Part IV

Learnability and diversity: How did languages emerge and diverge?

Introduction to Part IV: Computer modelling widens the focus of language study

James Hurford

The chapters in this part demonstrate a remarkably unified and coherent vision of a new emerging approach to the study of language, despite being somewhat diverse in their investigative styles and specific target problems. There have been hints and socially isolated pockets of this vision around for decades, probably even for centuries, but the advent of powerful high-speed computing has given it the boost it needed to raise it above a hopeful ideology. The vision is that Language (deliberately conceived broadly, hence the capitalization) is a complex dynamic system resulting from the interplay of biological evolution of the species, ontogenetic development of individuals, and cultural processes shaping the histories of societies. The truth of this vision has been recognized even by those who have deliberately steered their scientific methodology away from its daunting implications. This is evident is such passages as the following:

[W]hat a particular person has inside his head [i.e. an idiolect] is an artifact resulting from the interplay of many idiosyncratic factors, as contrasted with the more significant reality of UG. ... [E]ach actual 'language' will incorporate a periphery of borrowings, historical residues, inventions, and so on which we can hardly expect to—and indeed would not want to—incorporate within a principled theory of UG (Chomsky 1981: 8).

The chosen research strategy of generative grammar has been to narrow the domain of enquiry as far as possible, excluding considerations from performance, irregularity, and 'periphery of grammar'. This narrow viewpoint is aptly characterized in Chapter 13 here, by Henry Brighton et al., as a 'Principle of Detachment'. They argue that there are good reasons to reject this Principle of Detachment, thus opening the doors (some might

say floodgates) to consideration of how factors other than innate individual psychological biases could influence the shape of languages. Orthodox generative linguists have also, to the puzzlement of many observers, typically refused to speculate about the evolutionary biological pressures shaping the hypothesized domain-specific Language Acquisition Device. In Chapter 14, Ted Briscoe boldly faces up to the theoretical problem of the possibility of interaction between biological evolution and the histories of languages in populations, with interesting results. Chapter 15, by Matthew Roberts et al., also breaks away from the narrow research approach, in arguing, very plausibly, that a striking feature of languages, their stubborn proneness to irregularity, actually results from the historical process of social transmission by learning, through a tight bottleneck (i.e. from an impoverished stimulus) over many generations. And Chapter 16, by Zach Solan et al., also addresses an inevitably social question, namely the existence of so many languages. A research program only concerned with how a single individual acquires her linguistic competence cannot provide an explanation for why languages differ. Postulating that the Language Acquisition Device sets parameters merely describes the problem. Finally, Chapter 17, by Andrew Smith, probes another vital area which is typically not the concern of generative linguists, though it does exercise developmental psychologists, namely the growth of conceptual structures and corresponding lexicons in response to signals from other users and experience of the context of communication. Smith shows that a certain (presumably innate) learning bias, the Principle of Mutual Exclusivity, allows the growth of successful shared communication.

Both narrow and broad research strategies have their advantages and disadvantages. Dissecting a dead animal in a sterile lab can't tell you much about its social life, and watching it roam the savannah can't tell you much about its cells. Narrow and broad approaches should complement and learn from each other. A heartening aspect of the chapters in this section is the maturity with which they do not throw out the baby with the bath water. Promising insights from generative linguistics have been kept.

Thus, at a very general level, all contributors agree that there are significant learning biases which contribute (but do not wholly account for) the shape of languages. This is a welcome change from some early adverse reactions to generative grammar, which often did not seem to realize that all learning is inevitably biased in some more or less complex way. The idea of UG is not dismissed; rather, it is situated in a wider context. As

Brighton and his co-authors put it, UG is more opaque to the facts of languages than was thought, due to the interposition of historico-cultural processes. Briscoe, too, explicitly adopts a central theoretical device of (one phase of) generative grammar, namely the idea of invariable principles and settable parameters, though he gives these a (more plausible?) statistical interpretation. And Roberts and his co-authors accept as their starting point a conception central to generative theorizing, the 'logical problem of language acquisition', in particular how a learner can withdraw from overgeneralizations in the absence of negative evidence. Note that Roberts and his colleagues accept the generative insistence that children learn from positive instances only, but their approach to the standard acquisition problem is, like Briscoe's, statistical and based on Bayesian learning.

Four of the five chapters introduced here report on specific series of computational simulations carried out by the authors, and the first chapter (Chapter 13) serves as a theoretical introduction to such studies. All these studies adopt the methodology of simulating populations of agents which communicate with other agents and learn from their behaviour, thus collectively constructing a language. Of course, all computational simulation involves massive idealization. Thus Roberts et al. in Chapter 15 make a simplification similar to Chomsky's idealization to an individual speaker–hearer when they postulate a 'population' consisting of only one individual per generation. The crucial new feature is the transmission of language over many generations, and they show a very natural mechanism by which irregular gaps in grammar can arise as a result of this process. The existence of such irregular gaps cannot be explained by a more narrow theory which focuses on the prototypically exceptionless generalizations made by language learners. If an innate generalizing device were all that shaped languages, there should be no exceptions to generalizations, no irregularities.

Only one chapter, Briscoe's, Chapter 14, discusses the co-evolutionary interaction between biogenetic evolution and the cultural transmission of language. Most generative linguists have accepted the biological basis for the human language capacity, but not got to grips with the evolutionary mechanisms which could have shaped it. It has been a commonplace that the relationship between the language faculty and genes is very complex. Certainly there is no single 'gene for grammar', or even a very small number of such genes. This has generally been taken as an indication

that it is premature to theorize about specific possible relationships between genes and language. Without committing himself to any particular propositions about chromosomes or loci, Briscoe models a range of possible relationships between genes and the language capacity. In his terminology, the relationship can be more or less 'decorrelated'. To the extent that it is impossible to allocate a small groups of genes to certain specific features of the language capacity, the relationship between genotype and phenotype (the LAD) is decorrelated. Briscoe shows that it is in principle possible to build models of such complex interaction, and more specifically comes up with the rather surprising result that parameters with default settings would be, under his assumptions, expected to increase in the course of biological evolution. Thus another question which we can give generative linguistics credit for posing, but not answering, namely 'Why are there principles and parameters?' begins to get an answer. Of course, Briscoe's modelling of the gene–language relationship is idealized and simplified, but it does establish certain basic principled results about this relationship, seen in the context of a train of simultaneous biological (via DNA) and cultural (via learning) evolution of language.

Chapter 16, by Zach Solan and colleagues, also models the shifting fortunes of languages over time, accounting for the diversity of languages through a mechanism that resembles, but crucially is not, biological evolution. As in Briscoe's work, in their model individuals who communicate better have more offspring. (This is a very common, but not unchallenged, assumption of work in this vein in general.) But Solan et al.'s model has no genes for language learning transmitted biologically. The kind of fitness that they discuss has consequences in the proportions of certain kinds of individuals in the next generation, but the characterizing features of these various kinds are not encoded in a genome. This contrasts with Briscoe's model, in which agents vary in their innate language-acquisition capacity. In Solan et al.'s model, any agent can migrate to another group in the population and have the same chances of learning its language as an agent native to that group. Thus this model, like that of Roberts et al., is a model of cultural evolution.

The final chapter, Chapter 17, by Andrew Smith, also has a simulated population of agents, but here there is no iterated learning, that is no simulation of a community's history. Smith's model is one-generational. Like the other studies, this study seriously considers factors external to

narrow language competence, in particular the structure of the external world, which can be modelled to be more or less 'clumpy'. Smith also makes the same assumption of no negative evidence (or feedback) as is made by generative theorists and by the other modellers writing here. An interesting result is that agents do not need to acquire identical competences, that is identical structures representing concepts, to be able to communicate with each other effectively. Again, considering factors external to competence, in this case the context of an utterance and the referents of signals in a simple simulated world, sheds light on a property of language previously unquestioned within the linguistic community. It has generally been taken for granted that having (near-)identical internal representations of the meanings of utterances would be a prerequisite for successful communication.

These studies are made possible by massive computing power. Modelling a single agent acquiring an internal representation on exposure to experience can be complicated enough. Multiply that by as many agents in the population as your computer can manage. Also introduce possible variation in their language-learning strategies. Further, if that is what you want to investigate, define a genome and some hypothetical more or less complex relationship between the genome and the language-learning strategy. Now define an external world, with whatever ontology you choose, and let the agents loose in this world according to some defined routines, communicating with each other perhaps entirely at random, or perhaps in ways determined by some hypothetical social structuring of the population. Now give each individual a life history, including a critical period in which it is susceptible to language learning, and a period of maturity when it behaves according to its internalized rules. Finally add population history, letting these individuals be born, live, and die, and influence each other in the process. Maybe it all sounds too much like a computer game, where the programmer gets to play God and watch the consequences. It can be done for fun, but the only point in doing it is to derive results from modelling the interaction of carefully defined parameters, and not too many at a time. If your model is too complex, you are unlikely to understand it. The computer models in this collection are thus necessarily simplified. But as the subject develops, and more practitioners become acquainted with previous studies, it becomes possible to build somewhat more complex models on simpler earlier models, without losing understanding. The models described in this collection are part of

that growing movement in the study of language. Such modelling studies do not replace, but complement, empirical research on language acquisition, language change, and the genetic basis for language.

13   Cultural selection for learnability: three principles underlying the view that language adapts to be learnable

Henry Brighton, Simon Kirby, and Kenny Smith

> If some aspects of linguistic behaviour can be predicted from more general considerations of the dynamics of communication in a community, rather than from the linguistic capacities of individual speakers, then they should be.
>
> (Ray Jackendoff)[1]

13.1 Introduction

Here is a far-reaching and vitally important question for those seeking to understand the evolution of language: given a thorough understanding of whatever cognitive processes are relevant to learning, understanding, and producing language, would such an understanding enable us to predict the universal features of language? This question is important because, if met with an affirmative answer, then an explanation for why language evolved to exhibit certain forms and not others must be understood in terms of the biological evolution of the cognitive basis for language. After all, such an account pivots on the assumption that properties of the cognitive mechanisms supporting language map *directly* onto the universal features of language that we observe. We argue against this position, and note that the relation between language universals and any cognitive basis for language is opaque. Certain hallmarks of language are adaptive in the context of

---

[1] Jackendoff (2002: 101).

cultural transmission; that is, *languages themselves* adapt to survive by adapting to be learnable. This phenomenon is termed *linguistic evolution.* Now, when seeking to understand the evolutionary prerequisites for language, this observation requires us to reconsider exactly *what* has evolved. For example, if significant aspects of linguistic structure can emerge through the evolution of languages themselves, then any theory of the biological evolution of language must seek to explain the basis, or pre-requisites, for linguistic evolution. In short, by proposing that significant linguistic evolution can occur in the absence of biological change, we are suggesting that the question of the evolutionary prerequisites for language needs to be recast to account for the opaqueness between the cognitive basis for language and the structural tendencies we observe in language.

Linguistics should explain why languages exhibit certain hallmarks and not others. In relation to this objective, the notion of cultural selection for learnability is far-reaching because, traditionally, cognitive science seeks a detached account of cognitive processes and their behaviour. The prevailing assumption is that cultural processes must be factored out as much as possible: the locus of study is the individual, with the relationship between observed input–output conditions explained by internal acts of cognition alone. Despite supporting this discussion with insights gained from several computational models, we aim to arrive at three principles that are independent of any particular model. In doing so, we attempt to frame in a wider context demonstrative results gained from computational evolutionary linguistics: the notion of selection for learnability.

First, in section 13.2, we set the scene by characterizing a principle of detachment: the position that an explanation for language universals can be gained through an exploration of the cognitive mechanisms underlying language. We discuss the motivation for deviating from this position, and sketch parallels between computational evolutionary linguistics and situated cognitive science. Next, in section 13.3, we outline some key results that support our argument. The main thrust of our argument is presented in section 13.4, where we consider three underlying principles. First, we propose an *innateness hypothesis*: to what degree are features of language explicitly coded in our biological machinery? Second, the *principle of situatedness*: how much of the characteristic structure of language can we explain without considering side-effects arising from cultural transmission? Finally, in the *function independence principle*, we make clear that our position is not rooted in any notion of language function: we seek a non-functional explanation for certain aspects of linguistic structure.

13.2 Explaining universal features of language

13.2.1 *Language universals*

Take all the world's languages and note the structural features they have in common. On the basis of these universal features of language, we can propose a *Universal Grammar*, a hypothesis circumscribing the core features of all possible human languages (Chomsky 1965). On its own, this hypothesis acts only as a description. But far from being an inert taxonomy, Universal Grammar (UG) sets the target for an explanatory theory. The kind of entities contained in UG that we will allude to consist of *absolute* and *statistical* language universals (Matthews 1997; O'Grady, Dobrovolsky, and Katamba 1997) Absolute universals are properties present in all languages. Statistical universals are properties present in a significant number of languages. Several further distinctions naturally arise when describing constraints in cross-linguistic variation, but in the interests of clarity we will restrict this discussion to one of absolute and statistical universals.

On accepting UG as a natural object, we can move beyond a descriptive theory by asking why linguistic form is subject to this set of universal properties. More precisely, we seek an explanation for how and where this restricted set of linguistic features is specified. The discussion that follows will analyse the possible routes we can take when forming such an explanation. The hunt for an explanation of universal features is traditionally mounted by arguing that Universal Grammar is an innate biological predisposition that defines the manner in which language is learned by a child. The linguistic stimulus a child faces, whatever language it is drawn from, through the process of learning, results in a knowledge of language. For example, Chomsky states that this learning process is 'better understood as the growth of cognitive structures along an internally directed course under the triggering and partially shaping effect of the environment' (Chomsky 1980: 34).

So an innate basis for language, along with the ability to learn, permits the child to arrive at a knowledge of language. Just how influential the learning process is in arriving at knowledge of language is frustratingly unclear. At one extreme, we can imagine a highly specialized 'language instinct' (Pinker 1994) where learning only 'partially shapes' the yield of the language-acquisition process: the assumption here is that linguistic evidence faced by a child underdetermines the knowledge they end up

with. At the other extreme, we can imagine a domain-general learning competence which serves language as well other cognitive tasks. Here, the suggestion is that knowledge of language can be induced from primary linguistic data with little or no language-specific constraints (Elman et al.1996).

13.2.2 *Isolating the object of study*

Recall the conundrum we are considering: how and where are universal features of language specified? The explanatory framework discussed in the previous section concerns itself with the degree to which language-specific constraints guide language acquisition, and it is assumed that these constraints determine language universals. Nowhere in this analysis is the role of the linguistic population considered: an explanation for the universal features of a population-level phenomenon—language—has been reduced to the problem of the knowledge of language acquired by individuals. In short, the traditional route to understanding linguistic universals, to a greater or lesser extent, assumes that these universals are specified innately in each human, or at least, explainable in terms of detached linguistic agents. This de-emphasis of context, culture, and history is a recurring theme in the cognitive sciences, as Howard Gardner notes:

> Though mainstream cognitive scientists do not necessarily bear any animus [...] against historical or cultural analyses, in practice they attempt to factor out these elements to the maximum extent possible (Gardner 1985: 41).

Taking this standpoint is understandable and perhaps necessary when embarking on any practical investigation into cognition. The result of this line of explanation is that we consider universal features of language to be strongly correlated with an individual's act of cognition, which is taken to be biologically determined. Now we have isolated the object of study. Understanding the innate linguistic knowledge of humans will lead us to an understanding of why language is the way it is. For the purposes of this study, let us characterize this position:

Definition 1 (Principle of detachment). A thorough explanation of the cognitive processes relevant to language, coupled with an understanding of how these processes mediate between input (primary linguistic data) and output (knowledge of language), would be

sufficient for a thorough explanation of the universal properties of language.

Now, when considering knowledge of language, the problem is to account for a device that relates input (linguistic stimulus) to output (knowledge of language). For example, Chomsky discusses a language-acquisition device (LAD) in which the output takes the form of a system of grammatical rules. He states that:

An engineer faced with the problem of designing a device for meeting the given input-output conditions would naturally conclude that the basic properties of the output are a consequence of the design of the device. Nor is there any plausible alternative to this assumption, so far as I can see (Chomsky 1967).

In other words, if we want to know how and where the universal features of language are specified, we need look no further than an individual's competence, derived from primary linguistic data via the LAD. This position, which we have termed the principle of detachment, runs right through cognitive science and amounts to a general approach to studying cognitive processes. For example, in his classic work on vision, Marr makes a convincing case for examining visual processing as a competence understood entirely by considering a series of transformations of visual stimulus (Marr 1977, 1982). We will now consider two bodies of work that suggest that the principle of detachment is questionable.

13.2.3 *Explanation via synthetic construction*

One of the aims of cognitive science, and in particular, artificial intelligence (AI), is to explain human and animal cognition by building working computational models. Those working in the field of AI often isolate a single competence, such as reasoning, planning, learning, or natural language processing. This competence is then investigated in accordance with the principle of detachment, more often than not in conjunction with a simplified model of the environment (a micro-world). These simplifying assumptions, given the difficulty of the task, are quite understandable. So the traditional approach is centred around the belief that investigating a competence with respect to a simplified micro-world will yield results that, by and large, hold true when that agent is placed in the real world. General theories that underlie intelligent action can therefore be proposed by treating the agent as a detached entity operating with

respect to an environment. Crucially, this environment is presumed to contain the intrinsic properties found in the environment that 'real' agents encounter.

This is a very broad characterization of cognitive science and AI. Nevertheless, many within cognitive science see this approach as misguided and divisive, for a number of reasons. For example, we could draw on the wealth of problems and lack of progress which traditional AI is accused of (Pfeifer and Scheier 1999: 59–78). Some within AI have drawn on this history of perceived failure to justify a new set of principles collectively termed *Embodied Cognitive Science* (Pfeifer and Scheier 1999), and occasionally *New AI* (Brooks 1999). Many of these principles can be traced back to Hubert Dreyfus's critique of AI, twenty years earlier (Dreyfus 1972). The stance put forward by advocates of embodied cognitive science is important: its proponents refine Dreyfus' stance, build on it, and crucially cite examples of successful engineering projects. This recasting of the problem proposes, among others, *situatedness* as a theoretical maxim (Clancey 1997). Taking the principle of situatedness to its extreme, the exact nature of the environment is to be taken as primary and theoretically significant. For example, the environment may be partly constructed by the participation of other agents (Bullock and Todd 1999). In other words, certain aspects of cognition can only be fully explained when viewed in the context of participation (Winograd and Flores 1986; Brooks 1999). It is important to note that this 'new orientation' is seen by many as opposing mainstream AI, or at least the branches of AI that claim to explain cognition.

Advocates of embodied cognitive science tell us that any explanation for a cognitive capacity must be tightly coupled with a precise understanding of the interaction between environment and cognitive agent. What impact does this discussion have on our questions about language universals? First, it provides a source of insights into investigating cognition through building computational models: a theory faces a different set of constraints when implemented as a computational model. Second, this discussion should lead us to consider that an analysis of cognitive processes without assuming the principle of detachment can be fruitful. In the context of language and communication, the work of Luc Steels is an example of this approach. Steels investigates the construction of perceptual distinctions and signal lexicons in visually grounded communicating robots (Steels 1997, 1998b). In this work, signals and the meanings

associated with these signals emerge as a result of self-organization. This phenomenon can only be understood with respect to an environment constructed by the participation of others.

13.2.4 *The evolutionary explanation*

Only humans have language. The communication systems used by animals do not even approach the sophistication of human language, so the evolution of language must concern the evolution of humans over the past five million years, since our last common ancestor with a non-linguistic species (Jones, Martin, and Pilbeam 1992). Consequently, examining fossil evidence offers a source of insights into the evolution of language in humans. For example, we can analyse the evolution of the vocal tract, or examine skulls and trace a path through the skeletal evolution of hominids, but the kind of conclusions we can draw from such evidence can only go so far (Lieberman 1984; Wilkins and Wakefield 1995).

One route to explaining the evolution of language in humans, which we can dub *functional nativism*, turns on the idea that language evolved in humans due to the functional advantages gained by linguistically competent humans. Language, therefore, was a trait selected for by biological evolution (Pinker and Bloom 1990; Nowak and Komarova 2001). Here, we can imagine an evolutionary trajectory starting from some biological predisposition present in protohumans for using some set of communication systems C_{proto}. From this starting point, biological evolution led to the occurrence of the set of communication systems C_{UG}, which includes all human languages. The story of language evolution can then unfold by claiming that the biological machinery supporting C_{proto} evolved to support C_{UG} due to functional pressures (see Figure 13.1). Implicit in this account is the principle of detachment. The biological evolution of cognitive capacities supporting language are equated with the evolution of languages themselves.

Over the past fifteen years, computational evolutionary linguistics has emerged as a source for testing such hypotheses. This approach employs computational models to try and shed light on the problem of the evolution of language in humans (Hurford 1989; Briscoe 2000a; Kirby 2002b). One source of complexity in understanding the evolution of language is the interaction between three complex adaptive systems, each one operating on a different timescale. More precisely, linguistic
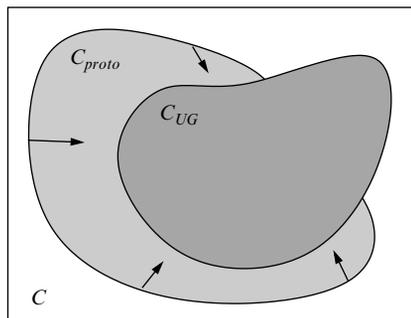
FIG. 13.1 Functional nativism. From the set of all communication systems C, the communication systems of protohumans, C_{proto}, evolved under some functional pressure towards C_{UG}.

information is transmitted on two evolutionary substrates: the biological and the cultural. For example, you are born with some innate predisposition for language which evolved over millions of years. The linguistic forms you inherit from your culture have evolved over hundreds of years. In addition to these evolutionary systems, your linguistic competence emerges over tens of years. Much of the work in computational modelling has analysed this interaction. By modelling linguistic agents as learners and producers of language, and then investigating how communication systems evolve in the presence of both biological and cultural transmission, computational evolutionary linguistics attempts to shed light on how language can evolve in initially non-linguistic communities. This approach draws on disciplines such as cognitive science, artificial life, complexity, and theoretical biology.

Recent work in this field has focused on developing models in which certain hallmarks of human language can emerge in populations of biologically identical linguistic agents. That is, characteristics of language emerge in the absence of biological change, where language is transmitted from generation to generation entirely through the cultural substrate. We detail this work in section 13.3, but mention it here as it impacts on the current discussion. In explaining how and why language has its characteristic structure, the evolutionary approach, by investigating the interaction between biological and cultural substrates, is in line with the claims made by proponents of embodied cognitive science. Because languages themselves can adapt, independent of the biological substrate, certain

features of language cannot be explained in terms of detached cognitive mechanisms alone.

13.2.5 *Summary: should we breach the principle of detachment?*

This discussion has outlined the basis for asking three questions. First, what kind of explanatory framework should be invoked when explaining universal features of language? Secondly, are any of the principles underlying situated cognitive science relevant to understanding the characteristic structure of language?[1] Thirdly, what kind of explanatory leverage can be gained by breaching the principle of detachment, and exploring issues of language evolution via computational modelling and simulation? On the validity of artificial intelligence, Chomsky notes 'in principle simulation certainly can provide much insight' (Chomsky 1993b: 30). Perhaps more relevant is the quotation located at the beginning of this chapter from another prominent linguist, Ray Jackendoff. Taking these two observations together, we should at least consider the role of the cultural transmission of language in explaining the universal features of language. The next section outlines recent work on exploring precisely this question.

13.3 Modelling cultural transmission

13.3.1 *Iterated learning models*

An iterated learning model (ILM) is a framework for testing theories of linguistic transmission. Within an ILM, agents act as a conduit for an evolving language—language itself changes or evolves rather than the agents themselves. An ILM is a generational model: after members of one generation learn a language, their production becomes the input to learning in the next generation. This model of linguistic transmission, providing that the transfer of knowledge of language from one generation to the next is not entirely accurate or reliable, will result in diachronic change. Importantly, certain linguistic structures will survive transmission, while other forms may disappear.

---

[1] We should make clear that when we refer to situatedness, we mean nothing more than a full consideration of the environmental context of cognition.

The range of linguistic structures that can be explained using an iterated learning model can vary substantially. There are two broad categories of model: those of *language change* and *language evolution.* An investigation into language change will often assume a model in which hallmarks of language are already present. Models of language change cannot, therefore, impact on our discussion of the principle of detachment. If the hallmarks of language are pre-programmed into the model, then the model cannot inform us how these hallmarks came about. Instead, models of language change can explain aspects of change within fully developed languages, and therefore aim to shed light on issues such as, for example, statistical universals—those properties present in many but not all languages. In contrast, iterated learning models of language evolution model the transition to language from non-linguistic communication systems, and can therefore shed light on how hallmarks of language came to be. For this reason, iterated learning models of language evolution are particularly relevant to our discussion of the principle of detachment.

13.3.2 *Language change*

In studying language change we often consider the trajectory of language through possible grammars. Any resulting explanation is therefore orientated neutrally with respect to explaining absolute linguistic universals. From one grammar to the next, we presume hallmarks of language are ever present (see Figure 13.2). Models of language change must invoke a situated component. A model must tackle the problem of language acquisition: a learner will deviate from the grammar of its teachers when the primary linguistic data fails to unambiguously represent the grammar from which it is derived. Knowledge of language is therefore not transmitted directly from mind to mind, but instead some external correlate—linguistic performance—must stand proxy for knowledge of language. Modelling language change must therefore consider some environment allowing the transmission of language competence via language performance. This environment, importantly, is constructed by other individuals in the culture.

Using iterated learning, we can construct computational models of language change. These studies are motivated by the observation that language change is driven by considerations arising from language acquisition (Clark and Roberts 1993; Niyogi and Berwick 1997; Briscoe 2002). For example, using a principles and parameters approach to language
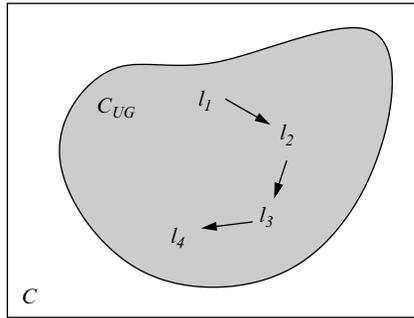
FIG. 13.2 Language change. An example trajectory of language change through languages l_1, l_2, l_3, and l_4.

specification, Niyogi and Berwick (1997) develop a population model with which they investigate the dynamics of language change. In particular, they use a probabilistic model of grammar induction to focus on the loss of verb second position in the transition from Old French to Modern French, which results directly, they claim, from misconvergences arising during language acquisition. In contrast, Hare and Elman (1995) address the problem of morphological change by examining connectionist simulations of language learning, which, when placed in the context of iterated learning, can be used to explain morphological changes, such as verb inflection in Modern English arising from the past-tense system of Old English. Importantly, the linguistic phenomenon these models attempt to explain is relatively well documented: the historical accuracy of models of language change can be tested.

13.3.3 *Language evolution*

These studies of language change tell us that the learnability of languages, over the course of cultural transmission, has a bearing on the distribution of languages we observe. Now we will discuss extending the range of explanation offered by models of iterated learning to include the possibility of explaining hallmarks of language. The dynamics of iterated learning can make certain properties of communication systems ubiquitous. This must lead us to consider the fact that, just as the dimensions of variation can be explored via iterated learning, the undeviating features of language may also depend on issues of learnability.

Importantly, the possibility that iterated learning models can shed light on an explanation of these properties will make a convincing case for questioning the principle of detachment. If the unvarying features of language can be explained in the same way as those that vary, then issues of innateness become problematic and less clear cut. For example, Christiansen, Deacon, and Kirby have each claimed previously that universals should, at least in part, be seen as arising from repeated transmission through learning:

In short, my view amounts to the claim that most—if not all—linguistic universals will turn out to be terminological artifacts referring to mere side-effects of the processing and learning of language in humans (Christiansen 1994: 127).

Grammatical universals exist, but I want to suggest that their existence does not imply that they are prefigured in the brain like frozen evolutionary accidents. In fact, I suspect that universal rules or implicit axioms of grammar aren't really stored or located anywhere, and in an important sense, they are not *determined* at all. Instead, I want to suggest the radical possibility that they have emerged spontaneously and independently in each evolving language, in response to universal biases in the selection processes affecting language transmission (Deacon 1997: 115–16).

The problem is that there are now two candidate explanations for the same observed fit between universals and processing—a glossogenetic one in which languages themselves adapt to the pressures of transmission through the arena of use, and a phylogenetic one in which the LAD adapts to the pressures of survival in an environment where successful communication is advantageous. (Kirby 1999: 132).

These arguments place an explanation for the universal features of language well and truly outside the vocabulary of explanation suggested by the principle of detachment. In the context of cultural transmission, we term the process by which certain linguistic forms are adaptive and therefore evolve and persist *cultural selection for learnability.* More precisely:

> **Definition 2 (Cultural adaptation).** By cultural adaptation, we mean the occurrence of changes in the language due to the effects of cultural transmission.

We should contrast the notion of cultural adaptation to that of genetic adaptation, where genetic changes occur as a result of natural selection. Importantly, our notion of cultural adaptation refers to the *language*

adapting, rather than the users of language. Next, we define cultural selection for learnability:

> **Definition 3** (**Cultural selection for learnability**). In order for linguistic forms to persist from one generation to the next, they must repeatedly survive the processes of expression and induction. That is, the output of one generation must be successfully learned by the next if these linguistic forms are to survive. We say that those forms that repeatedly survive cultural transmission are adaptive in the context of cultural transmission: they will be selected for due to the combined pressures of cultural transmission and learning.

In this context, the terms *adaptive* and *selection* only loosely relate to the equivalent terms used in the theory of biological evolution. Importantly, the idea that languages themselves adapt to be learnable, and in doing so organize themselves subject to a set of recurring structural properties, has been the subject of computational models that make explicit these assumptions. In particular, the experiments of Kirby (2002a) and Batali (2002) demonstrate that a collection of learners with the ability to perform grammar induction will, from an initially holistic communication system, spontaneously arrive at compositional and recursive communication systems. Because language is ostensibly infinite, and cultural transmission can only result in the production of a finite series of utterances, only generalizable forms will survive. These experiments suggest that certain hallmarks of language are culturally adaptive: pressures arising from transmission from one agent to another cause these hallmarks to emerge and persist. For example, adaptive properties such as compositionality and recursion, which we can consider absolute language universals, are defining characteristics of stable systems.

Breaching the principle of detachment requires us to adopt a conceptual framework in which the details of the environment of adaptation become crucial. The details form part of the focus of further work in this area. After all, if the precise nature of the environment of adaptation is to play a pivotal role, as suggested by situated theories of cognition, then the hope is that a wider range of linguistic forms can be explained within the iterated learning framework. For example, Kirby (2001) demonstrates that by elaborating the environment by imposing a non-uniform probability distribution over the set of communicatively relevant situations, regular/irregular forms emerge. Why is this? By skewing the relative frequency of

utterances, irregular forms can exist by virtue of the fact they are frequently used, and therefore are subject to a reduced pressure to be structured; see Chapter 15, Roberts, Onnis, and Chater. Similarly, Smith, Brighton, and Kirby (2003) show how clustering effects in the space of communicatively relevant situations leads to a stronger pressure for compositionality. These studies demonstrate that the precise nature of the environment of adaptation impacts on the resulting language structure. By understanding the impact of environmental considerations on the evolved languages, in tandem with an investigation into plausible models of language acquisition, we hope to shed further light on the relationship between cultural selection and the structure of evolved languages.

In this section we have discussed how models of language evolution and change based on a cultural, situated model of linguistic transmission can shed light on the occurrence of hallmarks of language. For a more thorough discussion and the modelling details we refer the reader to material cited, as well as a recent overview article (Kirby 2002b).

13.4 Underlying principles

We began this discussion by considering the manner in which language universals should be explained. We now present three principles that underlie the view that language universals are, at least in part, the result of cultural selection for learnability. We start by noting that any conclusions we draw will be contingent on an innateness hypothesis:

Principle 1 (Innateness hypothesis). Humans must have a biologically determined set of predispositions that impact on our ability to learn and produce language. The degree to which these capacities are specific to language is not known.

Here we are stating the obvious: the ability to process language must have a biological basis. However, the degree to which this basis is specific to language is unclear. Linguistics lacks a solid theory, based on empirical findings, that identifies those aspects of language that can be learned, and those which must be innate (Pullum and Scholz 2002). Next, we must consider the innateness hypothesis with respect to two positions. First, assuming the principle of detachment, the innateness hypothesis must lead us to believe that there is a clear relation between patterns we observe

in language and some biological correlate. If we extend the vocabulary of explanation by rejecting the principle of detachment, then the question of innateness is less clear cut. We can now talk of a biological basis for a feature of language, but with respect to a cultural dynamic. Here, a cultural process will mediate between a biological basis and the occurrence of that feature in language. This discussion therefore centres around recasting the question of innateness. Furthermore, this observation, because it relates to a cultural dynamic, leads us to accepting that situatedness plays a role:

> **Principle 2 (Situatedness).** A thorough understanding of the cognitive basis for language would not amount to a total explanation of language structure. However, a thorough understanding of the cognitive basis for language in conjunction with an understanding of the trajectory of language adaptation through cultural transmission would amount to a total explanation of language structure.

Of course, the degree of correlation between a piece of biological machinery supporting some aspect of language and the resulting language universal is hard to quantify. But in general, given some biological basis for processing language, some set of communication systems $C_{possible}$ will be possible. A detached understanding of language can tell us little about which members of $C_{possible}$ will be culturally adaptive and therefore observed. The principle of situatedness changes the state of play by considering those communication systems that are adaptive, $C_{adaptive}$, on a cultural substrate, and therefore observed. In short, cultural selection for learnability occurs with respect to constraints on cultural transmission. These constraints determine which members of $C_{possible}$ are culturally adaptive, observed, and therefore become members of the set $C_{adaptive}$.

By conjecturing an opaque relationship between some biological basis for language and some observed language universal, the notion of UG becomes problematic. Universal Grammar is often taken to mean one of two things. First, the term UG is sometimes used to refer to the set of features that all languages have in common (Chomsky 1965). Secondly, and perhaps more frequently, UG has been defined as the initial state of the language-learning child (Chomsky 1975). Figure 13.3 depicts how these two definitions relate to our discussion of the biological basis for language, the set of possible communication systems, and the set of observed communications systems.
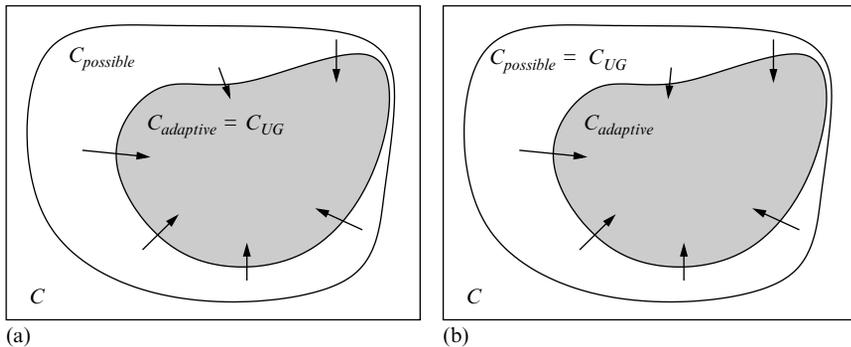
(a)   (b)

Fɪɢ. 13.3 Of the set of possible human communication systems $C_{possible}$, some set $C_{adaptive}$ are adaptive in the context of cultural transmission, and therefore observed. Depending on how we define UG, the set of communication systems characterized by *UG*, C_{UG}, are either precisely those we observe ($C_{adaptive}$), or those that are possible, but not necessarily observed ($C_{possible}$).

The set of communication systems that conform to the definition of UG are denoted as C_{UG}. Depending on which definition of UG we adopt, this set will be equivalent to either $C_{possible}$ or $C_{adaptive}$. These two alternatives are now explored:

1. *UG as the set of features common to all languages.* If we take UG as the set of features common to all observed languages, then C_{UG}, the set of communication systems conforming to UG, is identical to our set of culturally adaptive communication systems, $C_{adaptive}$. This must be the case, as only members of $C_{adaptive}$ are observed and can therefore contribute to a theory of UG under this reading. This position is represented in Figure 13.3(a).

2. *UG as the initial state of the language-learning child.* The alternative definition of UG, where UG defines the initial state of the learner, must encompass those communication systems which are possible, but not necessarily adaptive: $C_{possible}$. Because humans are equipped with the biological basis for using members of $C_{possible}$, their initial state must account for them. Hence, under this second reading of UG, $C_{possible} = C_{UG}$. As before, only some members of $C_{possible}$ will be culturally adaptive and therefore observed. Figure 13.3(b) reflects this relationship.

Irrespective of our definition of UG, an acceptance of the principle of situatedness allows us to explain a feature of language in terms of a biological trait realized as a bias which, in combination with the adaptive properties of this bias over repeated cultural transmission, leads to that feature being observed. However, if one accepts that cultural transmission plays a pivotal role in determining language structure, then one must also consider the impact of other factors which might result in the emergence of adaptive properties—for example, issues relating to communication and effective signalling. But as a first cut, we need to understand how much can be explained without appealing to any functional properties of language:

> **Principle 3 (Function independence).** Some aspects of language structure can be explained independently of language function.

A defence of this principle is less clear cut. Without doubt language is used for communication, but whether issues of communication determine all forms of language structure is by no means clear. The picture we are developing here suggests that constraints on learning and repeated cultural transmission play an important part in determining linguistic structure: the models we have discussed make no claims about, nor explicitly account for, any notion of language function. In short, the fact that, for example, compositional structure results without any model of language function suggests that this is a fruitful line of enquiry to pursue.

13.5 Conclusions

Universal features of languages, by definition, are adhered to by every user of language. We might then take the individual as the locus of study when seeking an explanation for why language universals take the form that they do. In line with this intuition, practitioners of cognitive science will often make the simplifying assumption that the behaviour of individuals can be understood by examining internal cognitive processes of detached agents. The principle of detachment characterizes this position.

In attempting to understand how and where language universals are specified, this discussion has focused on questioning the principle of detachment. We have explored two sources of ideas that suggest that an explanation of the characteristic structure of language could benefit from

breaching the principle of detachment. Firstly, advocates of situated cognitive science claim that the property of situatedness, a full understanding of the interaction between agent and environment, is theoretically significant. Secondly, recent work in the field of computational evolutionary linguistics suggests that cultural dynamics are fundamental to understanding why linguistic structure evolves and persists. We should stress here that in one respect languages are not stable because they are constantly changing. But in contrast, absolute linguistic universals are entirely stable, or at least they have been over the duration of modern linguistic inquiry.[2]

Taking these two sources as evidence, we outlined recent computational models that explore the relation between language universals and those linguistic features that are adaptive in the context of cultural transmission. On the basis of these experiments, we claim that cultural selection for learnability must form part of any explanation relating to how and where language universals are specified. We claim that, due to constraints on cultural transmission, languages adapt to reflect the biases present in language learners and producers. The relationship between these biases and the observed universal features of language is therefore opaque: a cultural dynamic mediates between the two.

Here is the message we wish to convey: selection for learnability is an important determinant of language universals, and as such should be understood independently of any particular computational model. Our aim is to outline the theoretical foundations of cultural selection for learnability. We do this by proposing three principles. First, the Innateness Hypothesis (Principle 1) states that there must be a biological basis for our language-learning abilities, but the degree to which these abilities are specific to language is unclear. The second principle, the principle of Situatedness (Principle 2), states that language universals cannot be explained through an understanding of the cognitive basis for language alone. Importantly, we claim that certain properties of language are adaptive in the context of cultural transmission. The third principle, that of Function Independence (Principle 3), makes clear that any functional properties of language are not necessarily determinants of language structure. We note that an explanation for certain universals, such as

---

[2] See Newmeyer (2002) for discussion of this and other issues that relate to uniformitarianism in linguistics.

compositional syntax, need not appeal to any notion of language function. In short, we seek a non-functional explanation for certain aspects of linguistic structure.

By questioning the principle of detachment and pursuing a line of enquiry guided by Principles 1–3 we have argued that the concept of cultural selection for learnability can provide important insights into some fundamental questions in linguistics and cognitive science. The work presented here should be seen as the first steps towards a more thorough explanation of the evolution of linguistic structure.

ACKNOWLEDGEMENTS

FURTHER READING

Christiansen and Kirby (2003) provide a selection of articles by influential thinkers covering a broad sweep of topics relevant to evolutionary linguistics. For those interested in the debate on cognitive science and embodiment, Pfeifer and Scheier (1999) constitutes an excellent source of further reading. Whereas in this chapter we have avoided appealing to language function when attempting to explain linguistic structure (in accordance with our principle of functional independence), Newmeyer (1998a) provides a thorough and thoughtful consideration of the place of such functionalist approaches in linguistics. Finally, computational models have played a key role in driving the development of the position that we present here. Briscoe (2002) provides a collection of articles describing recent research involving such models, and can be recommended to those seeking further insight into the topics, issues, and techniques of the field.