# CHAPTER 7

# Conclusions

I began this thesis by highlighting three features which, in combination, make language unique among the communication systems of the natural world — language is the only communication system which is culturally transmitted and symbolic and compositional. Theories of the evolution of language must explain the origins of these three distinctive features of language.

A recurring theme of this thesis has been to explain the latter two hallmarks of language in terms of the first — to explain the conventionalised symbolic vocabulary and compositional structure of language as features which arise during the cultural transmission of language. Pressures acting on language during cultural transmission result in the emergence of conventionalised symbolic vocabulary and compositional linguistic structure, through the adaptation of the linguistic system to the medium of cultural transmission.

A feature of this thesis has been the extensive reference to, and use of, formal models — as discussed towards the end of Chapter 1, theories of language evolution typically appeal to the interactions between several complex adaptive systems, and our intuitions about the behaviour of these kinds of systems are notoriously poor. Formal modelling techniques, such as those I have outlined here, allow us to investigate the outcomes of complex adaptive processes, and, by experimentation, to uncover underlying regularities and determinants of the behaviour of such systems. This kind of experimentation has allowed me to identify what I consider to be the two key determinants of linguistic structure — a bottleneck on cultural transmission and a one-to-one bias in language learners.

Can other features of language be explained as a consequence of pressures acting on language during cultural transmission? The comparison of design features of language and other communication systems carried out in Chapter 1 suggests that duality of patterning, whereby a small number of meaningless elements (phonemes) are combined to form a large number of meaningful elements (words), may be a fourth unique characteristic of language. Duality of patterning seems to be a prime candidate for explanation by this kind of cultural adaptation account. We can envisage a scenario under which agents produce sequences of articulatory gestures, initially without any underlying system of phonemic coding, which are reinterpreted over cultural time as sequences of gestures representing underlying phonemic representations. Formal models have demonstrated that, for vowel systems at least, this kind of cultural self-organisation can lead to the emergence of phonemic coding systems, via reinforcement learning (e.g. de Boer (2000), de Boer (2001)) or a coupling of sensory and motor systems (Oudeyer 2002).

These models of the emergence of phoneme systems need to be extended to explain the combination of phonemes to form words. However, the initial results of de Boer and Oudeyer, in conjunction with the research on symbolicism, compositionality and recursion reviewed and described in this thesis, suggest that cultural processes may offer a powerful and general explanation for most of the hallmarks of human language.

In Chapter 2 I introduced Boyd & Richerson's (1985) general model of cultural transmission, and discussed more language-specific treatments of cultural processes. There are three main issues in treating language as a culturally transmitted system. The first is to identify the units of cultural replication. It may be the case that there is no true replication in a cultural system — the cycle from E-language to I-language means that there is no direct replication of either E-language or I-language. The second problem is identifying the cultural traits of interest. Given that the formation of I-language is dependent on observation of E-language, and the production of E-language is dependent on an underlying I-language, both (or neither) external and internal language may be considered as culturally transmitted traits.

I have followed Boyd & Richerson's assumption, implicit in Boyd & Richerson (1985) and explicit in Boyd & Richerson (2000), that we can ignore such issues, at least until we have established the usefulness of cultural approaches to explaining the human behaviour of interest. As Boyd & Richerson put it,

> "[i]f the application of Darwinian thinking to understanding cultural change depended on the existence of replicators, we would be in trouble. Fortunately, culture need not be closely analogous to genes. Ideas must be gene-like to the extent that they are somehow capable of carrying the cultural information necessary to give rise to cumulative evolution of complex cultural patterns that differentiate human groups ... this can be accomplished by a

> most ungene like, replicatorless process of error-prone phenotypic imitation. All that is really required is that culture constitutes a system maintaining heritable variation" (Boyd & Richerson 2000:158).

In other words, we can ignore the precise mechanisms underlying cultural transmission, provided that we accept that cultural transmission, or cultural heritability, is possible in some sense. The Darwinian theory of evolution by natural selection was conceived, and broadly accepted, in spite of the lack of a precise understanding of the units underlying biological heritability, or the units of biological replication. Theories of cultural evolution should be allowed similar leeway, at least for the time being.

The third problem of modelling the cultural transmission of language concerns the nature of the primary linguistic data available to learners. I have assumed throughout this thesis that learners observe meaning-signal pairs produced by other members of their linguistic community. I have further assumed that the mental representations corresponding to meanings are shared within a population and, in the case of the models of structured language discussed in Chapters 5 and 6, that these meanings are structured. My position has been similar to that of Schoenemann in that I "take for granted that there are features of the real world which exist regardless of whether an organism perceives them ... [d]ifferent organisms will divide up the world differently, in accordance with their unique evolved neural systems ... [i]ncreasing semantic complexity [or meaning space structure] therefore refers to an increase in the number of divisions of reality which a particular organism is aware of" (Schoenemann 1999:318). Similarly, I take it that perception of structure in the environment is a consequence of focusing on regularities in the real world which exist regardless of whether organisms perceive them. This capacity for structured semantic representations may form a preadaptation for language, or as discussed below, may co-evolve with linguistic forms.

Under this set of assumptions, I have shown that cultural evolution can deliver up some of the characteristic structure of language. But how safe are these assumptions?

As discussed in Chapter 2, there is some evidence that human infants can deduce the meaning which should be associated with an utterance, through strategies such as intentional inference (Baldwin 1991; Baldwin 1993a; Baldwin 1993b; Tomasello & Barton 1994; Tomasello *et al.* 1996), and biases to attend to, for example, whole objects (Gentner 1982; Macnamara 1982). However, I am not suggesting that this process is error free. It has been argued that if this process *were* error free, and humans were effectively telepathic, then there would be no need to produce utterances at all. I think this is perhaps over-stating the case. We might accept that the process of identifying the referents of nouns like "book" or verbs like "jump" were error-free, while simultaneously accepting that the process of deducing the meaning associated with utterances such as "Last Tuesday, my brother told me he was in love with a widow" would be error-prone — the straightforward deduction of the meaning of simple terms might form a building block for the acquisition of some language, which then allows the acquisition of more opaque terms by syntactic context, more complex processes of deduction and so on. As a further defence, it has been demonstrated that the basic Expression/Induction framework can work even if we assume that the observation of meaning is error-prone (Hurford 1999) — if we weaken the assumption that learners receive error-free meaning-signal pairs, the whole E/I approach does not come crashing down.

Formal modelling approaches have been used to tackle the issue of language evolution in the absence of explicit meaning transfer (e.g. Steels (1997), Steels (1998), Smith (2001a) and Smith (forthcoming)). These demonstrate that stable communication systems can emerge if we do not assume that learners are presented with meaning-signal pairs. In these models, learners are instead presented with a signal and a set of objects in a simplified world, where one feature of one of the objects is the referent of the signal. These models have been broadly interpreted as showing that explicit meaning transfer is an unnecessary assumption, and therefore should not be made. I would draw the opposite conclusion: these models show that not too much rests on the assumption of explicit meaning transfer, therefore we should make this assumption in the interests of simplicity. However, these models are still extremely valuable in that they shed light on a second issue related to the observation of meaning: how do members of a population converge on a shared set of meanings, and how are these meanings structured?

Andrew Smith (Smith forthcoming) demonstrates that agents will only arrive at similar semantic representations through self-organisation if they 1) construct meanings in an intelligent way, so that they discriminate objects in the world from other objects in the world and 2) inhabit a world which is "clumpy", such that objects cluster together in object space, rather than being randomly distributed throughout the space of possible objects. I have assumed that all members of a population have access to a set of shared semantic representations. Smith shows that such shared semantic representations are essential for successful communication within a population, if we assume an observational, rather than reinforcement, learning paradigm. It is intriguing to note that Smith finds the emergence of such shared semantic representations only where the world is clumpy, and I find that compositional languages emerge most frequently when the shared semantic representations are "clumpy" (i.e. occupying a structured subspace of the space of possible meanings). This suggests a scenario where a world exhibiting a certain limited and regular degree of variation in the objects it permits leads to the formation of shared semantic representations in communicative agents, which in turn leads to the emergence of a shared, compositional language.

The second half of Chapter 2 was spent outlining Boyd & Richerson's taxonomy of pressures operating on cultural transmission — natural selection of cultural variants, guided variation, and biased transmission (which can be further subdivided into directly-biased transmission, indirectly-biased transmission and frequency-dependent bias). As discussed in that Chapter, all of these processes have been implicated in the cultural evolution of language. In this thesis I have focussed almost exclusively on directly-biased transmission, specifically cultural evolution resulting from learner biases with respect to the one-to-one nature of meaning-signal mappings, although in Chapter 3 I briefly consider the possibility of the natural selection of cultural variants. This force turns out to be weak in comparison to the pressure arising from learner biases — learner biases tend to eliminate the cultural variation which natural selection feeds upon.

Chapters 3–6 constitute the bulk of the thesis, and outline original research into the cultural and biological evolution of language and language-learning biases. In Chapter 3 I show that a bias in favour of one-to-one mappings is required for the cultural evolution of a symbolic, communicatively functional vocabulary. I further argue that this bias is present in humans. In Chapter 5, I demonstrate that a one-to-one bias is necessary for the cultural evolution of compositionally-structured language, although such languages only emerge frequently when there is a bottleneck on cultural transmission. Once again, I show that children appear to bring precisely this kind of bias to the language acquisition task.

The theme of Chapters 3 and 5 is that one-to-one biases play a crucial role in the evolution of language, and that one-to-one biases apply during language acquisition by human infants. The clear implication of these Chapters is therefore that linguistic structure may be a reflex of human learning biases applied to the cultural transmission of language, rather than a consequence of an innate specification of linguistic structure. A strong nativist might argue that linguistic structure is prespecified, and in order to understand language we need only to understand this innate prespecification. My position is that language is only prespecified in the sense that language learners come to the language learning task with an innate learning bias. Understanding this learning bias is important, but not the whole story — we also have to understand how the learning bias interacts with the medium of cultural transmission. It is only through this interaction that linguistic structure emerges. For example, if there is no bottleneck on cultural transmission then the one-to-one learning bias does not lead to the reliable emergence of compositional structure. However, a bottleneck on transmission radically changes this picture — the bottleneck forces language to be generalisable, and this pressure, in combination with the biases of learners, leads to the emergence of compositionality. Compositionality is not specified entirely in the learner, but emerges through the interactions between learner biases and transmission of linguistic structure on a cultural substrate (Brighton *et al.* forthcoming).

There are several important implications, and potential elaborations, of the research described in Chapters 3 and 5. Firstly, it highlights the importance of one-to-one biases in explaining the cultural evolution of linguistic structure. Learning bias has typically taken a back seat in explanations of linguistic evolution, with the role of the transmission bottleneck and the resultant pressure to generalise receiving more attention. However, Chapter 5 demonstrates (both through my original modelling and a review of the learning biases used in Kirby (2002), Batali (2002) and others) that one-to-one biases are crucial for the evolution of compositionality. I suspect that we might find them to be important on almost every level of linguistic structure.

Generalisation alone leads to the loss of linguistic structure — the generalisation that any meaning can be expressed by any linguistic form covers all possible combinations of data, and subsumes all other generalisations. We therefore might expect, given a pressure to generalise, that the most stable (or only stable) system would be one which maps any meaning to any form. This is clearly not what we see in language — there must be something counteracting the pressure to generalise in an unconstrained way. When considering this possibility, Hurford concludes that natural selection (of cultural variants or of genetic variants) must be responsible for preventing overgeneralisation:

> "in hearing a particular syllable used to express a particular atomic meaning, an acquirer might in theory make the absurd overgeneralization than *any* syllable can be used to express *any* meaning . . . [a]ny tendency to make overgeneralizations of such an absurd kind would presumably be eliminated by natural selection based on success in correctly divining a speaker's meaning and/or successfully signalling ones own meaning. Any mutant [cultural or biological] displaying any tendency to generalize from the primary linguistic data in ways which will lead to her being misunderstood, as she would be if she used *any* form to convey *any meaning*, will be at a disadvantage." (Hurford 2000:348).

We can probably rule out natural selection acting on overgeneral cultural variants — as discussed above, such pressures are fairly weak in comparison to those arising from learner biases, and are also likely to be weak in comparison to the pressure to generalise arising from a bottleneck on transmission. This leaves us with Hurford's (intended) conclusion, that natural selection must weed out individuals who are capable of learning in an overgeneral way. This of course implies that the learning capacity with respect to generalisation is genetically encoded. I would further argue that this genetically encoded constraint on generalisation comes in the form of a bias in favour of one-to-one mappings between meanings and signals — such biases constrain generalisations in *precisely* the right way. They allow the bottleneck on transmission to determine the degree of generalizability of the linguistic system, while at the same time constraining the system so that communicative function is maximised.

Let me take a more concrete example. In the G&B framework of syntactic analysis, reference is commonly made to a general movement operator, move-$\alpha$. Move-$\alpha$ states that any element in a syntactic structure can be moved to any other position in that syntactic structure. Of course, this generates syntactic structures which no speaker of a language would accept as grammatical. Move-$\alpha$ is therefore constrained in certain ways — for example, in terms of the possible landing sites of the moved element or in terms of "islands" from which elements may not move.

Move-$\alpha$ is clearly an overgeneralisation. This generalisation would have a high yield in cultural terms, and we might therefore expect it to emerge over cultural time, given a bottleneck on transmission. The fact that unconstrained movement does not occur suggests that there is some counteracting force at play. This could be a consequence of an arbitrary bias in learners, resulting from the structure of internal representations or from the wiring of the language centers of the brain. However, it could also be a consequence of a preference for one-to-one mappings between underlying semantic representations and surface forms. Unconstrained application of move-$\alpha$ would scramble such relationships, potentially beyond repair. A preference for one-to-one biases would constrain movement so that, by and large, the correspondence between underlying semantic forms and surface forms is preserved.

A second implication, if we accept that one-to-one biases have an important role to play in shaping linguistic evolution, is that more research should focus on identify these biases (or their absence) in the process of language acquisition. The Contrast/Mutual Exclusivity bias has been fairly well established, and I have suggested that similar experiments could be used to identify a bias against homonymy in lexical acquisition. One-to-one biases in the acquisition of structured linguistic form might be more difficult to isolate experimentally. However, language acquisition in naturalistic circumstances provides fairly good evidence on this, and the reanalysis of existing data in terms of one-to-one biases might shed further light on this issue.

Hauser *et al.* (2002) emphasise the importance of cross-species comparison in informing theories of language evolution. I would suggest that one-to-one biases provide an ideal candidate for this type of comparative approach. The tests for Contrast/Mutual Exclusivity, and the proposed test for a bias against homonymy, are fairly simple experiments which should be relatively straightforward to apply to non-human species, or at least the higher primates. However, to date only a single such experiment has been carried out (Lyn & Savage-Rumbaugh 2000), using two subjects, and the results are somewhat equivocal.

In the concluding sections of Chapters 3 and 5 I make the point that one-to-one biases in the learner are not the only possible biases at play in the cultural transmission of language — in particular, there are two further pressures, summarised by Langacker (1977) as pressures for signal simplicity (a least-effort preference on the part of individuals producing utterances) and code simplicity (a bias, in learners, against having to memorise large numbers of fixed expressions, such as words). I pointed out that these two pressures work *with* any one-to-one bias in learners in eliminating synonymy, but work *against* the one-to-one bias by favouring homonyms. This was advanced as one possible explanation for the fact that homonyms are fairly common in language, whereas synonyms are not. A worthwhile extension to the models in Chapter 3 and (particularly) Chapter 5 would be to introduce an explicit treatment of such pressures. This would show whether stability is a possibility given a tension between several learner/speaker biases and, if so, what structure the linguistic systems exhibits at stable states.

A second possible line of extension is to consider a fuller range of pressures operating at the level of the Arena of Use, the interface between the grammatical competence of one individual and the primary linguistic data of another. The transmission bottleneck is one aspect of the Arena of Use (an infinitely expressive competence is represented only by a finite number of utterances), as is the proposed bias in favour of signal simplicity. However, there are other possible pressures which could act at this level. Firstly, there are aspects of the poverty of the stimulus other than the transmission bottleneck. In Chapter 1 I summarised Pullum & Scholz's (2002) analysis of aspects of the poverty of the stimulus problem, which I repeat here.

1. Children are not specifically or directly rewarded for their advances in language learning.

2. Children's data-exposure histories are finite, but they acquire an ability to produce or understand an infinite number of sentences.
3. Children's data-exposure histories are highly diverse, yet language acquisition is universal.
4. Children's data-exposure histories are incomplete in that there are many sentences they never hear, yet can produce and understand.
5. Children's data-exposure histories are solely positive — they are never given details of what is ungrammatical.
6. Children's data exposure histories include numerous errors, such as slips of the tongue and false starts.

The transmission bottleneck corresponds to points 2 and 4 in this list, and possibly point 3 — diversity of input might be a consequence of taking a small sample from an infinite set. Diversity of input might also add a further pressure for generalisation. Consider the case where children receive learning input both from adults, and from other children. If each child receives different input from adults, this will tend to lead different children to converge on different grammars. However, if children also learn from one another, there will be a pressure for the children to converge on a grammar which is more general than that suggested by each individual's input from adults — learner-learner contact might force children's grammars to accept the union of the sets of sentences produced by the (adult) cultural parents of each child. Similarly, point 5 in the list above might introduce a bias towards generality — the lack of explicit negative evidence will tend to allow children to misconverge on superset grammars. The possible consequences of the first and final points in this list of aspects of the poverty of the stimulus are less obvious.

Finally, pressures arising from the passage of language through the Arena of Use might offer one possible explanation for *iconicity* in syntax. Iconicity has been used as a fairly general term which subsumes what I have called isomorphism, the one-to-one relationship between semantic representations and surface structures. Iconicity has also been invoked to cover cases where linguistic form appears to mirror non-linguistic reality. McMahon (1994) gives the scarcity of languages where objects precede subjects as a possible example of iconicity — this phenomenon "might be ascribed to the greater relevance or perceptual salience of the Subject in real-world situations; in linguistic representations of those situations, the Subject therefore comes first" (McMahon 1994:86). If this type of iconicity is to arise anywhere, it must be in the Arena of Use.

In Chapters 4 and 6 I describe models designed to test whether natural selection acting on genetic transmission can identify learning biases which lead, through cultural processes, to optimal (and compositional, in Chapter 6) communication. The main result of these Chapters is to show that this was not the case, or at least not reliably so. The cultural evolution of a functional communication system takes time, and consequently there is no immediate fitness payoff for individuals who are appropriately biased. This leads to random genetic transmission in the early stages of the construction process, which can stop cultural evolution before it gets started. These experiments support the general logical point that there is no advantage in being able to acquire an optimal communication system if there is no meaningful communication system to acquire.

These experiments indicate to me that the default assumption about the evolution of the human language acquisition biases should be non-adaptationist, rather than adaptationist. Even in the simple models in Chapters 4 and 6, with a fairly simple genetic search space and strong selection pressure acting in favour of successful communication, such biases are unlikely to evolve. It therefore seems safer to conclude either that the biases applied to the acquisition of language either a) originally evolved under selection pressure for some other function or b) did not evolve for any specific function, but are a spandrel, a coincident feature of some learning apparatus which was selected for. On the face of it, this is a rather frustrating conclusion — while I give a strong argument on the learning biases required for language, my account of the evolution of such biases appeals to extra-linguistic factors.

However, things are not as bleak as all that. Firstly, I am not ruling out the possibility of evolutionary reappropriation followed by adaptation specifically for language. For example, the human language learning biases might have evolved for some other purpose, then been pressed into service for language and subsequently specialised solely for language. I would also allow the possibility that the appropriate learning biases *arose* for some non-linguistic or general purpose, but were almost immediately identified as being good for communication and *spread* widely as a direct result of their utility for communication.

In part because of the rather unsatisfactory conclusions forced upon us by the results of Chapters 4 and 6, this area is ripe for further exploration. I have focussed on the assumption that learning biases are genetically encoded. An intriguing alternative possibility is that they are learned. Quinn (2001) shows that the capacity to communicate can evolve in populations which originally do not communicate, and are not pre-configured for communication. The cultural parallel to this would be to show that the capacity to learn from others was itself a learned capacity. More subtly, individuals might learn which type of learning is best for acquiring a communication system, in which case I anticipate they would learn to apply a one-to-one learning bias. Of course, the evolution of the meta-learning process, whereby individuals learn to learn from others, then has to be explained.

One possibility is that this meta-learning capacity is rooted in our understanding of others as intentional agents.

A second implication of these results relates to the role of the Baldwin effect in explaining language evolution. The Baldwin effect has received a great deal of attention of late (witness Jackendoff's comment that he "agree[s] with practically everyone that the 'Baldwin effect' had something to do with it [language evolution]" (Jackendoff 2002:237)), but it is not clear how the Baldwin effect relates to accounts of biological evolution of a learning bias which effects cultural evolution. The Baldwin effect predicts that behaviours which are initially learned tend to become innate — in other words, genes adapt to fit cultures. However, in my models the opposite is true — cultures adapt to fit genes, because genes encode learning biases which guide cultural evolution. What would Baldwinian evolution look like in this framework? We could envisage a scenario where the genes encode a range of strengths of bias. The emergence of a weak bias, say in favour of one-to-one mappings, results in the emergence of cultural traits conforming to this bias. Via the Baldwin effect, stronger and stronger forms of the one-to-one bias then evolve biologically, to allow individuals to reliably acquire the dominant cultural trait. Learned traits do not strictly speaking become innate, but learning becomes more restrictive and more biased. This seems to me to be a form of the Baldwin effect which is very appealing for explaining language evolution — we do not necessarily want to say that there is a tendency for language to become innate, but we might want to say that there has been evolution to more and more constraining forms of learning.

Finally, Chapters 4 and 6 indirectly highlight the importance of population dynamics in understanding the cultural evolution of language. In Chapter 4, the spatial organisation of populations was shown to have consequences for the speed of cultural evolution in those populations, which in turn impacted on the biological evolution of learning biases. In Chapter 6, in the population Iterated Learning Model, factors such as the number of cultural parents each individual has impacts on the possibilities for cultural evolution. It is reassuring that these models are sensitive to population structure and population dynamics — as discussed in Chapter 1, population dynamics play a role in creolization events, and the population dynamics have to be in some sense 'right' before creolization can take place (Ragir 2002). However, it is not clear what the right population dynamics are, or why precisely these matter. Computational modelling could be profitably applied to the investigation of such questions.

To summarise, the central theme of this thesis has been to explain the unique features of language in terms of the cultural adaptation, by language, to two pressures:

1. a pressure to be generalisable, arising from a bottleneck on cultural transmission
2. a pressure to conform to a learner preference for one-to-one mappings between meanings and signals, and parts of meanings and parts of signals.

Linguistic structure is not specified entirely in the learner, but emerges through the interactions between learner biases and transmission of linguistic structure on a cultural substrate.