CHAPTER 2

Models of cultural transmission

In this chapter I will review formal models of cultural transmission. These models fall into two groups — models which have been developed to account for cultural transmission in general, and models which have been developed to account for the cultural transmission of language in particular. The general model of cultural transmission given here is based heavily on the work of Robert Boyd and Peter Richerson, in particular Boyd & Richerson (1985) (henceforth B&R). B&R use mathematical techniques adapted from theoretical biology. The models of the cultural transmission of language have been developed by a fairly diverse group of researchers, typically (though not exclusively) working with computational models.

In Section 2.1 I review, in broad terms, two models of cultural transmission. In Section 2.2 I review in slightly more detail the mechanisms of cultural transmission acting in these models. Finally, in Section 2.3 we will see how different pressures acting on cultural transmission can drive cultural evolution and cultural adaptation, with particular reference to the cultural evolution of language.

There are two goals for this chapter. The first is to review the range of formal models which have been used to study cultural evolution in general, and the cultural evolution of language in particular. This review covers relevant techniques which have been used to address this question, and suggests areas which are worthy of further formal modelling. Secondly, some of the fundamental results of B&R's simple models will prove useful in interpreting the results of the more complex models introduced in later chapters.

2.1 General and linguistic models

2.1.1 A general model

A general and simple model of cultural transmission must account for three processes:

1. The cultural transmission of behaviour from a mature population to an immature population. The target of learning for the immature population is the behaviour of the mature population.
2. Individual learning by the immature population. The target of learning is determined by the environment, rather than the mature population.
3. Removal (possibly in a selective fashion) of individuals from the population.

The simplest scenario proceeds as follows. There is some set of immature individuals and some set of mature individuals. The immature individuals observe and learn from the mature individuals. The newly enculturated immature individuals then interact with the environment and adjust their behaviour according to processes of individual learning. Finally, the environment takes its toll on the population, removing some individuals and sparing others to produce a new, mature population. The process then repeats with a new immature population.

Following B&R, the simplifying assumption is made that cultural transmission, individual learning and selection can be separated out into these discrete stages. In a more realistic model these processes would be continuously modifying the population.

The distribution of phenotypes in a population at time t can be given by F_t, which specifies the proportion of each phenotype in the population. How does F_t change over time?

We will assume that F_t gives the initial phenotype distribution, prior to any social transmission, individual learning or environmental impacts. F'_t gives the distribution of phenotypes in the population after cultural transmission. This depends on three factors: 1) F_t, the distribution of phenotypes in the population prior to cultural transmission; 2) F'''_{t-1}, the distribution of mature phenotypes in the previous generation participating in cultural transmission; 3) the mechanism of cultural transmission.

F'''_t gives the distribution of phenotypes in the population once the interaction of individuals with the environment have been taken into account. These interactions can be separated into two parts: individual learning, which yields a phenotype distribution F''_t, and differential retention, which yields the final phenotype distribution F'''_t.

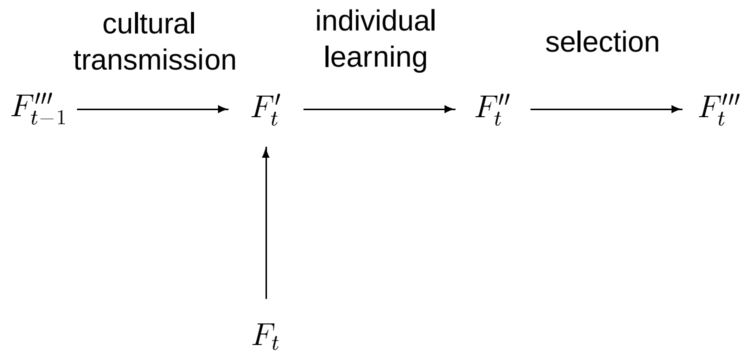

Figure 2.1: A general model of cultural transmission. F_t is the original distribution of phenotypes at time t (determined by factors other than culture). F'_t is the distribution of phenotypes after cultural transmission. F''_t is the distribution of phenotypes after individual interaction with the environment (specifically, learning). F'''_t is the distribution of phenotypes after individual interaction with the environment (specifically, death).

F''_t gives the distribution of phenotypes in the population once individual learning has been taken into account. This depends on two factors: 1) F'_t, the distribution of phenotypes prior to individual learning; 2) the process of individual learning, by which individuals change their phenotype in response to the environment.

F'''_t gives the distribution of phenotypes in the population after removal of individuals through death has been taken into account. This depends on: 1) F''_t, the distribution of phenotypes prior to death; 2) the process of differential retention, by which some individuals survive and some individuals are removed due to environmental factors.

The general model of cultural transmission is illustrated in Figure 2.1.

2.1.2 Linguistic models

As discussed in Chapter 1, in the dominant Nativist paradigm language is viewed as an aspect of individual psychology. Language acquisition is seen as the "growth of cognitive structures [linguistic competence] along an internally directed course under the triggering and partially shaping effect of the environment" (Chomsky 1980:34). The environmental triggers which guide the growth of a particular linguistic competence come from the Primary Linguistic Data (PLD), the language the child observes others using. Those working within the Nativist paradigm typically emphasises the degenerate nature of the PLD, in part to offer support for the hypothesised innate UG. However, the fact that the PLD plays some role in the formation of linguistic competence suggests that language is, to some extent, culturally transmitted — an individual's linguistic competence,

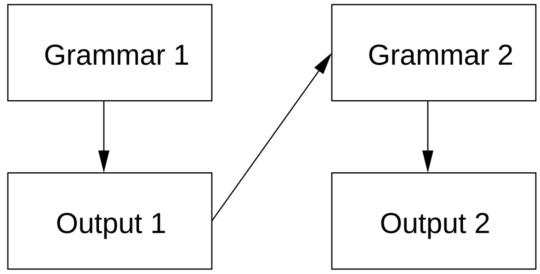

Figure 2.2: The transmission of language from generation to generation. The output of one grammar forms the PLD on which other grammars are based. Andersen (1973) points out that any attempt to make direct connections between grammar and grammar, or output and output, are spurious.

mediated by performance considerations, leads to linguistic behaviour, which forms the (degenerate) PLD for the formation of linguistic competence in other individuals.

Chomsky himself tends not to pursue this line of reasoning very far, being "concerned primarily with an ideal speaker-listener, in a completely homogeneous speech-community" (Chomsky 1965:3). However, the cultural transmission of language is of more interest to those concerned with language change, which necessarily involves a consideration of actual speaker-listeners in more or less heterogeneous speech communities.

Andersen (1973) presents an influential early account of phonological change. Andersen begins with the assertion that "[w]hat is needed is a model of phonological change which recognizes, on the one hand, that the verbal output of any speaker is determined by the grammar he has internalized, and on the other, that any speaker's internalized grammar is determined by the verbal output from which it has been inferred" (Andersen 1973:767). This scenario is sketched in Figure 2.2, adapted from Andersen's (1973) Figure 1.

Andersen applies this cultural approach to an account of phonological change in two dialects of Czech. Prior to 1300, Old Czech made a phonemic distinction between plain and palatalised ("sharped", in Andersen's terminology) dental and labial consonants. From 1300 to the end of the 1400s, most Czech dialects lost this phonemic opposition, palatalised dentals being replaced with plain dentals and palatalised labials being replaced with plain labials or plain labials plus /j/. However, in a group of dialects which Andersen terms the Teták dialects[1], palatalised labials became dentals before /i/, /e/ and /r/. The Teták dialects later lost the dental pronunciation and acquired the more

[1]This name derives from the pronunciation of the Czech word for "five" by speakers of such dialects. In more standard Czech five is /pjet/, whereas in Teták dialects it was pronounced /tet/. This difference is a consequence of the phonological change Andersen is concerned with.

standard labial pronunciation, perhaps due to the stigma associated with the Teták pronunciation — Andersen reports several standard manners for ridiculing Teták speakers, including /tiːte tiːvo ʃak je s tenou/ (meaning "Drink your beer, never mind the head"), which would be pronounced /piːte piːvo ʃak je s penou/ in non-Teták dialects. While Andersen provides an account of this later remedial change, we will focus here on his account of the first change which lead to the distinctive Teták pronunciation.

According to Andersen, the loss of a distinction between palatalised and plain labials and dentals in non-Teták dialects occurred due to errors by learners of those dialects in their analysis of the tonality of consonants. In Old Czech, the relevant consonants were either of high tonality (the dentals) or low tonality (the labials). Palatalisation further heightened the tonality of palatalised dentals or labials. This system is depicted as Stage 1 in Table 2.1a (based on Andersen's Table 2). Heightened high tonality was reinterpreted as non-heightened high tonality by learners of the non-Teták dialects, leading to the loss of the palatalised versus non-palatalised contrast for dentals (Stage 2 in Table 2.1a). Subsequently, heightened low tonality was reinterpreted as non-heightened low tonality, yielding the final non-Teták system (Stage 3), with no phonemic distinction based on palatalisation.

The first two stages of the change in Teták dialects proceeded in the same way, as illustrated in Table 2.1b (from Andersen's Table 3) — the palatalisation distinction was lost for dentals. However, the third stage of the change proceeded differently in the Teták dialects. The acoustic manifestation of heightened low tonality is ambiguous. In the non-Teták dialects this ambiguous tonality was interpreted by learners as representing underlying non-heightened low tonality. However, in the Teták case learners interpreted heightened low tonality as a realisation of underlying non-heightened high tonality. Adults produced linguistic behaviour which contain realisations of an underlying phoneme /p^{j}/. This behaviour constituted the PLD for learners. However, the acoustic ambiguity of the PLD led learners to interpret the consonant of interest as a manifestation of the underlying phoneme /t/.

Andersen gives examples of similar changes, induced by the ambiguity of the PLD, in the consonant system of early Latin and the vowel system of Old English. The account of the different paths taken by non-Teták and Teták dialects as sketched here is of course incomplete — it remains to be explained why learners of Teták dialects consistently reduced the phonemic distinction along different lines from learners of other dialects of Czech. This could be due to chance, or the phonological or phonetic properties of other parts of the Teták dialects. Andersen also proposes a system of *adaptive rules*, by which learners repair some of their misacquisition of the consonant system by realising certain instances

(a)

<table>
<tr><th></th><th>Stage 1</th><th>Stage 2</th><th>Stage 3</th><th></th></tr>
<tr><td>Heightened high tonality</td><td>$/t^j/$</td><td rowspan="2">/t/</td><td rowspan="2">/t/</td><td rowspan="2">High tonality</td></tr>
<tr><td>Non-heightened high tonality</td><td>/t/</td></tr>
<tr><td>Heightened low tonality</td><td>$/p^j/$</td><td>$/p^j/$</td><td rowspan="2">/p/</td><td rowspan="2">Low tonality</td></tr>
<tr><td>Non-heightened low tonality</td><td>/p/</td><td>/p/</td></tr>
</table>

(b)

<table>
<tr><th></th><th>Stage 1</th><th>Stage 2</th><th>Stage 3</th><th></th></tr>
<tr><td>Heightened high tonality</td><td>$/t^j/$</td><td rowspan="2">/t/</td><td rowspan="3">/t/</td><td rowspan="3">High tonality</td></tr>
<tr><td>Non-heightened high tonality</td><td>/t/</td></tr>
<tr><td>Heightened low tonality</td><td>$/p^j/$</td><td>$/p^j/$</td></tr>
<tr><td>Non-heightened low tonality</td><td>/p/</td><td>/p/</td><td>/p/</td><td>Low tonality</td></tr>
</table>

Table 2.1: The loss of phonemic distinctions in Czech. (a) sketches the situation in non-Teták dialects. At Stage 1 there is a distinction between palatalised and plain /t/ and /p/. The distinction between the palatalised and plain dental stops is lost at Stage 2. At Stage 3 the distinction is lost for the labial stop. This results in a reduction from four distinct levels of tonality at Stage 1 to two levels of tonality at Stage 3. (b) shows the situation in the Teták dialects. Stage 2 proceeds as in (a). However, at Stage 3 the distinction is lost in a different manner, with heightened low tonality being reinterpreted as plain high tonality.

of underlying /t/ as $/p^j/$, in line with their adult models. This additional detail is not important for our purposes — the crucial point of Andersen's work is his conclusion that misinterpretation of the PLD by language learners can lead to language change.

Lightfoot (1979) attempts to provide a fairly general account of syntactic change. Lightfoot's approach is similar to Andersen's (1973) in taking note of the cultural dimension of linguistic transmission. Under Lightfoot's account, languages gradually accumulate opacity in the derivation of surface forms from underlying syntactic structures. Once this opacity exceeds some threshold tolerance level, a therapeutic alteration to the underlying grammar is made by language learners, in order to improve the transparency of derivation. The details of Lightfoot's notions of transparency of derivation are not important here — the key point is that Lightfoot, working within the Chomskyan framework, addresses the cultural nature of language transmission:

> "An individual may be exposed to PLD which is different from the parents PLD . . . One individual may set some parameter differently from older people in her community; then it is likely that, because of the grammatical change, she will produce different utterances from other people in her community. These new expressions, in turn, affect the linguistic environment, and she will now be an agent of further change, by virtue of the fact that her younger siblings will have different PLD as a result of what she produces

with her new grammar. As the younger siblings also set the relevant parameter in the manner of the older sister, so other people's PLD will differ. Thus a chain reaction is created" (Lightfoot 1999:101)

One of Lightfoot's central concerns is to emphasise the importance of a theory of grammar in understanding language change. For Lightfoot, the main constraint on language change derives from the restrictions placed on possible grammars by UG:

> "Each generation has to construct a grammar anew, starting from scratch. Speakers of a given grammar construct a grammar on the basis of the primary data available [the PLD] ... A subsequent generation constructs a grammar in the same way, but if the primary data is now slightly different the grammar hypothesized will also be different, and there is no reason why it should bear any closer formal relation to that of the parent generation beyond the defining requirements of a theory of grammar; after all, small differences in output may result in large differences in the grammar, and vice versa" (Lightfoot 1979:147)

Lightfoot is particularly keen to rule out grammar-independent principles of change, which would predict how languages would change and do so in terms abstracted away from a particular theory of grammar. He criticises Traugott's (1965) position that "[t]he objectives of diachronic linguistics have always been to reconstruct the particular steps by which a language changes, and also to hypothesize about processes of language change in general", countering "[t]he neogrammarian legacy of a search for independent principles of change must be abandoned ... the distinction outlined here between theories of grammar and change requires a change of focus, such that these predictions are derived mostly from a theory of grammar" (Lightfoot 1979:153).

Lightfoot's position is interesting for several reasons. Firstly, it shows that the Chomskyan position can amenably be applied to a study of language change, in which cultural transmission, via the PLD, plays some role. Lightfoot's objection to independent principles of change will be returned to below in connection with Hurford's conception of the Arena of Use, and also in Section 2.3.6 and Chapter 5, where research suggesting a principle of change (from less to more generalisable grammars) is presented.

Hurford (1987) presents a further elaboration of Andersen's (1973) conception of language change. Hurford closes the loop between the PLD and grammatical competence via the substrate in which communication takes place, which Hurford dubs the Arena of Use. An individual's grammatical competence is expressed in the Arena of Use, which itself imposes further communicative pressures and social and cognitive constraints. This behaviour in the Arena of Use provides the PLD upon which grammatical competence is acquired. The closing of the loop between grammatical competence and PLD via the Arena of Use is illustrated in Figure 2.3. According to Hurford, the consequences resulting from the filter imposed by the Arena between competence and linguistic output are potentially non-trivial:

> “a speaker learns the most successful ways of expressing his meanings, and the statistical shape of his output is thus influenced by his experience. It is of course conceivable that the LAD is so rich that it makes full allowance for the effect of the arena of use on linguistic output. That is, the LAD might be able to compensate fully for the ‘distorting’ factors affecting output and be able to retrieve a more or less perfect replica of the competence(s) involved in producing the output. But this strikes me as very implausible” (Hurford 1987:22)

In other words, the pressures acting on language as it passes through the Arena of Use may skew the PLD to learners so as to effect language change. The acoustic ambiguities which Andersen hypothesises lead to the change in the phoneme inventory of Czech are one possible consequence of the passage of language through the Arena of Use. Hurford proposes another, hypothetical, case, where presuppositions about the world associated with a particular culture lead to inanimate objects rarely being referred to by grammatical subjects. This consequence of the Arena of Use might lead, Hurford suggests, to children internalising a grammatical rule prohibiting the appearance of inanimate noun phrases in subject positions — the way of seeing the world, determined by non-linguistic culture, impacts via the Arena of Use on the grammar. To the extent that the consequences of the Arena are predictable, they may result in, contra Lightfoot, independent principles of change.

To summarise the positions of Andersen (1973), Lightfoot (1979) and Hurford (1987), an individual’s grammatical competence is now not solely a matter of individual psychology, but how that individual psychology applies to data provided by other individuals. Linguistic behaviour, or E-language[2], begets linguistic competence, I-languages, which beget, via the Arena of Use and the LAD, E-languages. This cultural process, which has been dubbed the Expression/Induction (E/I) cycle (Hurford 2002a) is sketched in Figure 2.4.

[2] By E-language I mean the external linguistic behaviour of individuals. The same term has sometimes been taken to refer to language as an object external to human minds. This is not the interpretation which makes most sense in this context.

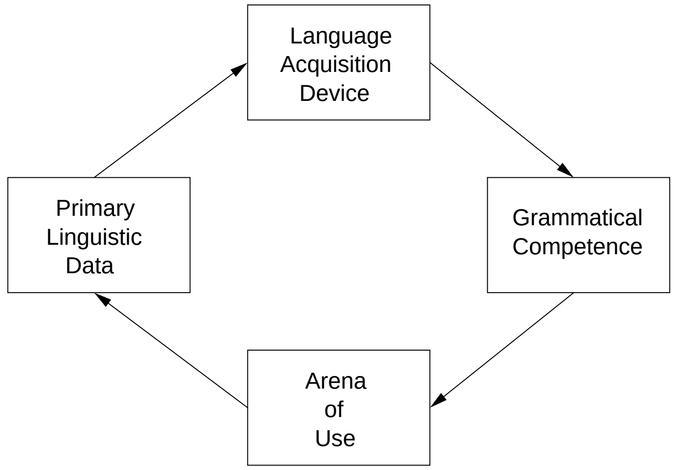

Figure 2.3: The Arena of Use in the cycle of linguistic transmission. The Arena mediates between an individual's grammatical competence and their expression of that competence as linguistic behaviour. Considerations arising from the Arena of Use may impact significantly on the PLD available to subsequent language learners, and may therefore result in language change.

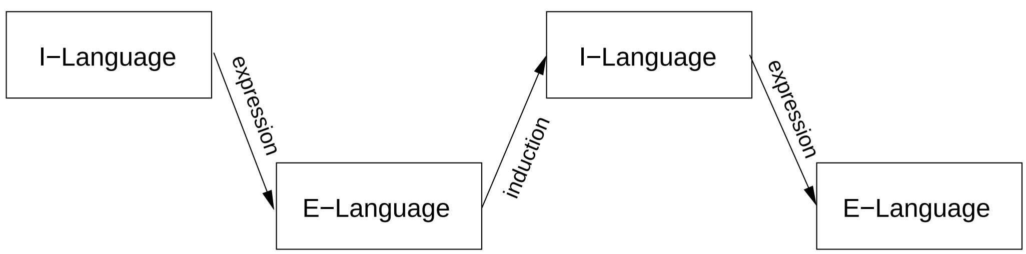

Figure 2.4: The Expression/Induction cycle. I-Language leads, via expression, to E-language. E-language leads, via induction (or acquisition, in more neutral terms) to I-language.

The E/I model constitutes a fairly general model of the cultural transmission of language. Its main shortcoming, as revealed by a comparison with B&R's general model of cultural transmission outlined in Section 2.1.1, is a lack of a formal specification of the pressures acting on language during its cultural transmission. This issue has been addressed recently with the advent of formal models of linguistic evolution, which adopt the E/I cycle as their starting point. These models can broadly be classified as either *Negotiation Models* (NMs) or *Iterated Learning Models* (ILMs). The primary methodological difference between the NM and ILM approaches is in their treatment of the verticality of cultural transmission.

B&R's general model given above is framed purely in terms of what is known as *vertical* transmission — a population is considered to consist of discrete, non-overlapping generations, with cultural transmission only taking place between generations. Intra-generational transmission is referred to as *horizontal* transmission, and is ignored in the model sketched above. However, the mathematical analyses which B&R use to investigate cultural evolution in the general model are essentially agnostic about the vertical-horizontal distinction — much of the analysis remains the same if we view a "generation" in the general model as snapshot of a monogenerational population interacting with itself.

Similarly, the E/I model is essentially agnostic regarding the vertical-horizontal distinction. In contrast, the distinction between vertical and horizontal transmission has become something of a defining characteristic in implementations of models of the cultural transmission of language. In the NM transmission is exclusively horizontal, whereas in the ILM vertical transmission is paramount. The transmission of language in the real world presumably lies somewhere between these two extremes.

2.1.2.1 The Negotiation Model

Populations in the NM consist of collections of individuals, who acquire their linguistic competence based on observations of the behaviour of other individuals. A population consists of a monogenerational collection of individuals. At each time-step members of the population produce observable linguistic behaviour, which is observed and learned from by other members of the population. This process is illustrated in Figure 2.5. There is no population turnover in the NM — new individuals do not enter the population and no individual leaves. Transmission is therefore exclusively horizontal. The NM framework is used by, for example, Hutchins & Hazelhurst (1995), Steels (1997), Batali (1998), Hazelhurst & Hutchins (1998), Steels (1998), Smith (2001a) and Batali (2002).

2.1.2.2 The Iterated Learning Model

In the ILM, as in the NM, populations consist of collections of individuals, who acquire their linguistic competence based on observations of the behaviour of other individuals. Unlike in the NM, there is population turnover in the ILM. This turnover can either be generational (see Figure 2.6 a), where the entire population is replaced at each time-step, or gradual (see Figure 2.6 b), where a single individual is replaced at each time-step. In the generational ILM individuals at generation $n + 1$ acquire their competence based on observations of the behaviour of the generation n population. In the gradual turnover ILM, each new individual acquires its competence based on observations of the behaviour of the population they enter. In either case, transmission is exclusively vertical, from fully enculturated individuals to naive individuals. The ILM framework is used in, for example, Hare & Elman (1995), Oliphant & Batali (1997), Kirby (1999), Livingstone & Fyfe

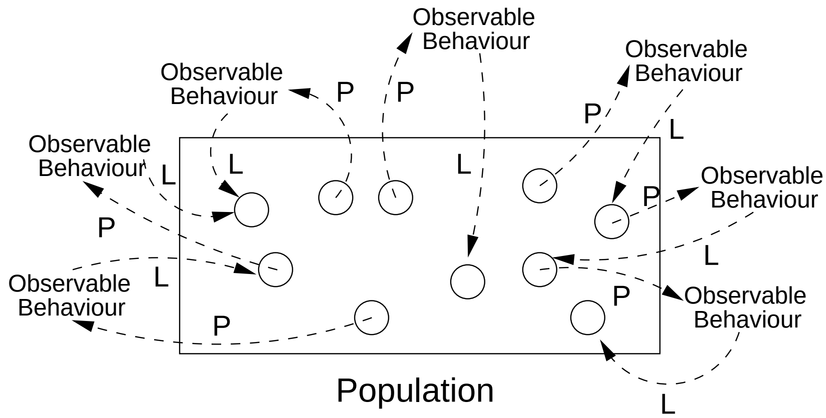

Figure 2.5: The Negotiation Model. The population consists of a collection of individuals, represented by circles. Individuals produce observable (linguistic) behaviour, which is learned from by other members of the population. There is no population turnover. Cultural transmission is therefore purely horizontal, acting within the monogenerational population.

(1999), Nowak *et al.* (1999), Oliphant (1999), Brighton (2000), Hurford (2000), Kirby (2001), Nowak *et al.* (2001), Kirby (2002), Smith (2002), Smith *et al.* (forthcoming) and Smith *et al.* (submitted).

2.2 Transmission and Cultural Traits

Given these two frameworks for studying cultural transmission (B&R's general model, and the E/I framework), several issues remain to be resolved. Firstly, how do we characterise the range of possible cultural variants? Secondly, how are these cultural traits transmitted? Section 2.2.1 describes B&R's two models of cultural traits and the transmission of such traits. Section 2.2.2 outlines the nature of cultural traits in the E/I model and the somewhat contentious issue of what information is available to learners during the transmission of linguistic structure.

2.2.1 Cultural traits and transmission in the general model

B&R provide two basic and very abstract treatments of possible cultural traits. In the *dichotomous trait* model there are two possible cultural traits, with individuals having one or the other trait. In the slightly more complex *continuous trait* model there are an unlimited number of possible cultural traits, each of which is assigned a numerical value. Each individual's cultural character is given by the numerical value of the trait they possess.

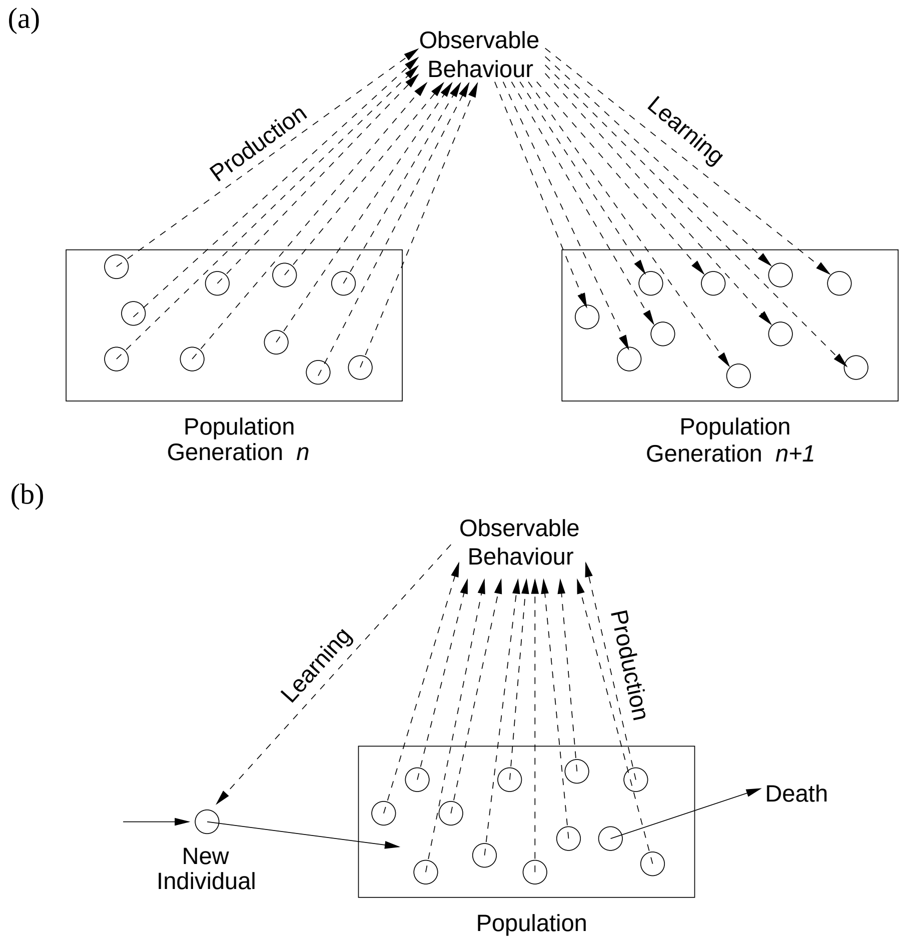

Figure 2.6: The Iterated Learning Model. In the generational version of the ILM (a), the observable behaviour produced by generation n individuals is observed and learned from by generation $n + 1$ individuals. In the gradual turnover version (b), a single individual is removed from the population at each time-step, to be replaced by a single individual. The new individual learns from the observable behaviour produced by the rest of the population. In both versions of the ILM cultural transmission is therefore purely vertical, with transmission proceeding from fully enculturated individuals to naive individuals.

B&R are not primarily concerned by how an individual's cultural trait manifests itself in that individual's behaviour. They assume that naive individuals can estimate the cultural trait possessed by mature individuals, possibly with some small error. B&R are more concerned with how the transmission of such traits can affect the cultural make-up of the population. In part, B&R can avoid being specific about how cultural traits are stored in the brain, manifested in behaviour and estimated by naive individuals because their model is very general. Firstly, their somewhat simplistic treatment of cultural traits is unlikely to specifically offend any particular specialism, or indeed offend all groups equally. Secondly, their model of cultural traits *should* be poorly specified, given their general aims of investigating how cultural transmission can result in cultural evolution, abstracted away from any particular putative cultural trait.

As shown in Section A.1.1 of Appendix A, B&R show that cultural transmission alone, of either dichotomous or continuous traits, does not result in cultural evolution — assuming that naive individuals acquire their cultural traits on the basis of a random sample of enculturated individuals, and individuals are unbiased with respect to which cultural trait they acquire, the distribution of cultural variants in the population will remain unchanged. Cultural evolution therefore does not automatically follow from cultural transmission — some pressure other than unbiased transmission is required for cultural evolution.

2.2.2 Cultural traits and transmission in linguistic models

B&R's simple characterisation of cultural variants and their assumption that the particular cultural variant an individual possesses is easily determinable from that individual's behaviour prove somewhat limiting when applied to the modelling of the transmission of language. There are three key difficulties in applying B&R's approach to models of linguistic transmission, outlined in the following three sections.

2.2.2.1 The simplicity of B&R's cultural traits

Firstly, B&R's characterisation of cultural traits (dichotomous or continuous) is too simple to capture most aspects of linguistic behaviour. Variations of their model have proved useful in certain cases. For example, Kirby (1999) (Chapter 2) considers the cultural evolution of grammars which exhibit either verb-object or object-verb order, and are either prepositional or postpositional. Kirby therefore treats grammars as pairs of dichotomous traits, and models how the distribution of the two traits impact on each other. Briscoe (2000a) characterises grammars as either head-initial or head-final, equivalent to a single dichotomous trait. Finally, Nowak *et al.* (2001) model the cultural evolution of multiple competing grammars. In B&R's terms, each grammar could be treated as a distinct integer, with grammatical competence then corresponding to a continuous trait. These models are described in more detail in Section 2.3.

However, B&R's model of cultural traits is insufficient to deal with more detailed questions of linguistic structure. Firstly, in a dichotomous or continuous trait model there is no notion of the structure *within* a particular trait. It would be possible to interpret such a model in this way, and say, for example, that trait t represents a language with agglutinating morphology and head-initial syntax. This is essentially the approach adopted by Kirby (1999) and Briscoe (2000a). There are two main problems with this approach:

1. There is no way to model how the internal structure of a morphological or syntactic system changes over time due to cultural transmission — a cultural trait is either present or absent in B&R's system, and the only possible change in a cultural trait is a change from presence in a particular individual to absence, or vice versa.
2. There is no way to investigate how the internal structure of a grammatical system impacts on its fecundity or fidelity during cultural transmission, short of explicitly imposing such factors. For example, Kirby (1999) defines a constant which determines to what extent verb-object order is favoured in the presence of prepositions, then sets this constant to some (theoretically well-motivated) value.

In addition to these problems relating to a lack of structure *within* a particular traits, there is an associated problem of a lack of structure *between* traits. In the dichotomous model there is no structural relationship between traits, other than one of dichotomy. In the continuous trait model, cultural traits are organised in a linear fashion, with the only structural relationship being one of numerical distance. Nowak *et al.* (2001) fall foul of this problem in modelling the degree of similarity between grammars, which they effectively treat as numerical values. Given the lack of structure within grammars in their model, it is meaningless to say to what extent users of two distinct grammars share grammatical structures, or to quantify the probability of a learner acquiring a particular grammar on exposure to expressions generated by another grammar. Nowak *et al.* circumvent this problem by assigning arbitrary (constant or random) degrees of similarity between grammars, a rather unsatisfactory solution.

Most implementations of E/I models therefore develop a more complex treatment of cultural traits. The nature of the model depends on the linguistic behaviour of interest. Broadly, models can be classified as investigating the cultural transmission of *vocabulary* systems, or *syntactic* systems.

In vocabulary models, an individual's competence typically consists of a mapping between a set of unstructured meanings and a set of unstructured signals. Such models are typically concerned with the impact of the population's vocabulary on communication within the population. These models therefore typically include some evaluative phase, where individuals attempt to communicate meanings to one another using their acquired mappings from meanings to signals. The communicative accuracy between two individuals is calculated according to a formula with the canonical form:

$$\text{communicative accuracy}\,(S,H) = \sum_i \sum_j p\,(s_j|m_i) \cdot r\,(m_i|s_j)$$

where S and H are speaker and hearer, $p\,(s_j|m_i)$ gives the probability of the speaker S producing signal s_j to communicate m_i and $r\,(m_i|s_j)$ gives the probability of the hearer H interpreting signal s_j as communicating meaning m_i. It should be noted that 1) the evaluation of communicative accuracy is typically distinct from the cultural transmission phase and 2) communicative accuracy typically plays no role in shaping the communication system of the population and 3) implicit in this measure is the assumption that meanings are functionally distinct — for example, if two meanings m_i and m_j result in the same behaviour on the part of the receiver and $r\,(p\,(m_i)) = m_j$ then the communication would be measured as a failure but could, at the behavioural level, be considered a success.

The definition of communicative accuracy above implies a definition of communication which is similar to that of Johnson-Laird (1990), who states that "the communicator [must] construct an internal representation of the external world, and then ... carry out some symbolic [not necessarily in the strict sense I have used in Chapter 1] behaviour that conveys the content of that representation. The recipient must first perceive the symbolic behaviour, i.e. construct its internal representation, and then from it recover a further internal representation of the state that it signifies" (Johnson-Laird 1990:2–4). Communication systems are therefore defined as systems for mapping between meanings and signals. Communication is a success where the internal representations of communicator and recipient are the same.

In syntactic models, an individual's competence typically consists of some system, possibly principled, for mapping between structured meanings and structured signals (for example, a context free grammar with semantic operations associated with rewrite rules). Syntactic models tend to be less concerned with how cultural evolution impacts on communication within a population, and more concerned with how the system for mapping from meanings to signals changes over time. A key question tends to be the extent to which the structure of meanings and signals is exploited during the meaning-signal mapping process — do compositional mappings emerge?

Examples of these models will be described in more detail later in this Chapter. The key point here is that a much more complicated model of cultural traits is used. This has several benefits. Firstly, interesting linguistic behaviour can be modelled in a less abstract way. An explicit model of a meaning-signal mapping, be it a vocabulary-type mapping or a syntactic mapping, allows measures of structure within and between different individual's acquired cultural traits to be fairly naturally defined. However, this enriched treatment does have drawbacks. Firstly, the notion of cultural traits becomes somewhat fuzzy. For example, in a vocabulary system, is a cultural trait an individual's whole system of mapping from meanings to signals, or could an individual's cultural character be considered as consisting of several traits, with each trait specifying a single meaning-signal association? Secondly, the more complex models make the type of mathematical analysis used by B&R difficult, if not impossible. There is a tradeoff between the richness of the treatment of cultural traits and the transparency of the model's results. One of the goals of this Chapter is to relate B&R's simple models to the more complex, linguistically-flavoured models, allowing the rich results generated by the latter to be interpreted in the simple, clear terms of the former.

2.2.2.2 Is the cultural transmission of language possible?

A second major problem with applying B&R's approach to linguistic evolution is identifying what cultural traits correspond to in linguistic theory. In terms of the E/I framework, there are two possibilities — a cultural trait could characterise an individual's linguistic competence, their I-language, or their linguistic behaviour, their E-language. Our preference, implicit in the discussion above, should perhaps be towards interpreting an individual's cultural trait as corresponding to their I-language. E-language is derived from I-language, and is contingent on the set of situations which require this internal competence to be pressed into service to produce linguistic behaviour. With this interpretation, the range of possible cultural variants is circumscribed by the limitations imposed by the representation of I-language. An individual's externally visible manifestation of their cultural trait is the E-language they produce, which is determined by factors related to considerations imposed by the Arena of Use.

However, this interpretation throws up another problem. I-language, as is apparent from Figures 2.2, 2.3 and 2.4, is not directly transmitted, but is acquired via the PLD, which in turn comes from E-language. B&R assume that an individual's cultural trait can be determined from their behaviour either straightforwardly (in the dichotomous trait model) or with some normally-distributed error (in the continuous trait model). This is too simplistic an approach to dealing with the transmission of I-language. To take an extreme view, it could be argued that the filtering through E-language makes it difficult, or inaccurate, to speak of I-language being transmitted at all. This would be the position favoured by Chomsky ("it seems that a child must have the ability to 'invent' a generative grammar" (Chomsky 1965:201)) and Lightfoot ("[e]ach generation has to construct a grammar anew, starting from scratch" (Lightfoot 1979:147)). This argument could be applied to the cultural transmission of any trait. For example, how can a naive individual guess at the internal state of its cultural parent that determines political persuasion, based on behaviour?

Two points should be made to at least cast doubt on this very strong negative position. There is a large body of evidence suggesting that cultural transmission is a fact for non-linguistic traits — B&R cite 57 articles in a brief review of evidence pointing towards the reality of cultural transmission (Boyd & Richerson 1985:47–55). It is therefore possible that cultural transmission also applies to language. With particular reference to language, the fact that speakers within a speech community tend to agree on what constitute valid grammatical sentences and so on should make us doubt that they all have radically different I-languages. The most parsimonious explanation for this is that I-language is, to some extent, culturally transmitted.

If we reject the strong negative view, we are left with a wide range of possible degrees of cultural transmission. The strongest positive view would be that the filter through E-language is completely irrelevant. This too seems unreasonable. Firstly, there may be several (or indeed infinitely many) possible grammars which are consistent with a particular PLD. A learner bound to be consistent with the observed PLD might therefore converge on a different grammar from the grammar which produced that PLD. Whether or not learners are bound to be consistent with their PLD is another issue. Both Andersen (1973) and Lightfoot (1979) assume that they are. However, radical restructuring events such as creolization suggest that this constraint may be fairly weak. A second point against the strongly positive view of cultural transmission of I-language would be Hurford's (1987) comments about the possible skewing effects introduced by the Arena of Use.

Where does this leave us? B&R's assumption of unproblematic cultural transmission is too simple when it comes to language — the filtering of I-language through E-language is unlikely to be so trivial as be ignorable. However, I-language probably is transmitted to some extent — we should reject Chomsky's and Lightfoot's position that children invent their I-language anew each generation.

2.2.2.3 The nature of the PLD

A subsidiary question is: what actually constitutes the PLD? The only aspect of linguistic behaviour which is uncontroversially determinable is the acoustic (or visual) sequence produced by an individual when that individual speaks (or signs). However, most implementations of the E/I model (NMs and ILMs) assume that this observable signal is paired with the communicative intention of the individual producing the signal — the PLD consists of meaning-signal pairs. This is not an uncontroversial assumption, nor is it an assumption which is always made.

In most implementations of the E/I model there is no explicit modelling of an environment outwith the individuals that make up the population — meanings and signals in the model are arbitrary agent-internal representations. It is typically assumed that there is some shared, stable mapping between external situations and internal representations of those events — indeed, cultural transmission of linguistic structure would be impossible without an external manifestation of internal representations of signals. However, this mapping is not typically the focus of E/I models, which concentrate instead on the mapping between internal representations of meanings and signals. An account of language as a mapping from aspects of the environment to other aspects of the environment must account for two additional mappings (see Figure 2.7):

- the mapping between states of the environment representing situations to be communicated and internal representations of those states (meanings).
- the mapping between communicative alterations of the environment and internal representations of those alterations (signals).

The nature of the mapping between environment and meaning forms a key part of the symbol grounding problem (Harnad 1990) — the meaning of internal representations must, at some point, be related to the objects or situations in the world which they refer to. The mapping between environment and signal corresponds to a mapping between strings of phonemes and articulatory movements. Most E/I models simply assume that these two mappings from internal representations to the environment are shared by all simulated individuals.

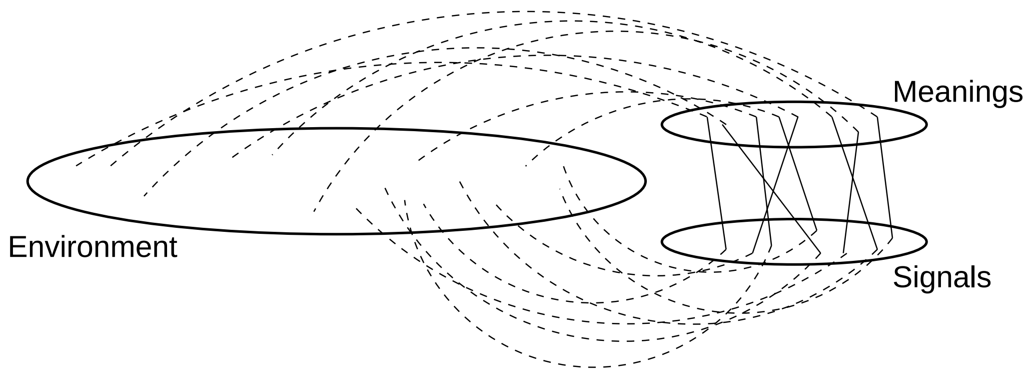

Figure 2.7: Language as a mapping between three spaces (represented as ellipses). Typically, E/I models focus on the mapping between two spaces — the internal representational spaces of meanings and signals. This mapping is given in solid lines. A complete model must account for two additional mappings — the mappings between the environment space and the internal representational spaces (dashed lines).

Given the absence of an explicit model of the environment, the assumption that learners observe meaning-signal pairs is unavoidable. In a fuller model, the more reasonable assumption could be made that learners are exposed to an environment which includes a state being communicated about and a set of articulatory gestures intended as a communicative alteration to the environment. The learner then has the task of identifying the communicatively relevant state and the communicative alteration, representing both internally and then learning the mappings between the internal representations and possibly the mappings between the internal representations and the environment.

There is a body of evidence which suggests that children have various strategies for mapping from the environment to internal representations of relevant parts of the environment. Much of this points to the importance of joint attention and intentional inference. Studies by Baldwin (Baldwin 1991; Baldwin 1993a; Baldwin 1993b) show that infants cannot learn words for toys simply by hearing the word for the toy while attending to the toy. The child must witness an intentional agent direct their attention to the toy while naming it. Under these circumstances the infant will learn the word for the toy, even if there is a delay between witnessing the intentional agent directing their attention at the toy and being able to attend to the toy directly themselves.

While most E/I models ignore the issue of the environment-internal representation mappings, some computational modelling work within the E/I framework has sought to tackle this problem. Neural network models show that genetic evolution can lead to the formation of internal representations which correspond to a categorisation of the environment (Cangelosi & Parisi 1998). These internal representations may form the basis of a (partially) culturally-transmitted communication system (Cangelosi 1999). Hazelhurst & Hutchins (1998) show that the negotiation of ritualised shifts of joint attention subserves the emergence of a learned communication system. Symbolic computational models demonstrate that shared mappings from the environment to internal representations of meanings can emerge through individual learning, both with explicit feedback (e.g. Steels (1997), Steels (1998)) and without (Smith 2001a). Finally, it has been demonstrated that repeated expression and induction of strings of words can lead to the emergence of meaning, where meaning is defined in terms of the relationship between words and other words — the emergent mesh of word-word associations constrains and guides the interpretation of signals (Hashimoto 1998). These models and the data from real language acquisition outlined above give some hope that an integrated model, of the type depicted in Figure 2.7, will be achievable. However, for the purposes of this thesis I will assume that the PLD available to learners consists of meaning-signal pairs.

2.3 Forces acting on cultural transmission

The general model of cultural transmission provided by B&R and the more language-specific E/I models raise several interesting issues on the nature of cultural traits and the manner in which these traits are transmitted. However, these questions are typically not the central concern of such models, nor are they of great importance for the purposes of this thesis. Of more interest is how the processes involved in cultural transmission outlined in Section 2.1.1 (transmission itself, individual learning by enculturated individuals, and selective removal of enculturated individuals) impact on the distribution of cultural traits in the population. Do any of these processes lead to significant cultural evolution?

The models of B&R are used to frame much of this discussion. They provide mathematical accounts of how three pressure acting on transmission can result in cultural change and cultural evolution. These are:

1. Natural selection of cultural variants, resulting from selective removal of enculturated individuals.
2. Guided variation, resulting from individual learning by enculturated individuals.
3. Biased transmission, resulting from the strategy of learners during cultural transmission. The forces of biased transmission can be further subdivided into three forms:
   (a) Directly biased transmission, resulting from a preference for learners to acquire one cultural variant over another.

(b) Indirectly biased transmission, resulting from a preference for learners to acquire cultural traits which are associated with other cultural traits.
(c) Frequency-dependent transmission, resulting from a disproportionate preference for learners to acquire the most (or least) frequent cultural trait in the population.

In Sections 2.3.1 to 2.3.5 B&R's models for these pressures are reviewed. In the interests of clarity, a separate section is devoted to each of the three subtypes of biased transmission. Mathematical details are given in Appendix A.

The main goal of this part of the thesis is to relate B&R's simple models to the more complex, linguistically-flavoured models. This allows the linguistically-interesting, somewhat complex results generated by the linguistic models to be interpreted in the simple, clear terms of B&R's models. To this end, each section describing one of B&R's pressures on cultural transmission is followed by one or more sections which introduce details of a linguistic (usually E/I) model, and discusses how that model can be interpreted in terms of B&R's model.

Finally, in Section 2.3.6, a new driving force for cultural evolution is introduced. This pressure, which arises from transmission through a bottleneck, does not feature in B&R's taxonomy, but is extremely relevant to the cultural transmission of linguistic structure.

As the structure of this Chapter is somewhat intricate, a full preview is perhaps in order.

Section 2.3.1 : Natural selection of cultural variants
- **Section 2.3.1.1** : B&R's model
- **Section 2.3.1.2** : Linguistic case study: the evolution of vocabulary under natural selection.
- **Section 2.3.1.3** : Linguistic case study: the evolution of grammar under natural selection.

Section 2.3.2 : Guided variation
- **Section 2.3.2.1** : B&R's model
- **Section 2.3.2.2** : Linguistic case study: Tomasello's cultural ratchet

Section 2.3.3 : Directly biased transmission
- **Section 2.3.3.1** : B&R's model
- **Section 2.3.3.2** : Linguistic case study: the evolution of consistent head ordering under direct bias
- **Section 2.3.3.3** : Linguistic case study: the evolution of vocabulary under direct bias

2.3.1 Natural selection of cultural variants

Natural selection was originally conceived of by Darwin as a mechanism for explaining the appearance of design in biological organisms.

> "Owing to this struggle for life, variations, however slight and from whatever cause proceeding, if they be in any degree profitable to the individuals of a species, in their infinitely complex relations to other organic beings and to their physical conditions of life, will tend to the preservation of such individuals, and will generally be inherited by the offspring. The offspring, also, will thus have a better chance of surviving, for, of the many individuals of any species which are periodically born, but a small number can survive. I have called this principle, by which each slight variation, if useful, is preserved, by the term Natural Selection." (Darwin 1859/1964:61)

Modern definitions appeal to three factors mentioned in Darwin's original formulation — variation, inheritance and selective survival or propagation. While there is some ongoing debate about the precise formulation of a definition of natural selection, Futuyma (1998) concludes that "[m]ost authors agree that the definition must include the following concepts: some attribute or trait must vary among biological entities, and there must be a consistent relationship, within a defined context, between the trait and one or more components of reproductive [implying heredity] success, where 'reproductive success' includes both survival (a prerequisite for reproduction) and the reproductive processes themselves." (Futuyma 1998:349).

This definition is actually rather general, excepting the single reference to "biological entities". Donald Campbell (Campbell 1965; Campbell 1975) presents an extension of the principle of natural selection to account for cultural evolution. Campbell's main contribution is to present an argument that the three central factors (variation, inheritance and selective retention or reproduction) required for natural selection occur in cultural systems. If cultural systems exhibit variation within or between populations, if cultural systems are in some sense heritable, and if there is selection either in the survival of cultural systems or selection in the elevation to roles allowing influence in the enculturation of others, then we should expect to see cultural evolution under natural selection. Campbell is optimistic that culture exhibits variation and selection, but acknowledges that "retention and duplication [inheritance], is also more problematic for social evolution than for biological evolution. What are required are mechanisms for loyally reproducing the selected variations." Campbell concludes that such inheritance systems are possibly present in culture — "through social mechanisms of child socialization, reward and punishment, socially restricted learning opportunities, identification, imitation, emulation, indoctrination into tribal ideologies, language and linguistic meaning systems, conformity pressures, social authority systems, and the like, it seems reasonable to me that sufficient retention machinery exists for a social evolution of adaptive social belief systems and organizational principles to have taken place" (Campbell 1975:1107).

2.3.1.1 B&R's model

B&R model the natural selection of cultural variants by assuming that there are a set of distinct social roles (e.g. mother, father, uncle, priest, teacher). Each naive individual acquires their cultural characteristic based on observation of a subset of these roles. In order to model natural selection we must assume that the probability that an individual attains a particular social role depends on the cultural variant that that individual possesses. As shown in Section A.1.2.1 of Appendix A, if a particular cultural variant c offers a selective advantage when averaged over social roles (i.e. individuals with variant c are, on average, more likely to attain a social role than individuals with some other cultural variant) then it will increase in frequency in the population. In other words, if possessing variant c makes an individual more likely to occupy a role which allows them to enculturate others and transmit that variant, then c will increase in frequency in the population. The rate of increase of the favoured variant is dependent on cultural variation in the population — in the extreme case, where the population exhibits no variation, natural selection of cultural variants has no impact.

2.3.1.2 Linguistic case study 1: the evolution of vocabulary under natural selection

Martin Nowak and colleagues have applied techniques from theoretical biology to the study of language evolution. With the exception of one published work (Nowak *et al.* (2000), discussed in Chapter 6), the work of Nowak *et al.* falls within what B&R describe as natural selection models of cultural evolution. Nowak *et al.* make the assumption, following the lead of Pinker & Bloom (1990), that "[i]t pays to talk. Cooperation in hunting, making plans, coordinating activities, task sharing, social bonding, manipulation and deception all benefit from an increase in expressive power. Natural selection ... can certainly see the consequences of communication" (Nowak & Komarova 2001:288). Based on this assumption, Nowak *et al.* use essentially the same model to study the evolution of symbolic vocabulary (described here) and universal grammar (as described in Section 2.3.1.3).

In Nowak *et al.* (1999), individuals are required to communicate about n objects using m signals. Each individual is characterised by an *association matrix*, A, which is an $n \times m$ matrix where entry a_{ij} gives the number of times during learning that that individual has observed its cultural parents communicating about object i using signal j. From this A matrix P and R matrices can be derived[3]. An individual's P matrix is an $n \times m$ matrix, where entry p_{ij} gives the probability of that individual producing signal j to communicate about object i. An individual's R matrix is an $m \times n$ matrix, where entry r_{ij} gives the probability of that individual associating signal i with object j when acting as a receiver. P and R are derived from A by normalising over rows and columns respectively:

$$p_{ij} = \frac{a_{ij}}{\sum_{l=1}^{m} a_{il}}$$

$$r_{ji} = \frac{a_{ij}}{\sum_{l=1}^{n} a_{lj}}$$

The communicative payoff for two individuals I_1 and I_2, where individual I_k is characterised by A matrix A_k, is:

[3]Nowak *et al.* refer to the R matrix as the Q matrix. However, R is used here to avoid confusion with the Q matrix introduced in Nowak *et al.* (2001), which serves a completely different purpose.

$$F\left(I_1, I_2\right) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m}\left(p_{ij}^{(1)} r_{ji}^{(2)} + p_{ij}^{(2)} r_{ji}^{(1)}\right)$$

where $p_{ij}^{(k)}$ is an entry in the P matrix derived from A_k and $r_{ij}^{(k)}$ is an entry in the R matrix derived from A_k. This equation states that the payoff for communication between I_1 and I_2 is the average of I_1's ability to communicate about objects to I_2 and I_2's ability to communicate about objects to I_1, averaged over all objects. The communicative payoff for an individual I_I with respect to a population of N individuals I_1 to I_N is:

$$F_I = \sum_J F\left(I_I, I_J\right)$$

where $J = 1, \ldots, N$ but $J \neq I$.

An individual arrives at their A matrix by sampling the production behaviour of K cultural parents. For each cultural parent, the individual observes them produce a signal for every object k times, according to the parent's P matrix. For each observation of a cultural parent producing signal j in association with object i the learning individual increments the value of a_{ij} with probability $1 - \rho$ and increments the value of $a_{ik \neq j}$ with probability ρ. ρ therefore gives the probability of errors during learning.

Nowak *et al.* consider two models of cultural transmission. In the first case, individuals select K cultural parents at random from the preceding generation of the population. In this case there is no natural selection on cultural transmission — an individual's communicative accuracy, a consequence of their culturally-acquired communication system, does not influence the probability of that individual acting as a cultural parent. Nowak *et al.* report two results for this case:

1. when $\rho = 0$ (learning is error free) the populations converge on sub-optimal communication systems. The population converges on a shared random binary (all entries are either 0 or 1) A matrix. Typically, these random matrixes will result in some intermediate level of communicative payoff. Fairly minor effects were found for different values of k (number of exposures to object-signal pairs) and K (number of cultural parents). For $k = 1$ or $K = 1$ convergence is rapid. For $k > 1$ or $K > 1$ convergence is somewhat slower, but the populations converge to similar levels of communicative payoff.

2. when $\rho > 0$ communicative payoff decreases, with shared communication systems failing to emerge when $\rho \geq 0.01$.

In their second model of transmission, each individual selects K cultural parents from the preceding generation, with the probability of any individual I being selected as a cultural parent being $F_I / \sum_J F_J$. This amounts to natural selection during cultural transmission, with systems which result in higher communicative payoff being more likely to be transmitted. Nowak *et al.* report three results for this case:

1. when $K = 1$ (individuals have a single cultural parent) and $0 < \rho < 0.01$ the populations converge on a range of communication systems, some of which offer maximal payoff and some of which are sub-optimal. The overall level of payoff is higher than for the no-selection case. $k = 1$ (a single exposure to each meaning-signal pair) results in the fastest convergence, but the eventual level of communicative payoff is independent of k.
2. when $K > 1$ (individuals have multiple cultural parents) and $\rho = 0$ (learning is error-free) the populations converge on a range of communication systems, some of which are suboptimal. Convergence is slower than in the $K = 1$ case, but the average payoff of the final systems is higher.
3. when $K > 1$ and $\rho > 0$ (learning is subject to errors) the average payoff of the populations depends on ρ. When $0 < \rho < 0.01$ the populations behave approximately as they did when $\rho = 0$, with some populations converging to optimal systems and some converging to suboptimal systems. When $\rho > 0.01$ the populations fail to converge on any shared communication system. However, when $\rho = 0.01$ all populations converge on an optimal system. Nowak *et al.* do not explore further, to identify the width of this "sweet spot" for ρ.

These results essentially meet the predictions of B&R's dichotomous trait model. In the absence of selection, the distribution of cultural variants remains unchanged except for changes introduced by random factors, which B&R do not consider. For the selection case, cultural variants which maximise some fitness function become more frequent, although this increase in frequency is dependent on variance in the population. Both $K > 1$ (multiple cultural parents) and $\rho > 0$ (errors during learning) introduce variance in the population, although there appears to be a sweet spot for ρ. Nowak *et al.* only experiment with a relatively limited range of values of K however, so it is not possible to tell if there is a sweet spot for K.

The model does offer several advances on the very simple model provided by B&R, however. There is a far wider range of possible cultural variants (there are a potentially infinitely many A matrices, although the set of culturally-stable A matrices is much smaller, being restricted to the m^n possible binary matrixes which have a single signal associated with each meaning). Secondly, the cultural fitness function depends on the structure of the cultural variants involved, rather than being arbitrarily assigned.

2.3.1.3 Linguistic case study 2: The evolution of universal grammar

Nowak *et al.* (2001) use a similar technique to study the evolution of universal grammar. In their model, a universal grammar U consists of a set of n grammars, numbered 1 to n. Each grammar G_i could be described as a rule system that defines a mapping between syntactic representations and semantic representations. σ_1 is the finite syntactic alphabet and σ_2 is the finite semantic alphabet. σ_1^* is the countably infinite set of all possible strings of characters drawn from σ_1, representing the set of all possible syntactic representations. Similarly, σ_2^* is the countably infinite set of all possible meanings, where meanings are strings of characters drawn from σ_2. A grammar G_i specifies a (potentially infinite) subset of $\sigma_1^* \times \sigma_2^*$, a mapping between semantic and syntactic representations. However, Nowak *et al.* do not exploit this notion of a grammar, as we will see — the behaviour of grammars with respect to communication and learning is essentially arbitrary.

As in Nowak *et al.* (1999), a measure of similarity among different communication systems is required. For the case of the grammar model, this is given by the probability that a meaning-signal pair present in G_i is present in grammar G_j. This probability is denoted[4] by c_{ij}, and it is assumed that $c_{ii} = 1$. This measure of overlap between two grammars allows for a straightforward definition of the communicative payoff between two grammars, $F(G_i, G_j)$:

$$F(G_i, G_j) = \frac{1}{2}(c_{ij} + c_{ji})$$

Notice the similarity to the communicative payoff equation given in the previous section — the form is identical, but the summation and production-reception calculations are parcelled up into the measure c_{ij}. The fitness of a grammar G_i with respect to a population is given by

[4] Nowak *et al.* (2001) use a_{ij} to denote the probability that meaning-signal pairs generated by G_i are acceptable to G_j. However, the notation c_{ij} is used to avoid confusion with the A matrix notation in Nowak *et al.* (1999).

$$f_i = \sum_j x_j F(G_i, G_j)$$

where x_j is the frequency of grammar j in the population. This is exactly equivalent to the equation used by Nowak *et al.* (1999) in their model of the evolution of vocabulary.

Finally, a model of learning is required. Nowak *et al.* assume that children attempt to learn the grammar of their parents. Q_{ij} is the probability that a child whose parent uses grammar G_i will acquire grammar G_j. The dynamics of the population are then specified by:

$$\dot{x}_i = \sum_{j=1}^{n} x_j f_j Q_{ji} - \phi x_i$$

where ϕ is the average fitness of the population, $\phi = \sum_i x_i f_i$. The change in frequency of a grammar G_i therefore depends on the product of the frequency of some grammar G_j, the fitness of that grammar relative to the average fitness of the population, and the probability of learning grammar G_i based on exposure to grammar G_j, summed over all grammars. This is clearly a model of natural selection acting on cultural transmission, given that the change in frequency of a grammar depends on the fitness of that grammar. However, the frequency of a grammar also depends on its learnability — a grammar with below-average Q_{ji} will decrease in frequency due to its being difficult to learn, unless it offers above-average communicative payoff. This model can therefore, depending on the choice of values of Q, model directly-biased transmission (to be discussed in Section 2.3.3) in addition to natural selection. However, Nowak *et al.* do not pursue this avenue.

Nowak *et al.* report three results for this model.

1. For the case where learning is error-free (i.e. $Q_{ii} = 1$, $Q_{ij} = 0$ where $j \neq i$) there are n stable equilibria where one grammar completely dominates the population ($x_i = 1$ and $x_j = 0$ for all $j \neq i$).
2. For high error rates, the only stable solution is the case where each grammar occurs in the population with approximately equal frequency ($x_i \approx 1/n$).
3. For the special case where all grammars are equidistant (i.e. $c_{ii} = 1$ and $c_{ij} = c$ for all $j \neq i$, where $0 \leq c \leq 1$, and $Q_{ii} = q$ and $Q_{ij} = (1-q)/(n-1)$ for $j \neq i$) there is one symmetric solution where all grammars occur with equal frequency ($x_i = 1/n$). There are also several possible asymmetric solutions where one grammar G_i is dominant (although does not necessarily completely dominate

the population) and all other grammars occur with equal frequency. These asymmetric solutions will be stable provided that a *coherence threshold*, q_1 is met — if $q > q_1$ (the probability of acquiring the same grammar as your parents exceeds some value given by q_1) then the asymmetric solutions will be stable. When q exceeds a second threshold q_2 the symmetric solution becomes unstable and only the asymmetric solutions are stable. The key point here is that q will tend to decrease as n increases, given that more possible grammars leads to a greater probability of selecting the wrong grammar during acquisition. Nowak *et al.* conclude that Universal Grammar must restrict the range of possible grammars such that q remains above the threshold q_1, therefore allowing the possibility of grammatical coherence within a population.

Nowak *et al.* then consider how q changes for incremental and batch learners as the number of sample sentences increases. They also consider how the population behaves where all grammars are not equidistant from one another. For this case they assume that the values of c_{ij} where $j \neq i$ are randomly selected from a normal distribution in the range 0 to 1. Given this assumption, the overall behaviour is broadly similar to the case where all grammars are equidistant.

Their model allows Nowak *et al.* to discover a fairly fundamental result regarding the relationship between the number of possible grammars, the number of sentences a learner is exposed to, the probability of acquiring the correct grammar and the behaviour of populations of such learners. However, there are several undesirable aspects to the model. The notion of a grammar is a fairly impoverished one — grammars specify allowable mappings between meanings and signals, which is a fairly standard assumption, but there is no notion of structure in a grammar or of any interplay between the sentences a grammar allows and the sentences it does not. The model of a grammar essentially boils down to a model of a very large communication system of the type modelled in Nowak *et al.* (1999). This is at odds with the commonly-held view that language is of a different type to agrammatical vocabulary, rather than simply being an extremely large vocabulary.

The other undesirable feature of their model, partially due to their simple model of a grammar, is that the utility and learnability of a grammar is entirely arbitrary. In their earlier work on communication systems, the structure of the object-signal mapping impacted on its functionality and, indirectly, on its learnability, with the most functional and stable A matrix being a binary matrix with a distinct signal for each object. In their paper on UG, values of q and c are arbitrarily assigned. While it would, in principle, be possible to calculate values of q and c for grammars which specified a non-infinite set of sentences, this would be undesirable for two reasons. Firstly, in restricting the grammars to only those which specified finite sets of sentences, the model of grammars would obviously reduce to a model of a large communication system, which as outlined above is undesirable. Secondly, the calculation would still depend on the sets of sentences allowed by grammars, rather than any internal structure of the grammar that specified the set of allowable sentences. The fact that grammars are modelled as essentially arbitrary sets of sentences, rather than collections of rules generating sets with some logical internal structure, makes any calculation of values of q and c essentially meaningless.

2.3.2 Guided variation

In B&R's taxonomy of pressures on cultural transmission, guided variation "results from the cultural transmission of the results of learning and acts to increase the frequency of traits that best satisfy the learning criteria" (B&R p 174). Individuals acquire their initial value for the phenotype through cultural transmission, then modify this value through individual, adaptive learning. The population of fully-matured phenotypes then acts as cultural parents for the next generation. If individual learning prefers one particular phenotype then that phenotype will be disproportionately represented in the distribution of phenotypes observed by learners at the next generation.

2.3.2.1 B&R's model

A model of guided variation requires a model of individual learning. B&R assume that individual learning takes as its starting point a culturally-acquired trait, and then moves this trait towards some optimal value specified by the external environment — individual learning is a process of adaptation to the environment. After cultural transmission and individual learning, an individual is considered mature, and can act as a cultural parent. Cultural parents transmit their traits, which are determined both by cultural transmission and individual learning, to naive individuals.

B&R show (see Section A.1.2.2 of Appendix A) that when individual learning is powerful (individuals are free to make large adjustments to their culturally-acquired trait during individual learning) the population moves towards the value of the phenotype favoured by the environment, due to the transmission of cultural traits favoured by individual learning. In contrast, when individual learning is weak (individual learning tends to conserve their culturally-acquired trait) the mean value of the population's cultural trait remains unchanged by individual learning — no cultural evolution takes place.

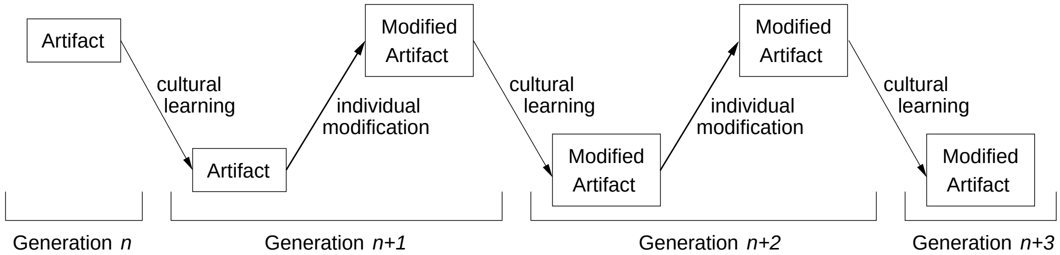

Figure 2.8: The cultural ratchet. A cultural artifact is passed from generation to generation by cultural learning. Each generation makes modifications to the artifact, which are subsequently transmitted.

2.3.2.2 Linguistic case study: Tomasello's cultural ratchet

Tomasello (Tomasello 1993; Tomasello 1999) has proposed a fairly general model of cultural transmission, which he terms "cumulative cultural evolution" or "the [cultural] ratchet effect". While this model might be equally at home in the section on general cultural models, it is included here as Tomasello's focus is primarily on the transmission of the products of individual learning.

The cultural ratchet is depicted in Figure 2.8 and proceeds as follows. Some cultural artifact is acquired, through cultural learning, by children. Those children then mature and make modifications to the cultural artifact to improve its functionality. The modified cultural artifact is then acquired, via cultural learning, by the next generation of children. This cultural transmission must be of sufficiently high fidelity to prevent the loss of earlier modifications — the ratchet must not slip backwards.

This model of the accumulation of modifications which are introduced by goal-directed adjustment of culturally-transmitted artifacts is clearly similar to B&R's model of guided variation, where goal-directed individual learning modifies culturally-transmitted traits. Tomasello's cultural ratchet theory is not supported by a formal model but his assertion that this process will result in well-adapted cultural artifacts is supported by B&R's model, which shows that a combination of social and individual learning can result in characteristics which are suited to the requirements of the environment.

Tomasello views this cultural ratchet theory as an explanation for all complex artifacts, such as tools, religious rituals, mathematics and governmental institutions. He makes particular reference to language, which in his view is established through the same processes as other cultural artifacts:

> "the way human beings have used objects as hammers has evolved significantly over human history. This is evidenced in the artifactual record by various hammer-like tools that gradually widened their functional sphere ... it is presumably the case that some cultural conventions and rituals (e.g. human languages and religious rituals) have become more complex over time as well, as they were modified to meet novel communicative and social needs" (Tomasello 1999:37)

Tomasello suggests that speakers introduce new words or constructions in order to meet novel communicative needs, or grammaticalise loosely-structured, commonly-occurring discourse structures into syntactic constructions, to improve the functionality of the linguistic system.

The explicitly teleological source of modifications to cultural artifacts (where by "teleological" I mean "designed with purpose in mind") is relatively unproblematic when applied to the cumulative modification of artifacts such as hammers and governmental institutions. However, the assumption that language changes through repeated, conscious, goal-directed modification is somewhat controversial. Lass (1980) rules out the possibility of purposeful modification — "[linguistic] change does not involve (conscious) human purpose (which I think can be accepted without argument)" (Lass 1980:82).

In addition to appealing to the established orthodoxy in linguistics that teleology plays little role in language change, a further criticism can be made of Tomasello's view of linguistic evolution. The burden of this type of teleological change must rest with the speaker — a hearer cannot decide to understand innovatively, therefore new constructions must be introduced by speakers. However, unless we assume speaker altruism this may lead to innovations which are highly non-functional with respect to the hearer.

Teleology may reasonably be expected to play a role in the introduction of new words into a new language — new technological innovations, for example, typically require new words to name them. However, while teleology may account for the introduction of such terms it perhaps cannot account for their subsequent diffusion. For example, we might be tempted to explain the preference for the term "mobile phone" over "cell phone" in English speakers in the British Isles in terms of conscious choice by language users in favour of the more functional variant. However, English speakers in North American typically favour "cell phone". It is possible that "cell phone" is more functional than "mobile phone" in the context of American English but not British English, in which case the teleological explanation of diffusion of the new compound could still apply. This would require a case-by-case analysis of the functionality of words which differed between varieties of English (e.g. "pavement" and "sidewalk" in addition to "mobile phone" and "cell phone" between British and American English, "turnip" and "swede" between Scottish English and non-Scottish English etc). However, it may be more parsimonious to allow that teleological explanations play a fairly limited role in the cultural evolution of language.

2.3.3 Directly biased transmission

Biased transmission arises when naive individuals are more likely to adopt one cultural variant than another, or when mature individuals are more likely to produce one cultural variant than another when acting as a model. In B&R's model, biased transmission is typically conceived of as "arising from the attempts of [individuals] to evaluate the adaptiveness (that is, their effects on genetic fitness) of the different cultural variants" (B&R, p134). B&R identify three subclasses of bias on transmission: direct bias, indirect bias and frequency-dependent bias. Directly-biased transmission will be discussed in this section, with models of indirect bias and frequency-dependent bias being discussed in Sections 2.3.4 and 2.3.5 respectively.

Direct bias occurs when one cultural variant is intrinsically more attractive than others. This intrinsic attractiveness makes naive individuals more likely to acquire that cultural variant. This attractiveness could derive from several sources. For example, individuals could prefer to acquire variants which they believe will be most successful, in which case the direct bias will be in favour of the cultural variants which offer the greatest fitness payoff. Alternatively, individuals could preferentially acquire one variant over another due to an arbitrary preference not necessarily related to functionality.

2.3.3.1 B&R's model

Direct bias can be simply modelled by assuming that individuals are disproportionately likely to acquire one cultural trait, to the detriment of other cultural traits. If individuals are disproportionately likely to acquire a particular cultural trait at the expense of all other cultural traits, then individuals are directly biased in favour of that trait. As shown in Section A.1.2.3, directly biased transmission will increase the frequency of the favoured variant in the population. The rate of increase depends on the strength of the bias — a weak bias will result in slow convergence of the favoured variant, whereas a strong bias will result in rapid convergence. The rate of increase also depends on the variance in the population — in the most extreme case, in a population which is culturally homogeneous, directly-biased transmission has no impact.

2.3.3.2 *Linguistic case study 1: evolution of consistent head ordering*

Kirby (1999) (Chapter 2 in particular) uses a simple ILM to develop an explanatory account of word-order universals. Hawkins (e.g Hawkins (1990)) notes a statistical tendency for languages of the world to be consistently head-initial or head-final across phrasal categories. Hawkins suggests that this word-order universal can be accounted for in terms of a preference for language users, when parsing utterances, to construct trees as rapidly as possible. This principle is termed Early Immediate Constituent recognition (EIC), and languages which exhibit consistent head ordering score more highly on the EIC metric and are easier to parse. However, as noted by Kirby, this does not constitute an explanation of the observed statistical universal — how does the preference of individual language users for constructions which score well on the EIC metric translate to a word-order universal?

Kirby constructs a generational ILM to attempt to answer this question. The linguistic competence of individuals in Kirby's model consists of a specification of a simple grammar which states the preferred ordering of head and complement in verb phrases (verb-initial, VO, or verb-final, OV) and adpositional phrases (prepositional, PreP, or postpositional, PostP). There are therefore 4 possible grammars — VO and PreP, VO and PostP, OV and PreP, OV and PostP. The first and last of these possibilities have consistent head ordering and therefore score more highly on the EIC metric.

Individuals produce utterances consistent with their grammars, where an utterance is not an actual sentence but a specification of the head-ordering in verb and adpositional phrases. Utterances are essentially direct externalisations of an individual's I-language. The next generation of learners take a random sample of these utterances, and according to the procedures outlined below, select their own grammar.

Selection of a grammar is based on the frequency of utterances exhibiting particular word orders in the sample pool of utterances, but also on the parsability of those utterances. The probability of an individual selecting a variant v for their grammar is given by $p\left(v\right)$ and the number of utterances exhibiting that variant in the sample pool is given by n_v. w_v gives the degree to which variant v is preferred. A learner selects their grammar features probabilistically according to the formulae:

$$\begin{aligned} p\left(PreP\right) &= \frac{w_{PreP}n_{PreP}}{w_{PreP}n_{PreP} + w_{PostP}n_{PostP}} \\ p\left(PostP\right) &= \frac{w_{PostP}n_{PostP}}{w_{PreP}n_{PreP} + w_{PostP}n_{PostP}} \end{aligned}$$

$$
\begin{aligned}
p\left(VO\right) &= \frac{w_{VO}n_{VO}}{w_{VO}n_{VO}+w_{OV}n_{OV}} \\
p\left(OV\right) &= \frac{w_{OV}n_{OV}}{w_{VO}n_{VO}+w_{OV}n_{OV}}
\end{aligned}
$$

The degree to which a particular variant is preferred depends on the frequency of other variants in the utterance pool and the EIC metric. For example, if, as mentioned above, the EIC for VO-PreP is better than VO-PostP then $w_{PreP} > w_{PostP}$ if VO order is common, and $w_{PreP} < w_{PostP}$ if OV order is common. The preferences are calculated as follows:

$$
\begin{aligned}
w_{PreP} &= \alpha n_{VO} + \left(1-\alpha\right) n_{OV} \\
w_{PostP} &= \alpha n_{OV} + \left(1-\alpha\right) n_{VO} \\
w_{VO} &= \alpha n_{PreP} + \left(1-\alpha\right) n_{PostP} \\
w_{OV} &= \alpha n_{PostP} + \left(1-\alpha\right) n_{PreP}
\end{aligned}
$$

where α is a constant reflecting the EIC metric of the various combinations. Kirby reports results for $\alpha = 0.6$, which corresponds to the situation where consistent head ordering is somewhat more parsable than inconsistent head ordering.

Kirby conducts multiple runs of the ILM and finds that there are two stable final states, corresponding to the two consistent head-orderings, VO-PreP and OV-PostP, with the simulation runs converging at random on one of these two states. The learner's bias in favour of consistent head ordering results in the emergence of word-order universals. In terms of B&R's scheme, this is a clear example of how directly biased transmission can be applied in a linguistic context to generate clear and parsimonious accounts of linguistic universals — a theoretically-well motivated bias of learners to preferentially acquire a particular cultural variant leads to that cultural variant dominating the population. Kirby also notes an S-shaped trajectory of change, as predicted by B&R's model of directly biased transmission.

2.3.3.3 Linguistic case study 2: the evolution of vocabulary

Hutchins & Hazelhurst (1995) present an early model of the negotiation of vocabulary in a population. Their key concern is to show that conventionalised symbolic vocabulary can arise through cultural processes. Hutchins & Hazelhurst model individuals using autoassociator neural networks. Autoassociator networks consist of an input layer of nodes,

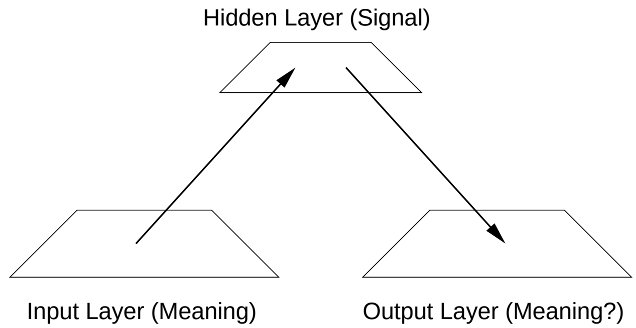

Figure 2.9: Hutchins & Hazelhurst's autoassociator network. The flow of activation (indicated by arrows) proceeds from the input layer via the hidden layer to the output layer. The goal of learning in an autoassociator network is to reproduce the input pattern of activation at the output layer. This necessitates forming a compressed representation of the input pattern at the hidden layer. Hutchins & Hazelhurst interpret the input pattern of activation (and the output pattern of activation) as a visual stimuli (meaning) and the pattern of activation at the hidden layer as an observable signal.

a smaller hidden layer and an output layer of the same size as the input layer. Autoassociator networks are trained (using the backpropagation learning algorithm in Hutchins & Hazelhurst's model) to map from an input pattern of activation, via an internal representation at the hidden layer, back to a pattern of output activation which exactly matches the input pattern — autoassociator networks essentially associate an input pattern of activation with itself.

In Hutchins & Hazelhurst's model there is a set of scenes which agents are required to communicate to one another about using a set of signals. Hutchins & Hazelhurst (1995) consider input/output patterns of activation to represent visual scenes, equivalent to meanings in the canonical E/I vocabulary model, while patterns of activation over the network's hidden layer are considered to represent observable signals. The network structure and interpretation of the various layers is illustrated in Figure 2.9.

At every time-step two individuals are selected at random from the population and one scene is chosen from the set of scenes, again at random. One individual acts as speaker and produces a signal given the selected scene. The second individual acts as a learner and is trained to associate the input scene with an identical output pattern of activation (i.e. perform the autoassociator learning task), while at the same time learning to associate the input scene with the signal produced by the speaker. The two individuals are then returned to the population. A simulation run consists of several thousand such pairwise interactions. Note that, as there is no turnover of population and therefore no separation between learners and producers, Hutchins & Hazelhurst's (1995) model is a classic implementation of the Negotiation Model. There is no explicit measurement of communicative accuracy in this model.

Hutchins & Hazelhurst report that, initially, the individuals in the population do not have distinct signals for each distinct scene and there is no consensus across the population as to which signals should be associated with which scenes. This is due to the random initialisation of the weights in each individual's network. However, after several thousand pairwise interactions, every individual in the population associates each distinct scene with a distinct signal and there is consensus between individuals as to which signal should be associated with which scene. The population converges on a communication system which would be optimal in terms of communicative payoff, as defined in Section 2.2.2.1.

What drives the emergence of these optimal communication systems? We can rule out natural selection of cultural variants, given the absence of any reward for successful communication. Guided variation can also be discounted, given the absence of any individual learning, as can indirect bias (to be discussed in Section 2.3.4). As will be discussed in Chapter 3, the autoassociator model of individuals used by Hutchins & Hazelhurst is strongly biased in favour of communication systems which are optimal in terms of communicative payoff — autoassociator agents are more likely to acquire such systems than systems which happen to be suboptimal in terms of communicative function. As predicted by B&R, the application of a direct bias on cultural transmission results in the increase in frequency of the cultural variant favoured by that bias. Hutchins & Hazelhurst observe the increase in frequency of scene-signal mappings which are optimal in terms of the biases of their chosen model of individual learners. These mappings also happen to be optimal in terms of communicative function.

The model outlined in Hutchins & Hazelhurst (1995) represents an interesting advance. The predictions of B&R are shown to hold with a far less abstract model of cultural variants and a bias which arises naturally from the chosen model of a learner. They also illustrate that natural selection is not the only possible pressure on cultural transmission that can lead to the emergence of communicatively optimal vocabulary systems — in their model there is no direct pressure for communicative function, nor even any evaluation of communicative accuracy, yet communicatively optimal systems still emerge. However, it is unclear how general their results are. Will any model of a learner result in the emergence of shared vocabulary when placed in the context of the Negotiation Model? What properties must the learner have to ensure the emergence of shared vocabulary? These questions will be returned to in depth in Chapter 3.

2.3.3.4 *Linguistic case study 3: the evolution of morphology*

Batali (1998) presents a computational implementation of the Negotiation Model where a small population of agents negotiate a vocabulary to communicate a set of (fairly minimally) structured meanings. The structure of the meanings and the biases of the learners results in the emergence of partially regular morphology.

In Batali's simulation individuals are modelled using simple recurrent neural networks capable of mapping a temporal sequence of input patterns to an output pattern of activation, with the eventual output pattern of activation depending on both the actual input patterns presented and the sequence in which input patterns are presented. Input patterns are representations of characters, with sequences of such input patterns being words. The developed output pattern of activation is considered to be a meaning. Batali's networks therefore take as input a sequence of characters forming a word and arrive at a meaning, their interpretation of that word. Meanings are analysed as consisting of a predicate component and a referent component. There are 10 possible predicates, represented by distinct but overlapping activation patterns over part of the output layer, and 10 possible referents, once again represented by distinct but overlapping activation patterns over the remainder of the output layer.

The simulation model follows the classic pattern of the Negotiation Model, with members of the population in turn producing meaning-signal pairs and learning from the produced meaning-signal pairs of other agents. Learning is carried out using the backpropagation algorithm. During each exposure the learner is presented with a meaning and a character sequence constituting the observed signal and attempts, using the backpropagation algorithm, to learn the association.

In addition to acting as learners, individuals act as producers (for communication and for producing observable behaviour for other agents to learn from) and as receivers (for the purpose of evaluating communicative accuracy). When acting as a receiver, the network is presented with a character sequence and constructs a meaning. Production is somewhat more complex, given that recurrent neural networks are non-reversible and Batali's networks map from input signals to output meanings. To produce a signal for a given meaning, an agent considers all possible characters and selects the character which produces the lowest output error in its own network with respect to the target meaning. That character is sent to the learner as the first character of the word associated with the target meaning. The producer continues producing characters until the error with respect to the target meaning drops below some threshold value or the number of characters in the word exceeds some arbitrary limit. This process of an individual using themselves as a model of the hearer and sending the signal which they themselves would interpret as the target meaning is known as the *obverter* strategy and is encountered elsewhere.

The accuracy of a communicative episode between two such individuals can be computed by comparing the speaker's given meaning vector with a hearer's output meaning vector after being exposed to the signal produced by the speaker. The communicative accuracy of a single episode is the number of positions on the meaning vector for which speaker and hearer agree on the value, within some tolerance. This corresponds to a variant of the canonical measure of communicative accuracy, with a distance metric over meanings and partial payoff for partially correct meanings. However, as with most Negotiation Models, there is no actual payoff for being a successful communicator. Cultural evolution is not driven by natural selection.

In the initial population of agents the accuracy of communicative episodes is at chance levels and distinct meanings are not necessarily communicated using distinct signals. As with the model of Hutchins & Hazelhurst (1995), this is due to the random starting values in each agent's initial network. Batali (1998) presents two results for this simulation model:

1. After several thousand learning interactions between members of the population, the population arrives at a near-optimal communication system. The accuracy of communicative episodes is high, each distinct meaning is communicated using a distinct signal and different individuals largely agree on the meaning associated with each signal.
2. The population converges on a semi-regular morphological system, with the predicate component of a meaning typically being associated with a root morpheme portion of a signal, and the referent component of a meaning typically being associated with a suffix component of the signal. For example, the pattern of activation corresponding to the predicate *happy* is usually, but not always, communicated using the root sequence "ba-", while the pattern of activation corresponding to the first person plural referent *we* is typically communicated with a suffix "-d" or some variant (e.g. "-dc" or "-ddc").

As with the model described in Hutchins & Hazelhurst (1995), the convergence of the population on a near-optimal communication system can be accounted for in terms of a direct bias on cultural transmission resulting from the learner model. This arbitrary bias of the learner results in the emergence of systems of meaning-signal mappings which happen to be optimal in terms of communicative function. An in-depth discussion of the bias is postponed till Chapter 3.

The emergence of semi-regular morphology is an interesting and novel aspect of the model. Batali also attributes this emergent structure to the biases of the agent architectures and the negotiation task. The negotiation task forces agents to arrive at shared mappings from signals to meanings. Sharing such mappings makes it likely that agents will share mappings from partially-presented signals (partial sequences of characters) to output patterns of activation, given that any large divergence in output patterns of activation midway through processing will be difficult to remedy later in the string. Agents will come to use fairly regular sequences to guide other agents (or themselves) into appropriate regions of output vector space. This is reflected in the semi-regularity of the final morphological system — there is a tendency to systematically use a particular sequence when expressing a particular portion of meaning, but this tendency is not so strong as to force a completely regular morphological system.

This model represents an interesting and significant development in the formal models of the cultural evolution of linguistic structure. Moderately sophisticated linguistic structure is investigated, using a methodology that has been applied to the cultural evolution of simpler, unstructured communication systems. The first result, that communicatively-useful communication systems can arise in the absence of natural selection, provides support for Hutchins & Hazelhurst's (1995) earlier conclusion. The use of a different model of a learner shows that neither Batali's nor Hutchins & Hazelhurst's results are dependent on a particular model of a learner. However, there is no exploration of the nature of the bias that leads, via negotiation, to the emergence of optimal communication and no attempt to relate the biases embodied in this model to those in earlier models. Furthermore, the explanation of the emergence of semi-regular morphological structure is perhaps somewhat underdeveloped, appealing to vague tendencies for agents to coordinate their movement through output vector space.

2.3.4 Indirectly-biased transmission

In directly-biased transmission one cultural variant is preferred over alternatives due to its own intrinsic properties. In indirectly biased transmission individuals must acquire two cultural traits — an *indicator* trait and an *indirectly biased* trait. Certain variants of the indicator trait are, as in the direct bias case, intrinsically preferable. No variant of the indirectly biased trait is intrinsically more preferable than any other. However, individuals prefer to acquire the indirectly biased trait of individuals who have an attractive indicator trait. For example, an indicator trait might be clothing styles (assuming an intrinsically-preferable style) and an indirectly biased trait might be political persuasion. Individuals will preferentially acquire the preferred clothing style (the indicator trait) and will preferentially acquire the political viewpoint of individuals who wear the preferred style of clothes — people will prefer to dress like smart dressers, and tend to vote like them too.

2.3.4.1 B&R's model

B&R assume that each individual is characterised by two cultural traits — an indicator trait and an indirectly-biased trait. They further assume that there is a direct bias in favour of a particular variant of the indicator trait — individuals preferentially acquire the preferred variant of the indicator trait. As discussed above, this will lead to the population converging on that variant of the indicator trait. It is furthermore assumed that individuals acquire the indirectly-biased trait of individuals who are have the preferred variant of the indicator trait — if an individual has the preferred indicator trait, they are disproportionately likely to transmit this, but also disproportionately likely to transmit their other cultural trait.

As shown in Section A.1.2.4, this results in the convergence of the population on values of the indirectly-biased trait which happen to be correlated with the preferred variant of the indicator trait — "variants of the indirectly biased trait that are positively correlated with the admired variants of the indicator trait will increase in frequency" (B&R p254). The rate of increase of the correlated trait depends on the strength of the correlation — indirectly-biased traits which are only weakly correlated with the preferred indicator trait will increase in frequency slowly, whereas indirectly-biased traits which are strongly correlated with the preferred indicator trait will rapidly come to dominate the population.

2.3.4.2 Linguistic case study: language evolution through acts of identity

Croft (2000) introduces an "utterance-based selectional theory" of linguistic evolution. In Croft's theory, differential replication of "linguemes" results in linguistic evolution, where a lingueme is a linguistic structure embodied in an utterance. Croft proposes that differential reproduction arises not through natural selection, where the functionality of a cultural variant determines access to roles which yield opportunities to transmit culturally, but through an evaluation of the social affordances of a particular lingueme by language users — biased transmission.

There is an established tradition in the sociolinguistic literature in accounting for the linguistic behaviour of individuals in terms of *prestige* (Labov 1966) and *covert prestige* (Trudgill 1972). In these terms, the choice of a particular linguistic form in preference to alternative forms constitutes an act of identity (LePage & Tabouret-Keller 1985) on the part of the speaker. A population consists of several, possibly overlapping, social groups — for example, a population can be subdivided according to gender, religion, social class or geographic region. Different social groups may use different linguistic systems. For example, residents of Japan, by and large, use Japanese whereas residents of the Korean Republic typically use Korean. Middle-class residents of Edinburgh typically have a different accent from working-class residents of Edinburgh. If an individual wishes to identify themselves with a particular social group they will adjust their linguistic behaviour to conform more closely with the linguistic behaviour associated with that group. In Croft's model it is such acts of identification that determine the selective advantage of one lingueme over another:

> "the factor in language use granting selective advantage to an individual speaker (and thus to the way she talks) is the desire of hearers who interact with her to identify with the community to which she belongs. This selective advantage ensures the differential perpetuation of the replicators she produces, that is, the propagation of the linguistic variants associated with the linguistic community to which she belongs" (Croft 2000:183).

To put it in the terminology Croft himself introduces, linguemes which are associated with social groups which individuals wish to identify themselves with will have a selective advantage. The social groups with which individuals wish to associate themselves are contingent on context and the notions of prestige and covert prestige. For example, in situations where covert prestige (say "working-classness") was important, linguemes associated with groups which rate highly on the covert prestige scale (e.g. linguemes associated with "working-classness") would have a selective advantage.

This informal model is highly reminiscent of the indirectly-biased transmission model of B&R, and Croft's conclusions are equivalent to the predictions of B&R's simple mathematical model — "variants of the indirectly biased trait [=linguemes] that are positively correlated with the admired variants of the indicator trait [= admired social groups] will increase in frequency [= have a selective advantage]". Croft's model is of course more complex, given that his notion of cultural traits is rather more complex and his equivalent of the biasing function on indicator traits, groups with which individuals wish to identify themselves, changes from occasion to occasion. It is also somewhat debatable as to whether an individual's social group can truly be treated as a cultural trait. An acceptable circumlocution might be to say that an individual's social group is determined by a complex of traits, many of which are culturally acquired. In this case, the comparison with B&R's model passes at least a cursory inspection.

There is a further complication, however. In B&R's model the correlation between an indicator trait and an indirectly biased trait can be presupposed, or can arise through cultural transmission of indicator and indirectly biased traits, assuming that errors during cultural transmission increase the correlation between indicator and indirectly biased traits. Croft has nothing to say on how this correlation between a particular social group and a particular set of linguistic behaviours gets off the ground in his model. For groups defined in terms of regional boundaries this might be unproblematic, given that spatial separation leads over time to linguistic separation. However, it is less clear how social groups defined in terms of religion or social class might come to be associated with particular linguistic behaviours. It is perhaps sufficient to simply presuppose this, or to assume it can be introduced by chance factors, or correlated errors between the cultural transmission of measures of prestige and linguemes.

2.3.5 Frequency-dependent bias

Cultural transmission is of course frequency dependent in all of the models considered so far — in the case of unbiased transmission, the probability of acquiring a particular cultural trait depends linearly on the frequency of that trait in the population, whereas in the direct and indirect bias cases, the relationship between the frequency of a trait and the probability of acquiring it is non-linear, with the non-linearity being introduced by the bias in favour of a particular trait. B&R reserve the term "frequency-dependent bias" for the case where the probability of acquiring the *most common* cultural variant in the population, regardless of which variant it is, is greater than (in the case of conformist frequency-dependent bias) or less than (in the case of non-conformist frequency-dependent bias) the probability of acquiring that variant in the unbiased transmission scenario.

2.3.5.1 B&R's model

B&R assume that individuals are disproportionately likely to acquire the majority cultural variant in the population (*conformist* frequency dependent bias). Assume that there are two possible cultural variants, c and d. As shown in Section A.1.2.5 of Appendix A, under conformist transmission, variant c will increase in frequency if the frequency of c is greater than the frequency of variant d. Conversely, c will decrease in frequency if its frequency is less than that of d. The rate of change of the most common variant is at its lowest as its frequency approaches 1 (where the population is converged on the most frequent variant) or 0.5 (the point where the population is perfectly split between the two variants). Conformist transmission results in the spread of the most common cultural variant.

2.3.5.2 *Linguistic case study: the spread of head order*

Briscoe (2000a) considers a model, similar in some aspects to that of Kirby (1999) described in Section 2.3.3.2, of the spread of a particular head ordering parameter within a population. Briscoe's (2000a) model is rather more complex than Kirby's model — learners acquire categorial grammars based on exposure to sets of triggers. However, the details of this model are not particularly relevant for the purpose of this section. It will suffice to say that learners acquire a grammar which is either head-initial (which we shall call G_I) or head-final (G_F). Triggers are specified as being of type I or F. Individuals using head-initial grammars (G_I) produces triggers of type I, and individuals using head-final grammars (G_F) produces triggers of type F. Learners observe a set of n triggers, and decide on whether to acquire G_I or G_F. Briscoe considers three possible learning procedures:

- The learner acquires G_I if the *first* of the n triggers it observes is of type I, and G_F if the first of the n triggers is of type F.
- The learner acquires G_I if the *last* of the n triggers is of type I, and G_F if the last of the n triggers is of type F.
- The learner acquires G_I if the *majority* of the n triggers are of type I, and G_F is the majority of the n triggers are of type F.

Learning procedures 1 and 2 are in fact equivalent and lead to a probability of acquiring grammar G_x of approximately $p\,(G_x)$, where $p\,(G_x)$ is the proportion of grammar G_x in the population. This is equivalent to the linear transmission rule for a dichotomous character as discussed in Section A.1.1.1 of Appendix A, where there is a single cultural parent. In the context of Briscoe's simulation model, where populations are finite, this leads to random fluctuation in the frequencies of G_I and G_F until one grammar reaches fixation.

Learning procedure 3 has rather different behaviour. Assuming that n is odd (learners observe an odd number of triggers), and that k is the first integer such that $j > n/2$ (i.e. k is the lowest value such that k represents more than half the number of models n), the probability that an agent will acquire grammar G_I is:

$$Prob\left(G_I\right) = \sum_{j=k}^{n} \binom{n}{j} p\left(G_I\right)^j \, p\left(G_F\right)^{n-j}$$

In other words, the probability of acquiring grammar G_I is equal to the probability of picking a set of n models from a population characterised by $p\left(G_I\right)$ and $p\left(G_F\right)$ such that individuals of type G_I are in the majority. The probability of acquiring G_F can be similarly defined.

These equations results in population dynamics equivalent to B&R's equation describing the dynamics of a population acquiring a dichotomous character under frequency-dependent bias, in the case of a strongly conformist bias function.

Briscoe's results confirm those of B&R — the frequency of the initially most frequent cultural variant increases, with the rate of change decreasing as the population approaches saturation. In Briscoe's simulations, where the population is finite, random sampling errors can move the population away from the case where both variants are equally frequent, which in B&R's infinite population model would be an unstable fixed point.

2.3.6 Transmission through a bottleneck

The final pressure acting on cultural transmission to be discussed in this chapter is a novel discovery, arising from computational models of the cultural transmission of language. The character of this bias suggests that it may be a specific bias acting on the transmission of infinite systems, such as language.

2.3.6.1 The transmission bottleneck in the Iterated Learning Model

Kirby (2002) presents a generational ILM which deals with the cultural evolution of compositionality and recursiveness, two of the design features of language discussed in Chapter 1. This model also reveals a new mechanism of cultural evolution, not covered in B&R's set of possible mechanisms.

In Kirby's model an individual's linguistic competence consists of a definite-clause grammar with attached semantic arguments. These definite-clause grammars consist of a set of rules, where the left hand side consists of a non-terminal category and the semantic label for that category and the right hand side consists of zero or more non-terminal categories, with semantic labels, and zero or more strings of characters, which correspond to phonetically-realised components of a signal. Semantic representations are predicate-argument structures, which may have hierarchical structure. Two example grammars might be:

Grammar 1:

S / sees′(mark′,loves′(lynne′,garry′)) $\rightarrow$ `markseeslynnelovesgarry`

Grammar 2

S / p(x,y) $\rightarrow$ *N/x V/p S/y*
S / p(x,y) $\rightarrow$ *N/x V/p N/y*
V / sees′ $\rightarrow$ `sees`
V / loves′ $\rightarrow$ `loves`
N / mark′ $\rightarrow$ `mark`
N / lynne′ $\rightarrow$ `lynne`
N / garry′ $\rightarrow$ `garry`

Atomic semantic elements are marked with primes, characters are represented in `typewriter font`, upper case italics represent non-terminal categories and lower case italics represent variables over semantic elements.

Both grammars would produce the string `markseeslynnelovesgarry` meaning sees′(mark′,loves′(lynne′,garry′)), but clearly do so in rather different ways — grammar 1 would do so in a holistic manner, whereas grammar 2 would do so in a compositional (each subpart of the meaning corresponds to a subpart of the signal) and recursive (the first *S* rule rewrites as a string of non-terminals including an *S*) manner.

Learners in Kirby's model are presented with a set of utterances, where utterances consist of meaning-signal pairs, and induce a grammar based on these utterances. Grammar induction consists of two main processes — rule incorporation and rule subsumption. In an incorporation event a learner is presented with a meaning-signal pair $\langle m, s\rangle$ and forms a rule: *S*/$m \rightarrow s$. This amounts to simply memorising an observed utterance.

Subsumption involves two main sub-processes, chunking and merging. During chunking, pairs of rules are examined in the search for meaningful chunks, which are then separated out into new syntactic categories. For example, if a grammar contains two incorporated rules:

S / loves′(lynne′,garry′) $\rightarrow$ `lynnelovesgarry`
S / loves′(lynne′,beppe′) $\rightarrow$ `lynnelovesbeppe`

chunking would identify that there is a single difference in semantics between the two rules (the second argument) and a single difference in signal (the string-final `garry` and `beppe`). The two sentence level rules are then replaced with a single rule

S / loves$'$(lynne$'$,x) $\rightarrow$ `lynneloves` N/x

and two new rules are added referring to the newly-introduced syntactic category:

N / garry$'$ $\rightarrow$ `garry`
N / beppe$'$ $\rightarrow$ `beppe`

Merging compares pairs of rules and attempts to reduce the number of distinct syntactic categories in the grammar. For example, suppose that a learner with the grammar just given makes a third observation of the meaning-signal pair $\langle$loves$'$(garry$'$, beppe$'$), `garrylovesbeppe`$\rangle$. This will lead, via incorporation and chunking, to the grammar:

S / loves$'$(y,x) $\rightarrow$ M/y `loves` N/x
N / garry$'$ $\rightarrow$ `garry`
N / beppe$'$ $\rightarrow$ `beppe`
M /garry$'$ $\rightarrow$ `garry`
M /lynne$'$ $\rightarrow$ `lynne`

The merging operation will notice that the first example of category N and the first example of category M are identical, and will therefore replace all mentions of M with N. After removal of redundant rules, this leads to the grammar

S / loves$'$(y,x) $\rightarrow$ N/y `loves` N/x
N / garry$'$ $\rightarrow$ `garry`
N / beppe$'$ $\rightarrow$ `beppe`
N /lynne$'$ $\rightarrow$ `lynne`

Merging therefore leads to generalisation — the learner with this grammar can now parse or produce utterances for nine meanings, based on three observations.

When called upon to produce an utterance for a meaning, Kirby's agents attempt to find a combination of rewrite rules which will cover the given meaning. If such a set of rules exist then the producer will produce an utterance consisting of the meaning and the string of terminals produced by the application of these rules. However, if such a set of rules does not exist then the producer will be forced to apply an invention procedure. Invention involves using existing rules as much as possible, with parts of the meaning which are not expressible using the grammar being expressed with random strings. During invention, the inventor learns from its own invention via incorporation, subsumption and so on.

Using these models of production and learning, Kirby conducts a series of generational ILM simulations, with a population size of 1 at each generation. At each generation individuals produce utterances for 50 randomly selected meanings involving no embedded predicates (e.g. loves′(lynne′,garry′)), then 50 meanings involving a single embedded predicate (e.g. sees′(mark′, loves′(lynne′,garry′))) then 50 meanings involving two embeddings (e.g. thinks′(beppe′, sees′(mark′, loves′(lynne′,garry′)))). These utterances then constitute the learning data for the learner at the next generation. Given that there are 5 embedding predicates, 5 non-embedding predicates and 5 possible atomic arguments each individual sees only a tiny fraction of the possible language of the previous generation — learners suffer from one aspect of the poverty of the stimulus problem. In Section 1.2.1, one aspect of the poverty of the stimulus problem was identified as the fact that children's data-exposure histories are finite, yet they acquire the ability to produce or understand an infinite number of sentences. Kirby's learners are exposed to a small number of utterances drawn from a system which is at least very large (covering all possible meanings involving two embeddings) and in principle infinite — embedding could be continued to an arbitrary depth. Kirby terms this aspect of the poverty of the stimulus the *transmission bottleneck*.

The language in the early stages of a simulation is almost entirely holistic, including rules such as:

S/ thinks′(beppe′, sees′(mark′, loves′(lynne′,garry′))) $\rightarrow$ `im`

Such grammars are obviously fairly large, as each distinct meaning is represented by a single rule. However, after 1000 generations of the ILM the grammars reduce to maximally compressed, fully compositional grammars with a single non-recursive sentence rule (for the non-embedded predicates) and a single recursive sentence rule (for the embedded predicates), and three further non-terminals (one for arguments, one for non-embedding predicates and one for embedding predicates). What drives this evolution of recursive compositionality?

One might be tempted to attribute the emergence of compositionality to directly biased transmission — learners compress their grammars wherever possible and are therefore biased in favour of acquiring cultural variants corresponding to compressed grammars. While this is almost certainly true to a degree, directly biased transmission is only part of the story. Were each learner exposed to utterances for the full set of meanings, the learner bias in favour of compressed grammars would lead to a partially compressed grammar, but not the radically compressed grammars that do emerge.

This radical compression is driven by the transmission bottleneck — utterances produced using holistic rules must be observed to be reproduced, whereas utterances produced using general rules can be reproduced even if they have not been seen, provided some other utterances produced using those same general rules have been observed. More general rules are more likely to be represented in the data presented to the learner, and are therefore more likely to survive cultural transmission. The maximally stable grammar is maximally general, and the bottleneck on transmission filters out non-stable grammars until the maximally stable, maximally general grammar is found.

This therefore represents a new mechanism for cultural change, one which is particularly relevant to the cultural transmission of language. Batali (2002) (discussed below) argues that such a dynamic is also present in his model.

This mechanism by which languages change from less to more generalisable due to the pressure introduced by the bottleneck might be seen as an independent principle of language change. Recall Lightfoot's assertion that the "search for independent principles of change must be abandoned . . . predictions are derived mostly from a theory of grammar" (Lightfoot 1979:153). This does not seem to be the case in Kirby's model — the initial, idiosyncratic, holistic systems in Kirby's simulations are as compatible with Kirby's theory of grammar (such as it is) as the final, systematic, compositional languages. Of course the representational restrictions of the grammar and the inductive biases of the learner, which could be seen as part of the theory of grammar, play a role in shaping language emergence in Kirby's simulations. But his results are not predictable solely from these factors — the bottleneck on transmission imposes a key pressure. As such, the change from holism to generalizability could be seen as an independent principle of change.

While Kirby's findings are obviously significant, one or two minor criticisms might be made. Firstly, the transmission bottleneck is enforced in all of his simulation runs. Some runs with no bottleneck on transmission would allow us to identify how much linguistic evolution would take place in the absence of the bottleneck, which would in turn allow us to identify the strength of the learning bias in his model and the strength of the dynamic arising from the bottleneck. Secondly, it is not clear to what extent the incremental increase in complexity of embedding during production (no embedding first, then degree-1 embedding, then degree-2 embedding) assists the emergence of recursive compositionality — to what extent would these results still hold if the "starting small" (Elman 1993) procedure was not applied. Finally, how important is the assumption that producers learn from their own invention? This may be a reasonably plausible assumption, but would we still see the evolution of recursive compositionality if we weakened this assumption and assumed that producers only occasionally learned from their inventions, or never did?

2.3.6.2 The transmission bottleneck in the Negotiation Model

Batali (2002) presents another computational implementation of the NM. Unlike in his earlier implementation (Batali 1998), Batali takes a symbolic, rather than connectionist, approach to modelling agents and the emergent communication systems in his model are striking in their intricacy.

In his symbolic model Batali treats meanings as formula sets, equivalent to predicate logic constructions. Formula sets consist of conjunctions of zero or more predicates, each of which take one or two arguments, where the arguments are variables represented numerically. For example, the feature set $\{(lizard\ 1)\,(likes\ 1\ 2))\}$ is equivalent to the logical formula $\exists x \exists y\,[lizard\,(x) \wedge likes\,(x, y)]$. Signals are simply strings of characters, of unbounded length.

Linguistic structures are mappings between formula sets and signals. The simplest type of linguistic structure simply has a formula set paired with a signal. Batali terms these basic structures "tokens", and they are equivalent to lexical items. More complex structures consist of combined sets of tokens. Tokens are combined in binary branching structures, with the formula set for the whole structure being the union of the formula sets for its constituent tokens and the signal for the whole structure being the concatenated signals associated with each token. Argument maps rename variables during the unification of the formula sets for two distinct tokens. Some example structures are given in Figure 2.10.

Agents learn by observing meaning-signal pairs produced by other members of the population, in the classic Negotiation Model framework. Individuals acquire linguistic knowledge by building up a set of exemplars, where each exemplar is a structure of the type shown in Figure 2.10 with an associated cost. When presented with a meaning-signal pair a learner has several choices.

- Store the observed meaning-signal pair in their memory as an unanalysed token.
- Search among existing exemplars and find an exemplar which can be modified so that it matches the observed meaning and signal. Exemplars can be modified by replacing one subpart of a complex exemplar with another exemplar, which may

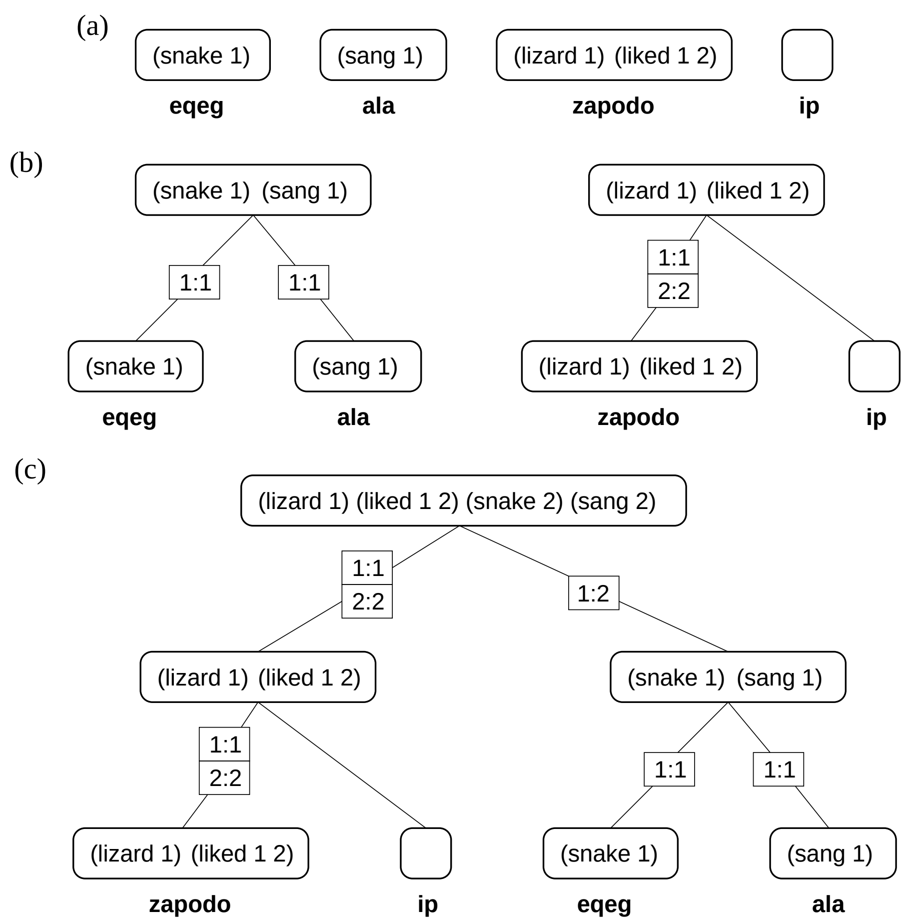

Figure 2.10: Exemplars. The exemplars in (a) are tokens, corresponding to lexical items. The formula set is enclosed in a box, while the string is given in sans serif font. The exemplars in (b) are constructed by combining the tokens from (a), with argument maps (in square boxes) specifying how to relabel variables from the tokens in the formula set for the complex exemplars. The exemplar in (c) is constructed by combining the exemplars in (b). Note that one argument map rewrites occurrences of 1 as 2.

have to be created from scratch. The modified exemplar and any newly created subparts are stored in memory.

- Combine two or more exemplars to form a new complex exemplar, which has the same meaning and signal as the observed utterance. This may involve creating new exemplars. The new, complex exemplar and any newly-created subparts are added to the store of exemplars.

Every newly-created exemplar has a initial cost of 1. A learner searches for the cheapest exemplar or combination of exemplars with the lowest total cost which matches the given meaning and signal. Exemplars which are used during learning have their cost reduced. During learning, agents also perform a search for inconsistent exemplars. If two exemplars or combinations of exemplars have matching signals but non-matching formula sets, they have their costs increased by a fixed amount. This amounts to a penalisation of homonymy. Finally, exemplars which have not been used in the last 200 episodes of production, reception or learning are removed from an agent's memory.

Agents are called upon to produce a signal for a given meaning, either when producing behaviour for another individual to observe or when taking part in a communicative interaction (as described below). Production involves searching through an agent's set of exemplars for an exemplar or combination of exemplars which will have as its formula set the meaning to be conveyed.

Agents have several options during production:

- Exemplars can be retrieved as a whole from the agent's set of exemplars. Such exemplars have an associated cost.
- Exemplars can be modified, by replacing one subpart with another exemplar. Such exemplars have a cost equal to the sum of the costs of the two exemplars used, plus a fixed modification cost.
- Exemplars can be combined, in which case the combined exemplar has a total cost equal to the sum of the costs of the component exemplars plus a fixed combination cost.

During any one of these processes, new unstructured token can be created. Such tokens consist of a formula set and a random string, and have a cost proportional to the sum of the length of the string and the number of predicates in the formula set. The producer searches for the cheapest combination of exemplars. Reception (the search for an exemplar which matches a given signal) proceeds in the same fashion. Any new exemplars which are created during production or reception are *not* stored in an individual's memory — unlike in Kirby's model, individuals do not learn from their own inventions.

90% of the interactions an individual participates in are of the learning type, with one individual called upon to produce a signal for a randomly-selected formula set containing between 2 and 7 predicates and 1 and 3 different variables, and the other individual learning from the produced behaviour. 10% of an agent's interactions are of a purely communicative nature, with one individual producing a signal and the other individual receiving the signal and arriving at a formula set. The success of such interactions are evaluated according to the formula:

$$\text{Communicative Accuracy} = \frac{1}{2}\left(\frac{c}{s} + \frac{c}{r}\right)$$

where s is the number of formulae in the sender's formula set, r is the number of formulae in the receiver's formula set and c is the number of formulae common to both sets. This is similar to Batali's (1998) communicative accuracy measure, again with partial payoff for partial communicative success. Communicative accuracy varies between 0 and 1, with a value of 1 representing perfect communication.

Batali reports several interesting results relating to the level of communicative accuracy within the population and the structure of the emergent communication systems:

1. The population's communicative accuracy increases from 0 to close to 1 over tens of thousands of rounds of negotiation. Individuals settle on fairly stable sets of exemplars, and typical exemplar cost is low. In the final stable system individuals rarely need to produce new structures during learning. The majority of tokens (unanalysed pairings between a signal and a formula set) contain a single formula in the formula set. The emergent systems are compositional — the meanings of complex exemplars depend of the meanings of their constituent exemplars, and the way those constituents are combined.
2. In some emergent systems empty tokens (unanalysed pairings of a string and an empty formula set) play an important role in the types of argument map applied to a complex exemplar. Some examples are given in Figure 2.11, which involve empty tokens which introduce a collapsing map and an inverting map.
3. Systems which do not use empty tokens rely on word order to construct argument maps. For example, in a system with no empty tokens, predicates relating solely

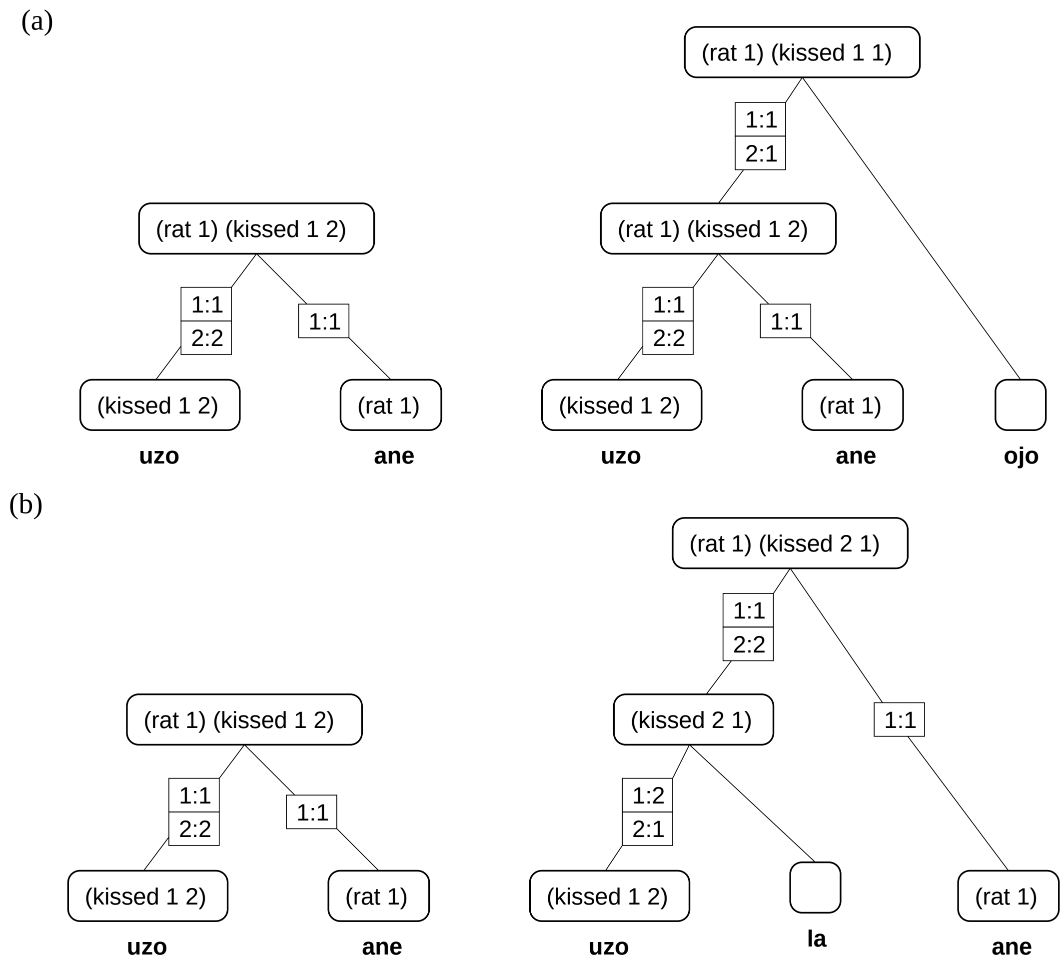

Figure 2.11: Argument maps in Batali's model. The larger complex exemplar in (a) is constructed by combining the smaller exemplar with the token with the string ojo. While this token contributes nothing to the formula set of the utterance as a whole, it does introduce another layer of structure. In the argument map introduced with this top layer of structure, the variable 2 is collapsed with variable 1. The ojo token could be said to function in this case as some kind of marker of a reflexive. (b) shows two more exemplars. The smaller of the two exemplars yields the formula set which could be glossed as meaning "rat kissed someone". In the second complex exemplar, the insertion of the semantically-empty token with the string la allows an intermediate layer of structure to be built. The associated argument map swaps the variables around. This exemplar, when combined with a further exemplar, yields the formula set which would be glossed as "someone kissed rat" — the la token could be considered a marker of a passive-like construction.

to individual x will appear first, followed by a two-place predicate involving x and another individual y, followed by all predicates relating solely to individual y.

What drives the emergence of coordinated communication and linguistic structure? In B&R's terms, the emergence of linguistic structure in Batali's model appears to be driven by a direct bias. Systems encodable using a small number of exemplars which are frequently reused and recombined are favoured by this bias. The exemplars encoding these systems will have a low cost and will therefore be unlikely to be replaced by random, less structured inventions, which have a high cost.

This bias will apply both to production — agents will prefer to reuse cheap exemplars, which will therefore become cheaper and more reusable — and to learning — analyses of novel meaning-signal pairs which involve existing exemplars will be preferred by learners, reinforcing the reusable components and forcing out analyses which involve large, non-reusable chunks. The division between systems which use empty tokens to control argument maps and those which use a rigid word order is an example of two alternative solutions to the problem of arriving at a particular formula set when combining exemplars. Presumably the state of the final system with respect to this choice is contingent on coincidences in the early rounds of the negotiation process.

We should also expect the transmission bottleneck to have an impact on the population's emerging language — learners acquire the ability to communicate approximately 10^{13} meanings after a few thousand exposures. However, the NM framework makes it difficult to tell how much cultural evolution is due to direct bias and how much is due to the transmission bottleneck. The sharp discontinuity between generations in the ILM highlights the importance the transmission bottleneck. In the NM, there is no notion of a generation, no sharp discontinuity and no clear indication of when the bottleneck applies. In fact, we would expect the impact of the bottleneck to be at its most severe at the start of an NM simulation — at this point learners have made few observations and are still called upon to produce utterances. The strength of the bottleneck decreases as observations are accumulated. Is the structure of the population's language determined early on, when the bottleneck is very tight, or later, after its severity is diminished?

Batali's model represents a significant development of the very simple models of cultural transmission proposed by B&R. Cultural variants are highly structured aggregations of smaller culturally-transmitted variants. However, as discussed above, the framework of the Negotiation Model makes the role of bias during cultural transmission difficult to separate from forces arising from cultural transmission itself.

2.4 Summary of the Chapter

In this Chapter I have introduced B&R's general model of cultural transmission, and the more language-specific Expression/Induction Model. The latter is typically implemented as either an Iterated Learning Model, where vertical transmission is paramount, or a Negotiation Model, where horizontal transmission is paramount. Treating language as a culturally transmitted trait requires us to question to what extent language is indeed culturally transmitted, what constitute the units of cultural replication, and what form the Primary Linguistic Data available to learners takes.

I then provided a summary of B&R's taxonomy of pressures acting 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. Each of these pressures has been implicated in the cultural evolution of some aspect of language, and I have given examples of models which demonstrate how these pressures can drive linguistic evolution. Finally, I described a further pressure, which does not appear in B&R's taxonomy, which has been hypothesised to play a role is the evolution of linguistic structure — the pressure to generalise arising from the bottleneck on cultural transmission.

These causes of cultural evolution, in particular the pressures arising from direct bias and a transmission bottleneck, will play a recurring role in the thesis. I will show that some of the fundamental structural properties of language are a consequence of the interaction of these pressures during the cultural transmission of language.