakness of this po-

[12]Briscoe (1998) discusses Kirby's position and argues that learnability and parsability are not

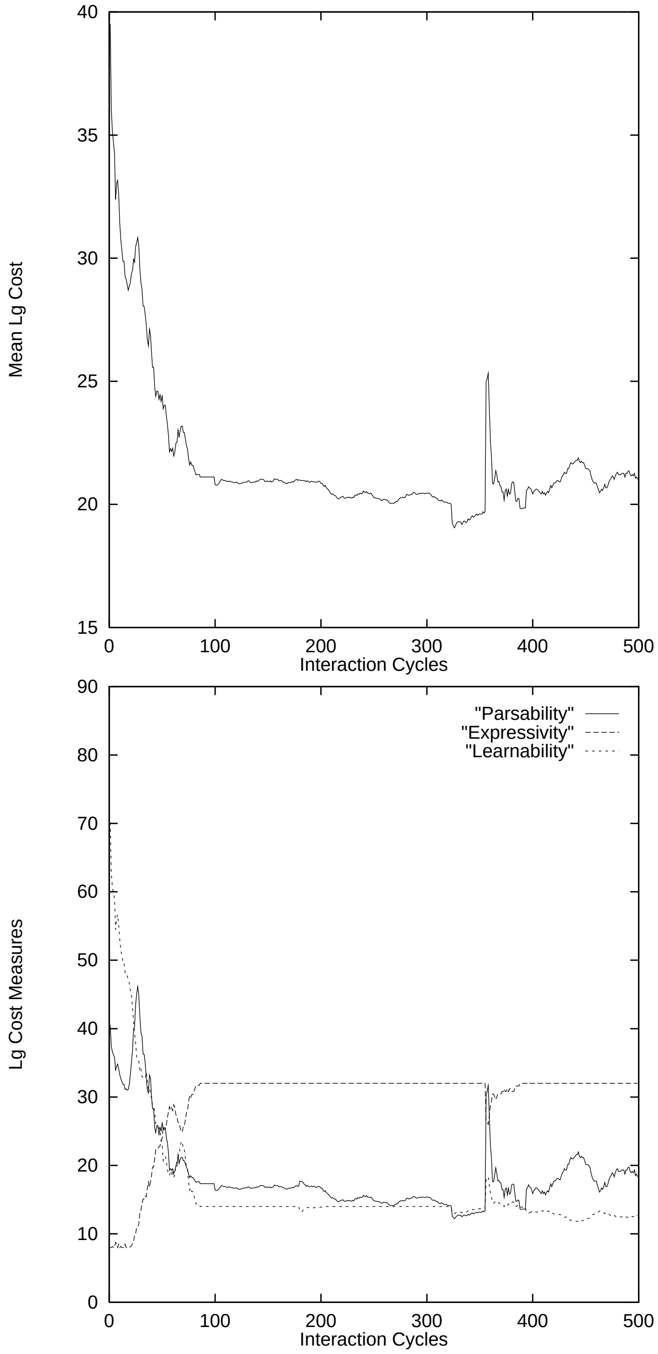

Figure 14: Language costs with random selection and migrations

sition is that Kirby only models differential learning between competing variants. Once a more complete learning procedure is defined, the possibility of simply *not* learning arises, and thus the possibility of converging on a subset language. This is exactly what is seen in runs of the simulation model without natural selection for LAgts – there is no pressure for LAgts to prefer a more expressive, and thus costly, language, so, even if the population is initialized to use such a language it soon selects for subset languages. A counteracting pressure for expressivity is needed to prevent this tendency.

Other runs were performed using communicative performance, memory load, generate subset costs and/or learning costs as a component of the fitness function on LAgt reproduction. In the runs where expressivity was a component of selection, the population did not converge on subset languages despite the linguistic variation in the learning environment created by frequent migrations. When the full fitness function was utilized, LAgts' mean fitness typically did not vary greatly, except where migrations removed them temporarily from a (local) optimum. The mean language costs for parsability, learnability and expressivity displayed in Figure 15 demonstrate consistent linguistic selection for more easily learnable and parsable full languages. This is typical of such runs where full natural selection for LAgts is utilized.[13] Comparing this with Figure 14 above demonstrates the contrast with linguistic selection and no natural selection.

5.2 Linguistic Selection between Language Pairs

In more circumscribed experiments, linguistic selection for more parsable and/or more learnable languages, and the interplay between these two pressures, can be demonstrated directly. For example, a series of 50 cycle simulations were run in which the population was initialized with equal numbers of default learner adult LAgts speaking two different full languages which contrasted in learnability and/or parsability. There were no differences in the initial p-settings in the population and no genetic variation.

For example, the population was initialized with equal numbers of default learner LAgts speaking either SOV or SOVv2. SOVv2 has a slightly lower mean memory load, and thus parsability, than SOV (largely because the freer constituent ordering options of Japanese relative to German are not modelled effectively in "Japanese" (see e.g. Hawkins, 1994)). Figure 16 shows the languages which emerge during one run with the full fitness function. SOVv2 comes to dominate the population after 5 interaction cycles. The other language which persists, SOVv2-N, is a subset language spoken by learners of SOVv2. SOVv2-GWP-COMP is also a subset language of SOVv2 so the 'recurrence' of this language at cycle 45 just reflects presence of one or two less successful learners at the end of an interaction cycle. The other non-v2 languages are eliminated within the first 5 interaction cycles. All runs exhibited the same clear effect. However, with parsability not a factor in LAgt fitness, the opposite result was obtained – in all runs SOV came to dominate with SOV-N(-GWP-COMP) subset languages, again spoken exclusively by learners. These clear opposing outcomes illustrate that parsability can be a factor in linguistic selection.

identical.

[13]In some cases, migrations still cause the population to settle on a less optimal language, though this is far less frequent when full selection for LAgts is utilized. The use of random interaction between LAgts idealizes a vast range of sociolinguistic factors which influence selection between linguistic variants, such as the prestige, charisma, economic power or ideology of the speakers of the variants, and so forth. In reality, these factors probably significantly outweigh considerations of selection for parsability or learnability in many situations; for example, where the migrants are conquering invaders. In addition, the simulation does not address differences in death rates between linguistic groups due to disease, genocide, and so forth. Dixon (1997) and Pullum (1981) provide an extended discussion of such factors.

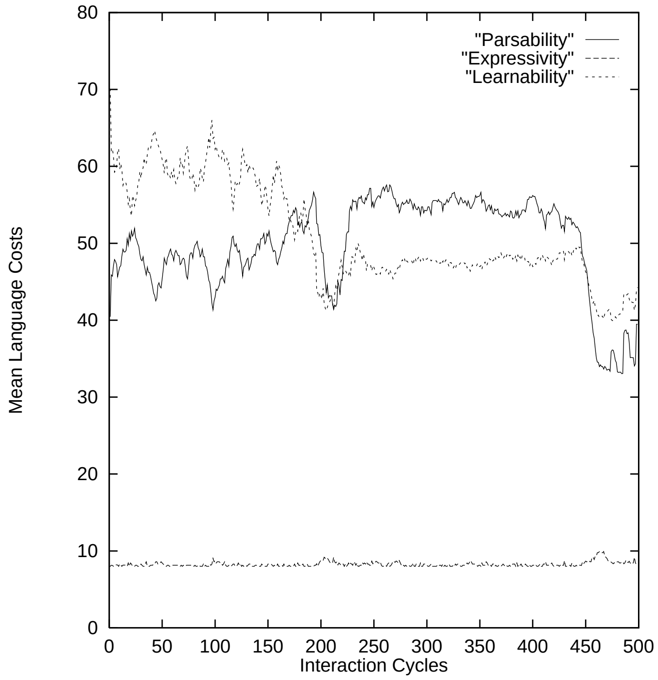

Figure 15: Language costs with natural selection and migrations

It is highly likely that even in the absence of some ranking of languages in terms of the learning or parsing procedures, populations will select a single language, as this increases individual LAgt's communicative performance. However, as SOV is consistently selected in all such runs when parsability is not a factor, this is most likely to be because the learning procedure for SOVv2 requires the setting of one further parameter over SOV. Thus, in the case of these two languages, ease of parsability for both learners and users creates greater overall linguistic selection pressure than that created by the need to set one less parameter. This result is consistent with the claim made by Hawkins (1994) that ease of processing is a factor in the distribution of constituent order types in the world's languages (see Briscoe 1997b, 1998 for further examples of linguistic selection and demonstrations that memory limitations during learning are a significant factor.)

6 Coevolution

The experiments of section 5 demonstrated evolution of language on a historical timescale within a genetically-invariant population of LAgts. To demonstrate coevolution, it is also necessary to allow the LAgt population to evolve. LAgts' initial p-settings were varied by allowing mutation of a single element of a LAgt p-setting (with probability 0.05) during LAgt reproduction. Successful variant initial settings could then propagate through the population via single-point crossover (with probability 0.9). The full fitness function was used.

6.1 Coevolution without Migrations

In the first series of such experiments, the initial population were all unset learner adults speaking one of the clearly-attested full languages. Mutations could intro-

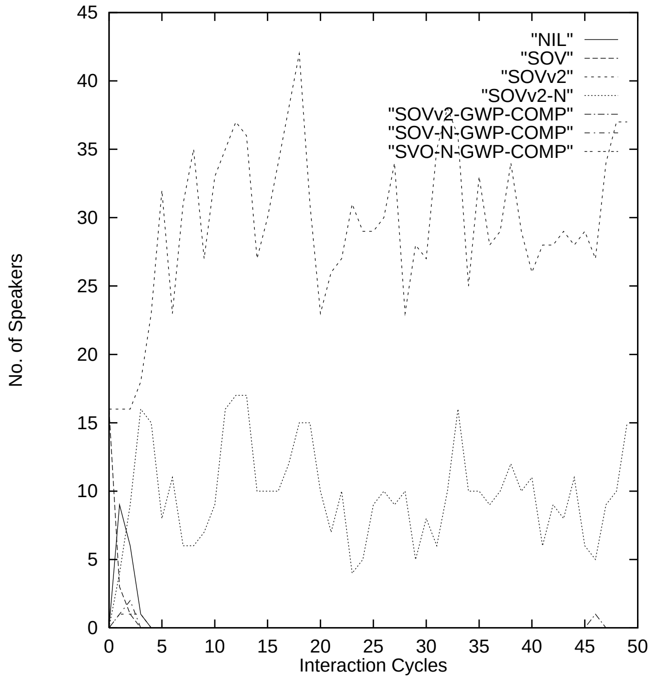

Figure 16: Selection for SOVv2 over SOV

duce a language variant by altering the default value of a parameter. However the degree of linguistic variation in such runs was typically minimal with populations sampling around 5 closely-related full languages over 500 interaction cycles. In these runs, the populations always evolved towards initial p-settings which enhanced the learnability of the dominant language in the environment. Figure 17 shows mean LAgt fitness for one such population – a measure of 0–1 integrating mean communicative performance, and mean LAgt learning costs, memory load and subset language costs (see section 3 above) – and the relative proportions of default parameters, unset parameters and principles in the same population. In all such runs, the proportion of default parameters grew at the expense of unset ones, with default values reflecting the language of the environment. In addition, the mean number of (re)settable parameters per trigger fell until typically by cycle 500 the whole population converged on a value of 2 or 3, depending on the language in the environment. Consequently, LAgt fitness improved over the course of the run as a result of reduction in learning costs, whilst mean parsability, expressivity and communicative performance remained roughly constant. This is a clear example of genetic assimilation in which LAgts are evolving to be able to acquire the dominant language more effectively. By replacing unset with default parameters with initial settings compatible with the dominant language, the LAD is evolving an accurate language-specific learning bias which simplifies the acquisition of this language. At the same time, this bias itself will alter the relative complexity of languages by altering their relative learnability. However, linguistic variation in these simulations is very limited, caused only by occasional failures of convergence, mutations of default parameter values or mutations of parameters to principles. Consequently, the rate of linguistic change is very slow, creating a constant selection pressure for genetic assimilation to work on.

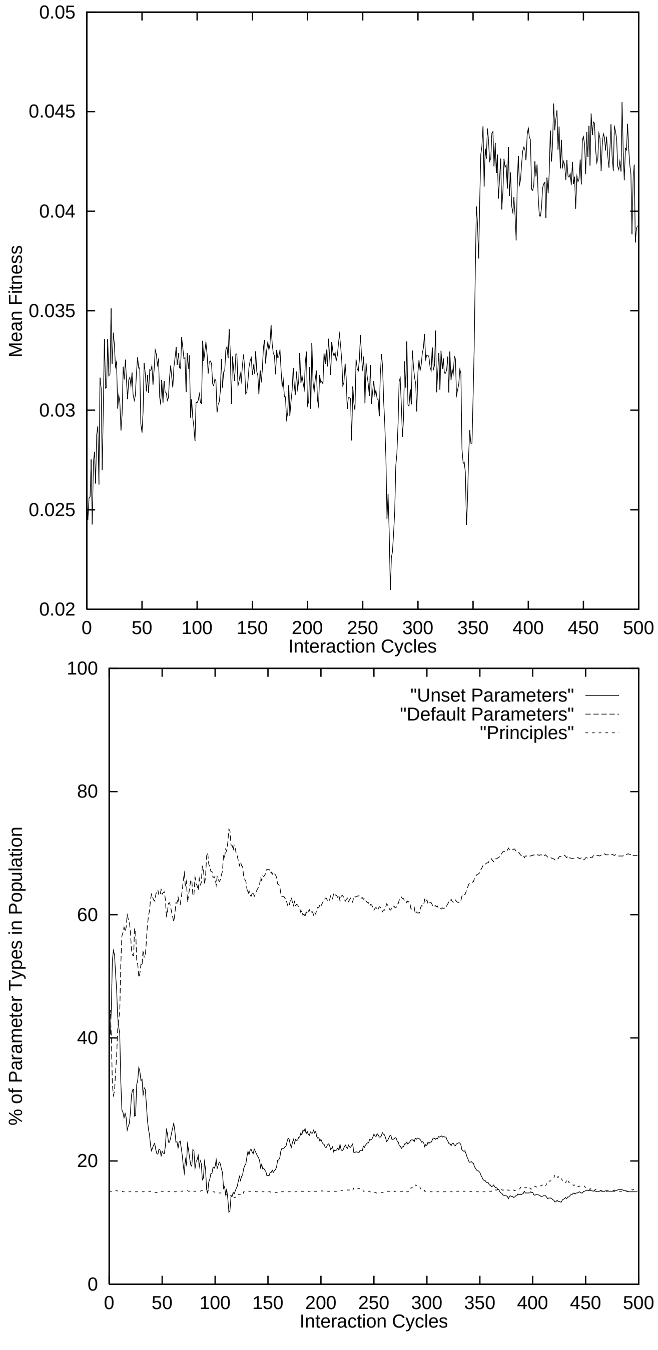

Figure 17: Mean fitness and p-setting types during coevolution without migrations

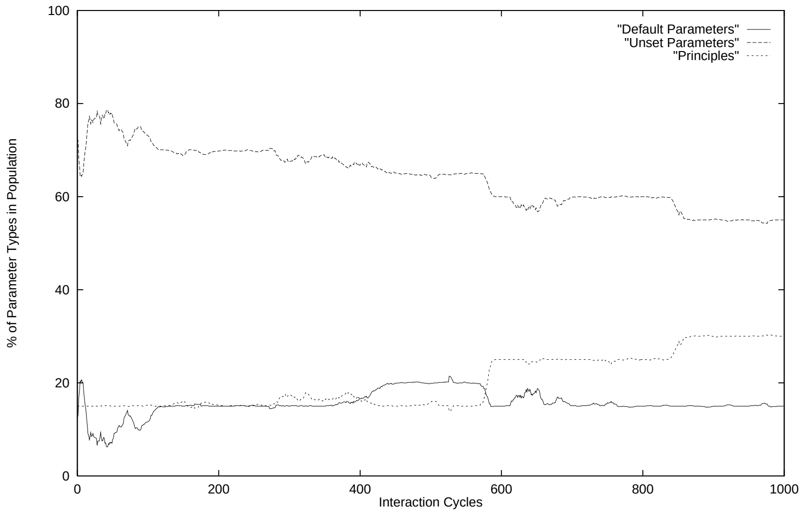

Figure 18: Proportions of p-setting types with migrations

6.2 Coevolution with Migrations

To see whether genetic assimilation would occur with maximal linguistic variation consonant with communicative success, a second series of experiments was run identical to those described above, except that migrations occurred as often as was compatible with a mean 90% communicative performance over the entire 1000 cycle run. Figure 18 shows the relative proportions of default parameters, unset parameters and principles for one such run.

In these runs, LAgts still evolved LADs which improved learnability despite the fact that typically the dominant language changes about 20 times and approximately 50 languages are sampled by the population. However, as in the run shown, there was a greater tendency to replace unset parameters with principles rather than just with default parameters. Over 10 such runs, the proportion of unset parameters declined by a mean 35% leading to around 50% of p-settings being principles or default parameters with roughly an equal number of new principles and default parameters. In other respects, results were identical to the first series of runs with LAgt fitness improving as a consequence of reduced learning cost. However, the greater degree of linguistic variation also allowed more linguistic selection for more optimal languages.

The replacement of unset parameters by principles is an example of the type of genetic assimilation which Pinker and Bloom (1990) envisage, in which the class of learnable languages is (further) constrained by the LAD in the interests of enhanced learnability. Thus, in these runs we see examples of genetic assimilation of both learning biases (defaults) and constraints (principles), albeit at a slower rate than when the linguistic environment was more constant. To see how long genetic assimilation would continue in a heterogeneous linguistic environment, several simulations were run for 10,000 cycles. In these, the mean decline in the proportion of unset parameters was 55% with 65% of p-settings being principles or parameters at the end of the runs. Once again approximately half of the replaced unset parameters were default parameters. Plots of the proportions of each type of parameter show a rate of genetic assimilation for default parameters and principles which slows through these runs. Finally, in similar runs with populations initialized to reproduce learn-

Old		New		
PS-Type	P-value	PS-Type	P-value	Expected Fitness
Absol	F	Def	F	$f >$
Absol	F	Def	T	$f >$
Absol	F	Absol	T	$f >$
Absol	F	Unset	?	$f >$
Absol	T	Def	F	$f <$
Absol	T	Def	T	$f =$
Absol	T	Absol	F	$f <$
Absol	T	Unset	?	$f <$
Def	F	Absol	F	$f <$
Def	F	Absol	T	$f >$
Def	F	Def	T	$f >$
Def	F	Unset	?	$f =$
Def	T	Absol	F	$f <$
Def	T	Absol	T	$f =$
Def	T	Def	F	$f <$
Def	T	Unset	?	$f <$
Unset	?	Absol	T	$f >$
Unset	?	Absol	F	$f <$
Unset	?	Def	T	$f >$
Unset	?	Def	F	$f =$

Table 2: P-setting Transitions and Fitness Effects

ers with all default parameters with values appropriate to the initial language, the population invariably evolved away from such 'total' genetic assimilation towards p-settings containing some unset parameters. Therefore, we can conclude that there is a limit to genetic assimilation in the face of such linguistic variability.

6.3 Discussion

Why then do we see (partial) genetic assimilation even in the face of great linguistic heterogeneity and rapid linguistic change? And why, when change is rapid, is there a greater tendency for the assimilation of principles as well as default initial parameter values? Firstly, consider the possible mutations which can occur within a p-setting and their expected fitness effects; Table 2 catalogues the possible transitions of individual initial p-settings (which can be created by a single mutation) and their expected fitness cost/benefit in terms of the 'truth/falsity' (T/F) of the resulting p-setting value in the current linguistic environment. The fitness cost/benefit is based on the expected effect on learnability. It is clear that any transition from a false principle (i.e. one which is inconsistent with the current dominant linguistic environment) will incur a fitness benefit, because it will allow a LAgt a chance to learn the dominant language. By contrast, a transition from a true principle to anything other than a true default will have a learning cost because it will either render learning impossible or increase the number of parameters to be (re)set. Likewise, no transition from a true default creates any benefit and three incur a cost. Three transitions from a false default incur learning benefit, only a transition to a false principle incurs a cost, by making learning impossible. Transitions from unset parameters to true default parameters or true principles are beneficial, whilst a false principle, as always, incurs a (fatal) cost. The transition to a false default incurs no cost (or benefit) because during learning it still takes one parameter (re)setting to obtain the correct value.

It should be clear from this discussion, that what we would expect to evolve is a population with correct principles, predominantly correct default initial parameter values, and possibly a minority of unset and/or default incorrect parameters. In an unchanging linguistic environment, we would expect the population to eventually fix on all true principles or default parameters. However, in all the experiments reported above the linguistic environment is never entirely homogeneous or static. Therefore, the 'truth/falsity' of a p-setting is an approximation: a value may be predominantly correct in the current environment given the dominant language, but become predominantly or completely incorrect over succeeding cycles (and vice versa). Whether an initially beneficial mutation achieves fixation, or even predominance, within the population will depend not only on the initial benefit it offers the mutated LAgt, but also on the continuing benefit to its descendents. It is here that coevolutionary effects will occur; for example, as a predominantly correct principle spreads through the population, it will create greatly increased linguistic selection for languages which obey this principle. This, in turn, will increase the chance that the principle will go to fixation in the population, rendering languages which do not obey the principle unlearnable. In a changing environment, we would expect there to be a preference for default parameters over absolute principles, because an initially predominantly correct principle which spread through a proportion of the population would incur a high, possibly fatal, cost to them if it subsequently became (predominantly) incorrect. By contrast, a default parameter which becomes incorrect, incurs no more cost than an unset parameter, given the learning procedure assumed in the current simulation. There does appear to be a bias towards genetic assimilation of default parameters in the experiments reported above with lowish rates of linguistic change. The migration mechanism, used in the simulation for introducing linguistic variation, tends to reinforce the status of principles which have spread through more than 50% of the population and accelerate their fixation (because it introduces adults with identical initial p-settings to those of the existing majority). So, further experiments are needed to explore the degree of genetic assimilation of principles as opposed to default parameters with variant migration mechanisms.

In the experiments reported above with mean 90% communicative performance, the fastest observed rate of change from one dominant language variant to another was 4 interaction cycles. The fastest observed rate at which a mutation in a p-setting reached fixation was 43 cycles. This suggests that linguistic evolution of grammatical parameters was only about one order of magnitude faster than 'genetic' evolution of p-settings. Increasing the speed of linguistic change would have resulted in a decrease in communicative performance below what is assumed reasonable in a language community. Nevertheless, the simulation tells us nothing about the true relative rates of linguistic and biological evolution – increasing the size of the population (in the simulation or real world) would, for example, slow down biological evolution. But, there can be no certainty about the size of the ancestor population in which the LAD might have evolved. Deacon (1997:329) suggests that linguistic evolution is 'many' orders of magnitude faster than biological evolution, arguing that languages can change their major grammatical properties over thousands of years (historically, 1-2 millenia for the types of constituent order properties modelled here). However, the time taken for a major grammatical change and the time taken for biological evolution will depend critically on population size, geographical dispersal, diffusion rates of genes and of variant grammatical forms, and so forth. In the simulation runs with rapid linguistic change, typically 2-3 major grammatical changes propagate through the population every 50 interaction cycles. Therefore, default parameters and absolute principles are being genetically assimilated and going to fixation in the population typically in the face of several such major linguistic changes.

The key to understanding why genetic assimilation is still likely to occur is that the sample space of possible grammars and associated languages is vastly larger than the number of grammars which can be sampled by a population in the time taken for a principle or default parameter to go to fixation. In the simulation, there are under 300 languages and only 70 distinct full languages, based predominantly on constituent order differences. Therefore, in the time taken for a p-setting to go to fixation typically around 5% of the space of grammars might be sampled. This means that 95% of the selection pressure for genetic assimilation of grammatical information remains constant at any one time. In his discussion, Deacon (1997:329f) ignores the issue of the space of grammatical possibilities and the degree to which this can be sampled in the time required for biological evolution. It is impossible to estimate the real size of this space properly, but few linguists would probably balk at the idea that 30 independent binary grammatical parameters will be required to capture the differences between the world's languages in an account of universal grammar. Given this, there are billions of distinct grammars to explore.[14] This guestimate is based on the existence of an evolved LAD. Most linguists would argue that prior to the existence of the LAD, the space of possible grammars was infinite, and some believe that it remains so now (e.g. Pullum, 1983). Given the likely vast, even if finite space, to be explored, rapid changes in the tiny subset of potential grammatical systems which the ancestral linguistic population was exposed to could not prevent genetic assimilation on the basis of the many potential systems which were not sampled; for example, all those potential grammatical systems which would have resulted in arbitrarily intersecting dependencies between constituents (see section 1).

7 Conclusions

The model of a language (learning) agent, described in section 2, embedded in the (co)evolutionary simulation described in section 3 demonstrates that, given the assumptions behind the model of a LAgt and the (co)evolutionary scenario, linguistic selection for more parsable, learnable and expressive languages will occur. Expressivity, though modelled crudely, is a critical factor in the simulation, since without it LAgts converge to subset languages which are easier to learn and to parse. Once expressivity is a component of fitness, LAgts converge to (locally) optimal languages which represent a trade-off between memory-efficient constituent order configurations and ones which require the setting of fewer parameters, both per trigger and overall, with respect to the dominant learning procedure. Thus, the model demonstrates that embedding a generative model of a LAgt in a population of LAgts leads naturally to an account of language in which idiolects are well-defined stringsets, but languages are complex adaptive systems.

When LAgts' p-settings can vary, under all experimental conditions genetic assimilation of more 'informative' default parameters or absolute constraints occurs. These not only improve the learnability of the dominant language(s) by incorporating learning biases and constraints into the language learning procedure, they also alter the nature of the linguistic selection pressure exerted by the evolving population of LAgts. As a more constrained and biased language acquisition device is genetically assimilated, so the class of learnable languages becomes more constrained and the ranking of learnability amongst them alters to reflect the evolving biases in the LAD. This effect is observed even in the face of as rapid grammatical

[14] This is another reason why ordered setting of parameters is probably the only feasible manner in which a learner can converge to a target grammar given realistic data (see also Clark, 1992).

change as is consistent with the maintenance of a successfully communicating population, because only a tiny subset of the range of grammatical possibilities can be sampled in the time it takes for a p-setting to go to fixation. Nevertheless, with linguistic heterogeneity, genetic assimilation is asymptotic and some parameters remain unset. Thus, the model demonstrates a clear coevolutionary dynamic between the historical evolution of language and the biological evolution of the LAD.

The conclusions drawn from a simulation model of the type presented here must remain highly conditional. Not only the assumptions behind the model but also the many contingent, accidental or chance factors in the actual, but prehistoric, evolution of language and its users may undermine the results. Nevertheless, models of this type have heuristic value in guiding us towards hypotheses which can then be tested by other means; for example, claims about the effect of working memory on parsing are testable, in principle, via psycholinguistic experimentation or typological investigations, even though claims about the prehistoric development of language are not. Furthermore, such models can be used to evaluate evolutionary theorizing about language which does not utilize a simulation methodology and to expose implicit and inadequate assumptions in such theorizing; for example, Deacon's (1997:329) arguments from rapid relative linguistic change to the implausibility of genetic assimilation of grammatical knowledge.

Acknowledgements

I would like to thank Bob Berwick, Ann Copestake, Gerald Gazdar, Miriam Eckert, Hans van Halteren, Jim Hurford, Simon Kirby, David Milward, and Geoff Pullum for helpful comments on various aspects of this work. Hans van Halteren's careful reading and comments on several drafts greatly improved the presentation. I am entirely responsible for any remaining errors or lack of clarity.

References

Ackley, D. and M. Littman (1991) 'Interactions between learning and evolution' in C. Langton, C. Taylor, J. Farmer and S. Rasmussen (ed.), *Artificial Life II,* Addison-Wesley, Redwood City, Ca., pp. 487–509.

Baddeley, A. (1992) 'Working Memory: the interface between memory and cognition', *J. of Cognitive Neuroscience, vol.4.3,* 281–288.

Baldwin, J.M. (1896) 'A new factor in evolution', *American Naturalist, vol.30,* 441–451.

Bertolo, S. (1995) 'Maturation and learnability in parametric systems', *Language Acquisition, vol.4.4,* 277–318.

Bickerton, D. (1984) 'The language bioprogram hypothesis', *The Behavioral and Brain Sciences, vol.7.2,* 173–222.

Brent, M. (1996) 'Advances in the computational study of language acquisition', *Cognition, vol.61,* 1–38.

Briscoe, E.J. (1987) *Modelling Human Speech Comprehension: A Computational Approach,* Chichester, UK: Ellis Horwood / New York: Wiley.

Briscoe, E.J. (1997a) 'Co-evolution of language and of the language acquisition device', *Proceedings of the 35th Assoc. for Comp. Ling.,* Madrid, pp. 418–427.

Briscoe, E.J. (1997b, submitted) 'Language Acquisition: the Bioprogram Hypothesis and the Baldwin Effect', *Language,*

Briscoe, E.J. (1998, in prep.) *Learning Language and Evolving Language,* University of Cambridge, Computer Laboratory, m.s..

Casti, J.L. (1994) *Complexification: Explaining a Paradoxical World through the Science of Surprise,* Harper Collins, New York.

Cecconi, F., Menczer, F. and Belew, R.K. (in press) 'Maturation and the evolution of imitative learning in artificial organisms', *Adaptive Behaviour, vol.4.1,*

Chomsky, N. (1981) *Government and Binding,* Foris, Dordrecht.

Chomsky, N. (1995) *The Minimalist Program,* MIT Press, Cambridge, Ma..

Clark, R. (1992) 'The selection of syntactic knowledge', *Language Acquisition, vol.2.2,* 83–149.

Wexler, K. and Culicover, P. (1980) *Formal Principles of Language Acquisition,* MIT Press, Cambridge, Ma..

Curtiss, S.R. (1988) 'Abnormal language acquisition and the modularity of language' in F. Newmeyer, vol. 2 (ed.), *Linguistics: The Cambridge Survey,* Cambridge University Press, Cambridge.

Cziko, G. (1995) *Without Miracles: Universal Selection Theory and the Second Darwinian Revolution,* MIT Press, Cambridge, Ma..

Darwin, C. (1871, 1989) *The descent of Man, Selection in Relation to Sex,* Pickering and Chatto, London.

Dawkins, R. (1982) *The Extended Phenotype,* Oxford University Press, Oxford.

Dawkins, R. (1983) 'Universal Darwinism' in D.S. Bendall (ed.), *Evolution: From Molecules to Men,* Cambridge University Press, Cambridge, pp. 403-425.

Dawkins, R. (1989) *The Selfish Gene,* Oxford University Press, Oxford, 2nd edition.

Deacon, T. (1997) *The Symbolic Species: Coevolution of Language and Brain,* MIT Press, Cambridge, Ma..

Dennett, D. (1991) *Darwin's Dangerous Idea: Evolution and the Meanings of Life,* Simon and Schuster, New York.

Dixon, R. (1997) *The Rise and Fall of Languages,* Cambridge University Press, Cambridge.

Durham, W. (1991) *Coevolution, Genes, Culture and Human Diversity,* Stanford University Press, Pala Alto, Ca..

Elman, J. (1993) 'Learning and development in neural networks: the importance of starting small', *Cognition, vol.48,* 71–99.

Frank, R. and Kipur, S (1996) 'On the use of triggers in parameter setting', *Linguistic Inquiry, vol.27.4,* 623–660.

Futuyma, D.J. and Slatkin, M. (1983) *Coevolution,* Sinauer Associates, Sunderland, Ma..

Gazdar, G. (1988) 'Applicability of indexed grammars to natural languages' in Reyle, U. and Rohrer, C. (ed.), *Natural Language Parsing and Linguistic Theories,* Reidel, Dordrecht, pp. 69–94.

Gibson, E. (1991) *A Computational Theory of Human Linguistic Processing: Memory Limitations and Processing Breakdown,* Doctoral dissertation, Carnegie Mellon University.

Gibson, E. and Wexler, K. (1994) 'Triggers', *Linguistic Inquiry, vol.25.3,* 407–454.

Gopnik, M. (1994) 'Impairments of tense in a familial language disorder', *J. of Neurolinguistics, vol.8,* 109-133.

Greenberg, J. (1966) 'Some universals of grammar with particular reference to the order of meaningful elements' in J. Greenberg (ed.), *Universals of Grammar,* MIT Press, Cambridge, Ma., pp. 73–113.

Harris, A.C. and Campbell, L. (1995) *Historical Syntax in Cross-Linguistic Perspective,* Cambridge University Press, Cambridge.

Harris, R. and Taylor, T.J. (1997) *Landmarks in Linguistic Thought,* vol1, Routledge, London.

Harvey, I. (1993) 'The puzzle of the persistent question marks: a case study of genetic drift' in S. Forrest (ed.), *Genetic Algorithms: Proc. of the 5th Int. Conference,* Morgan Kaufmann, San Mateo, Ca..

Hawkins, J.A. (1994) *A Performance Theory of Order and Constituency,* Cambridge University Press, Cambridge.

Hinton, G.E. and Nowlan, S.J. (1987) 'How learning can guide evolution', *Complex Systems, vol.1,* 495–502.

Hoffman, B. (1995) *The Computational Analysis of the Syntax and Interpretation of 'Free' Word Order in Turkish,* PhD dissertation, University of Pennsylvania.

Hoffman, B. (1996) 'The formal properties of synchronous CCGs', *Proceedings of the ESSLLI Formal Grammar Conference,* Prague.

Holland, J.H. (1993) *Echoing emergence: objectives, rough definitions and speculations for echo-class models,* Santa Fe Institute, Technical Report 93-04-023.

Hurford, J. (1987) *Language and Number,* Blackwell, Oxford.

Jablonka, E. and Lamb, M.J. (1995) *Epigenetic Inheritance and Evolution,* Oxford University Press, Oxford.

Joshi, A., Vijay-Shanker, K. and Weir, D. (1991) 'The convergence of mildly context-sensitive grammar formalisms' in Sells, P., Shieber, S. and Wasow, T. (ed.), *Foundational Issues in Natural Language Processing,* MIT Press, pp. 31–82.

Kauffman, S. (1993) *The Origins of Order: Self-Organization and Selection in Evolution,* Oxford University Press, New York.

Keller, R. (1994) *On Language Change: The Invisible Hand in Language,* Routledge, London.

Kirby, S. (1996) *Function, Selection and Innateness: The Emergence of Language Universals,* Doctoral Thesis, University of Edinburgh.

Kirby, S. (1997) 'Competing motivations and emergence: explaining implicational hierarchies', *Language Typology, vol.1,* 5–32.

Lascarides, A., Briscoe E.J. , Copestake A.A and Asher, N. (1995) 'Order-independent and persistent default unification', *Linguistics and Philosophy, vol.19.1,* 1–89.

Lascarides, A. and Copestake A.A. (1996, in press) 'Order-independent typed default unification', *Computational Linguistics,*

Lightfoot, D. (1991) *How to Set Parameters: Arguments from language Change,* MIT Press, Cambridge, Ma..

Maynard-Smith, J. (1993) *The Theory of Evolution,* 3rd ed., CUP, Cambridge.

Niyogi, P. and Berwick, R.C. (1995) 'A markov language learning model for finite parameter spaces', *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics,* MIT, Cambridge, Ma..

Ochs, E. (1982) 'Talking to children in Western Samoa', *Language in Society, vol.11,* 77–104.

Pinker, S. and Bloom, P. (1992) 'Natural language and natural selection' in J. Barkow, L. Cosmides and J. Tooby (ed.), *The Adapted Mind: Evolutionary Psychology and the Generation of Culture,* Oxford University Press, Oxford, pp. 451–493.

Pullum, G.K. (1981) 'Languages with object before subject: a comment and a catalogue', *Linguistics, vol.19,* 147–155.

Pullum, G.K. (1983) 'How many possible human languages are there?', *Linguistic Inquiry, vol.14.3,* 447-467.

Rambow, O. and Joshi, A. (1994) 'A processing model of free word order languages' in C. Clifton, L. Frazier and K. Rayner (ed.), *Perspectives on Sentence Processing,* Lawrence Erlbaum, Hillsdale, NJ., pp. 267–301.

Richards, R.J. (1987) *Darwin and the Emergence of Evolutionary Theories of Mind and Behaviour,* University of Chicago Press, Chicago, Il..

Rose, S. (1997) *Lifelines: Biology, Freedom, Determinism,* Penguin Press, London.

Roughgarden, J, (1983) 'The theory of coevolution' in D.J. Futuyma and M. Slatkin (ed.), *Coevolution,* Sinauer Associates, Sunderland, Ma., pp. 33–64.

Shieber, S. (1985) 'Evidence against the non-context-freeness of natural language', *Linguistics and Philosophy, vol.8,* 333–343.

Smith, N.V. and Tsimpli, I.A. (1991) 'Linguistic modularity? A case study of a 'savant' linguist', *Lingua, vol.84,* 315–351.

Steedman, M. (1988) 'Combinators and grammars' in R. Oehrle, E. Bach and D. Wheeler (ed.), *Categorial Grammars and Natural Language Structures,* Reidel, Dordrecht, pp. 417–442.

Steedman, M. (1996) *Surface Structure and Interpretation,* MIT Press, Cambridge, Ma..

Tomlin, R. (1986) *Basic Word Order: Functional Principles,* Routledge, London.

Waddington, C. (1975) *The Evolution of an Evolutionist,* Edinburgh University Press, Edinburgh.

Wood, M.M. (1993) *Categorial Grammars,* Routledge, London.