Iterated Learning: a framework for the emergence of language

Kenny Smith[1], Simon Kirby[1], Henry Brighton[1]

[1] Language Evolution and Computation Research Unit, Theoretical and Applied Linguistics, University of Edinburgh, Adam Ferguson Building, 40 George Square, Edinburgh, UK. {kenny,simon,henryb}@ling.ed.ac.uk

Language is culturally transmitted. Iterated Learning, the process by which the output of one individual's learning becomes the input to other individuals' learning, provides a framework for investigating the cultural evolution of linguistic structure. We present two models, based upon the Iterated Learning framework, which show that the poverty of the stimulus available to language learners leads to the emergence of linguistic structure. Compositionality is language's adaptation to stimulus poverty.

Linguists traditionally view language as the consequence of an innate "language instinct" (Pinker 1994). It has been suggested that this language instinct evolved, via natural selection, for some social function — perhaps to aid the communication of socially relevant information such as possession, beliefs and desires (Pinker & Bloom 1990), or to facilitate group cohesion (Dunbar 1996). However, the view of language as primarily a biological trait arises from the treatment of language learners as isolated individuals. We argue that language should be more properly treated as a culturally transmitted system. Pressures acting on language during its cultural transmission can explain much of linguistic structure. Aspects of language which appear baffling when viewed from the standpoint of individual acquisition emerge straightforwardly if we take the cultural context of language acquisition into account. While we are sympathetic to attempts to explain the biological evolution of the language faculty, we agree with Jackendoff that "[i]f some aspects of linguistic behaviour can be predicted from more general considerations of the dynamics of communication [or cultural transmission] in a community, rather than from the linguistic capacities of individual speakers, then they should be" (Jackendoff 2002:101).

We present the Iterated Learning Model (ILM) as a tool for investigating the cultural evolution of language. Iterated Learning is the process by which one individual's competence is acquired on the basis of observations of another individual's behaviour, which is determined by that individual's competence. This model of cultural transmission has proved particularly useful in studying the evolution of language. We present two models here. Both attempt to explain the emergence of compositionality, a fundamental structural property of language. In the first model, insights gained from the ILM suggest a mathematical analysis, which predicts when compositional language will be more stable than non-compositional language. In the second model, techniques adopted from artificial life are used to investigate the transition, through purely cultural processes, from non-compositional to compositional language. These models reveal two key determinants of linguistic structure. Firstly, the poverty of the stimulus available to language learners during cultural transmission drives the evolution of structured language — without this stimulus poverty, compositional language will not emerge. Secondly, compositional language is most likely to evolve when linguistic agents perceive the world as structured — structured pre-linguistic representation facilitates the cultural evolution of structured language.

Two views of language

In the dominant paradigm in linguistics (formulated and developed by Noam Chomsky, for example Chomsky (1965) and Chomsky (1995)), language is viewed as an aspect of individual psychology. The object of interest is the internal linguistic competence of the individual, and how this linguistic competence is derived from the data the individual is exposed to. External linguistic behaviour is considered to be epiphenomenal, the uninteresting consequence of the application of this linguistic competence to a set of contingent communicative situations. This framework is sketched in Figure 1 (a). From this standpoint, much of the structure of language is puzzling — how do children, apparently effortlessly and with virtually universal success, arrive at a sophisticated knowledge of language from exposure to sparse and noisy data? In order to explain language acquisition in the face of this poverty of the linguistic stimulus, the Chomskyan program postulates a sophisticated, genetically-encoded language organ of the mind, consisting of a Universal Grammar, which delimits the space of possible languages, and a Language Acquisition Device, which guides the formation of linguistic competence based on the observed data.

Following ideas developed in Hurford (1990), we view language as an essentially cultural phenomenon. An individual's linguistic competence is derived from data which is itself a consequence of the linguistic competence of another individual. This framework is sketched in Fig. 1 (b). Under this view, the burden of explanation is lifted from the postulated innate language organ — much of the structure of language can be explained as a result of pressures acting on language during the repeated expression and induction of linguistic forms. In this paper we will show how the poverty of the stimulus available to language learners is the cause of linguistic structure, rather than a problem for it.

The Iterated Learning Model

The Iterated Learning Model (ILM), as introduced in Kirby (2001) and Brighton (2002), provides a framework for studying the cultural evolution of language. The ILM in its simplest form is illustrated in Fig. 2. In this model H_i corresponds to the linguistic competence of individual i, whereas U_i corresponds to the linguistic behaviour of individual i and the primary linguistic data for individual $i+1$.

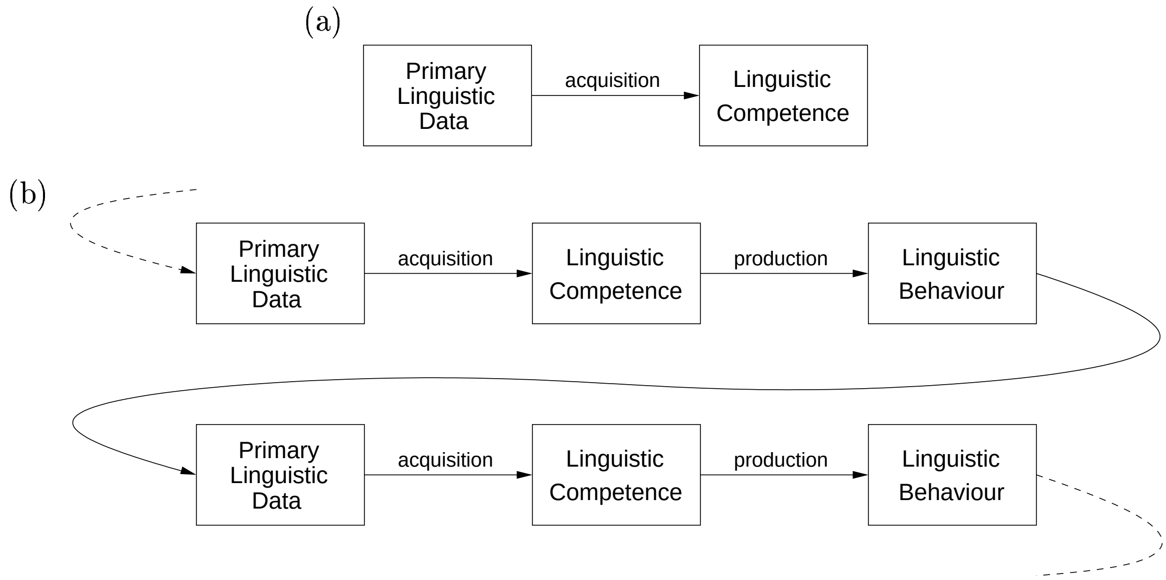

Figure 1: (a) The Chomskyan paradigm. Acquisition procedures, constrained by Universal Grammar and the Language Acquisition Device, derive linguistic competence from linguistic data. Linguistic behaviour is considered to be epiphenomenal and has no place. (b) Language as a cultural phenomenon. As in the Chomskyan paradigm, acquisition based on linguistic data leads to linguistic competence. However, we now close the loop — competence leads to behaviour, which contributes to the linguistic data for the next generation.

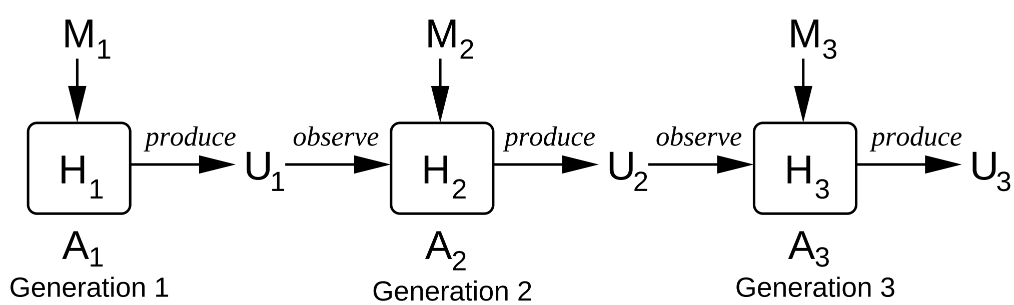

Figure 2: The ILM. The ith generation of the population consists of a single agent A_i who has hypothesis H_i. Agent A_i is prompted with a set of meanings M_i. For each of these meanings the agent produces an utterance using H_i. This yields a set of utterances U_i. Agent A_{i+1} observes U_i and forms a hypothesis H_{i+1} to explain the set of observed utterances, and the cycle repeats.

We make the simplifying idealisation that cultural transmission is purely vertical — there is no horizontal, intra-generational cultural transmission. This simplification has several consequences. Firstly, we can treat the population at any given generation as consisting of a single individual. Secondly, we can ignore the intra-generational communicative function of language. However, the Iterated Learning framework does not rule out either intra-generational cultural transmission (see Livingstone & Fyfe (1999) for an ILM with both vertical and horizontal transmission, or Batali (2002) for an ILM where transmission is purely horizontal) or a focus on communicative function (see Smith (2002b) for an ILM focusing on the evolution of optimal communication within a population).

In most implementations of the ILM, utterances are treated as meaning-signal pairs. This is obviously an oversimplification of the task facing language learners — if the meaning of every signal were self-evident then the signal itself would be rather pointless. However, empirical evidence suggests that language learners have a variety of strategies for deducing the communicative intentions of others during language acquisition (see Bloom (2000) for review). We will assume for the moment that these strategies are error-free, while noting that the consequences of weakening this assumption is a current and interesting area of research (see, for example, Steels (1998), Smith (2001) and Steels *et al.* (2002)).

This simple model proves to be a powerful tool for investigating the cultural evolution of language. While we have previously used the ILM to look at the emergence of word-order universals (Kirby 1999), the regularity-irregularity distinction (Kirby 2001), and recursive syntax (Kirby 2002), here we will focus on the evolution of compositionality. The evolution of compositionality provides a test-case to evaluate the suitability of techniques from mathematics and artificial life in general, and the ILM in particular, to tackling problems from linguistics.

The cultural evolution of compositionality

In a compositional system the meaning of a signal is a function of the meaning of its parts and they way they are put together (Krifka 2001). The morphosyntax of language exhibits a high degree of compositionality. For example, the relationship between the string *John walked* and its meaning is not completely arbitrary. It is made up of two components: a noun (*John*) and a verb (*walked*). The verb is also made up of two components: a stem and a past-tense ending. The meaning of *John walked* is thus a function of the meaning of its parts. Compositionality, in combination with recursive syntax, allows language users to produce and comprehend an infinite range of sentences.

Compositional language can be contrasted with non-compositional, or *holistic* communication, where a signal stands for the meaning as a whole, with no subpart of the signal conveying any part of the meaning in and of itself. Animal communication is typically viewed as holistic — no subpart of an alarm call or a mating display stands for part of the meaning "there's a predator about" or "come and mate with me".

We view language as a mapping between meanings and signals. A compositional language is a mapping which preserves neighbourhood relationships — neighbouring meanings will share structure, and that shared structure in meaning space will map to shared structure in the signal space. A holistic language is one which does not preserve such relationships — as the structure of signals does not reflect the structure of the underlying meaning, shared structure in meaning space will not necessarily result in shared signal structure.

In order to model such systems we need representations of meanings and signals. For both models outlined in this paper meanings are represented as points in an F-dimensional space where each dimension has V discrete values, and signals are represented as strings of characters of length 1 to l_{max}, where the characters are drawn from some alphabet Σ. More formally, the meaning space $\mathcal{M}$ and signal-space $\mathcal{S}$ are given by:

$$\mathcal{M} = \{(f_1\ f_2 \ldots f_F) : 1 \le f_i \le V \text{ and } 1 \le i \le F\}$$

$$\mathcal{S} = \{w_1 w_2 \ldots w_l : w_i \in \Sigma \text{ and } 1 \le l \le l_{max}\}$$

The world, which provides communicatively relevant situations for agents in our models, consists of a set of N objects, where each object is labelled with a meaning drawn from $\mathcal{M}$. We will refer to such a set of labelled objects as an *environment*.

A mathematical model

We will begin by considering, using a mathematical model[1], how the compositionality of a language relates to its stability over cultural time. For the sake of simplicity, we will restrict ourselves to looking at the two extremes on the scale of compositionality, comparing the stability of perfectly compositional language and completely holistic language.

Learning holistic and compositional languages

We can construct a holistic language L_h by simply assigning a random signal to each meaning. More formally, each $m \in \mathcal{M}$ is assigned a signal of random length l ($1 \le l \le l_{max}$) where each character is selected at random from Σ. The meaning-signal mapping encoded in this assignment of meanings to signals will not preserve neighbourhood relations, unless by chance.

Consider the task facing a learner attempting to learn L_h. There is no structure underlying the assignment of signals to meanings. The best strategy here is simply to memorise meaning-signal associations. We can calculate the expected number of meaning-signal pairs our learner will observe and memorise. After R observations of randomly-selected objects paired with signals a learner will have a set of O observed meanings. We can calculate the probability that any arbitrary meaning $m \in \mathcal{M}$ will be included in O, $Pr(m \in O)$, with:

[1] This model is described in greater detail in Brighton (2002).

$$Pr\left(m \in O\right) = \sum_{x=1}^{N} \begin{pmatrix} \text{Probability that} \\ m \text{ is used to label} \\ x \text{ objects} \end{pmatrix} \times \begin{pmatrix} \text{Probability of observing an} \\ \text{utterance being produced} \\ \text{for at least one of those } x \\ \text{objects after } R \text{ observations} \end{pmatrix}$$

When called upon to produce utterances, such a learner will only be able to reproduce meaning-signal pairs they themselves observed. Given the lack of structure in the meaning-signal mapping, there is no way to predict the appropriate signal for a meaning unless that meaning-signal pair has been observed. We can therefore calculate E_h, the expected number of meanings an individual will be able to express after observing some subset of a holistic language, which is simply the probability of observing any particular meaning multiplied by the number of possible meanings:

$$E_h = Pr\left(m \in O\right) \cdot V^F$$

We can perform similar calculations for a leaner attempting to acquire a perfectly compositional language. As discussed above, a perfectly compositional language preserves neighbourhood relations in the meaning-signal mapping. We can construct such a language L_c for a given set of meanings $\mathcal{M}$ using a lookup table of subsignals, where each subsignal is associated with a particular feature value. For each $m \in \mathcal{M}$ a signal is constructed by concatenating the appropriate subsignal for each feature value in m.

How can a learner best acquire such a language? The optimal strategy is to memorise feature value-signal substring pairs. After observing R randomly selected objects paired with signals, our learner will have acquired a set of observations of feature values for the ith feature, O_{f_i}. The probability that an arbitrary feature value v in included in O_{f_i} is given by $Pr\left(v \in O_{f_i}\right)$:

$$Pr\left(v \in O_{f_i}\right) = \sum_{x=1}^{N} \begin{pmatrix} \text{Probability that } v \\ \text{is used to label } x \\ \text{objects} \end{pmatrix} \times \begin{pmatrix} \text{Probability of observing an} \\ \text{utterance being produced} \\ \text{for at least one of those } x \\ \text{objects after } R \text{ observations} \end{pmatrix}$$

We will assume the strongest possible generalisation capacity. Our learner will be able to express a meaning if it has viewed all the feature values that make up that meaning, paired with signal substrings. The probability of our learner being able to express an arbitrary meaning made up of F feature values is then given by the combined probability of having observed each of those feature values:

$$Pr\left(v_1 \in O_{f_1} \wedge \ldots \wedge v_F \in O_{f_F}\right) = Pr\left(v \in O_{f_i}\right)^F$$

We can now calculate E_c, the number of meanings our learner will be able to express after viewing some subset of a compositional language, which is simply the probability of being able to express an arbitrary meaning multiplied by N_{used}, the number of meanings used when labelling the N objects:

$$E_c = Pr\,(v \in O_{f_i})^F \cdot N_{used}$$

We therefore have a method for calculating the expected expressivity of a learner presented with L_h or L_c. This in itself is not terribly useful. However, within the Iterated Learning framework we can relate expressivity to *stability*. If an individual is called upon to express a meaning they have not observed being expressed, they have two options. Firstly, they could simply not express. Alternatively, they could produce some random signal. In either case, any association that was present in the previous individual's hypothesis will now be lost. A shortfall in expressivity therefore results in instability over cultural time. We can relate the expressivity of a language to the stability of that language over time by $S_h \propto E_h/N$ and $S_c \propto E_c/N$. Stability is simply the proportion of meaning-signal mappings encoded in an individual's hypothesis which are also encoded in the hypotheses of subsequent individuals.

We will be concerned with the *relative stability* of compositional languages with respect to holistic languages, S, which is given by:

$$S = \frac{S_c}{S_c + S_h}$$

When $S = 0.5$ compositional languages and holistic languages are equally stable and we therefore expect them to emerge with equal frequency over cultural time. When $S > 0.5$ compositional languages are more stable than holistic languages, and we expect them to emerge more frequently, and persist for longer, than holistic languages. $S < 0.5$ corresponds to the situation where holistic languages are more stable than compositional languages.

The impact of meaning space structure and the bottleneck

S depends on the number of dimensions in the meaning space (F), the number of possible values for each feature (V), the number of objects in the environment (N) and the number of observations each learner makes (R). Unless R is very large, or N is very small, there is a chance that a an agent will be called upon to express a meaning they themselves have never observed paired with a signal. This is one aspect of the poverty of the stimuli facing language learners — the set of utterances of any human language is arbitrarily large, but a child must acquire their linguistic competence based on a finite number of sentences. We will refer to this aspect of the poverty of stimulus as the *transmission bottleneck*. The severity of the transmission bottleneck depends on R and N. It is convenient to refer instead to the degree of object coverage (b), which is simply the proportion of all N objects observed after R observations.

Together F and V specify the degree of structure in the meaning space. We will vary this structure, together with the transmission bottleneck b, while holding N constant. The results of these manipulations are shown in Fig. 3.

There are two key results to draw from these figures:

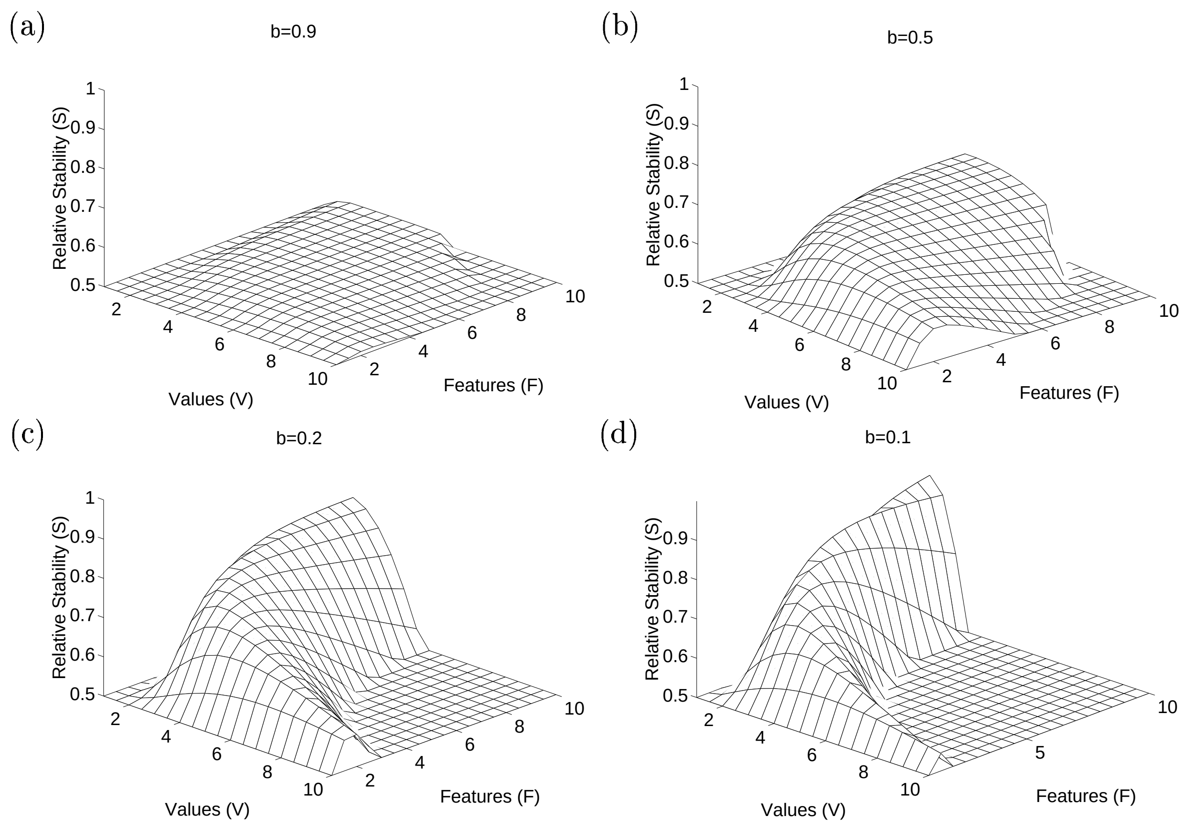

Figure 3: Severity of bottleneck and degree of meaning space structure impact on the relative stability of compositional language. The relative stability advantage of compositional language increases as the bottleneck tightens, but only when the meaning space exhibits a certain degree of structure.

1. S is at a maximum for small bottleneck sizes. Holistic languages will not persist over time when the bottleneck on cultural transmission is tight. In contrast, compositional languages are generalisable, due to their structure, and remain relatively stable even when a learner only observes a small subset of the language of the previous generation. The poverty of the stimulus "problem" is in fact required for linguistic structure to emerge.

2. High S only occurs when the meaning space exhibits a certain degree of structure, suggesting that structure in the conceptual space of language learners is a requirement for the evolution of compositionality. However, if the meaning space is too highly structured S is low, as few distinct meanings will share feature values and the advantage of generalisation is lost.

A computational model

The mathematical model outlined above, made possible by insights gained from viewing language as a culturally-transmitted system, predicts that compositional language will be more stable than holistic language when 1) there is a bottleneck on cultural transmission and 2) linguistic agents have structured representations of objects. However, the simplifications necessary to the mathematical analysis preclude a more detailed study of the dynamics arising from Iterated Learning. What happens to languages of intermediate compositionality during cultural transmission? Can compositional language emerge from initially holistic language, through a process of cultural evolution? We can investigate these question using techniques from artificial life, by developing a multi-agent computational implementation of the ILM.

A neural network model of a linguistic agent

Smith (2002b) presents a neural-network model of the evolution of holistic communication. We extend this model to allow the study of the cultural evolution of compositionality[2].

Agents are modelled using networks consisting of two sets of nodes $\mathcal{N}_M$ and $\mathcal{N}_S$ and a set of bidirectional connections $\mathcal{W}$ connecting every node in $\mathcal{N}_M$ with every node in $\mathcal{N}_S$. Nodes in $\mathcal{N}_M$ represent meanings and partially-specified components of meanings, while nodes in $\mathcal{N}_S$ represent partial and complete specifications of signals.

As with the mathematical model, meanings are sets of feature values, and signals are strings of characters. Components of a meaning specify one or more feature values of that meaning, with unspecified values being marked as a wildcard $*$. For example, the meaning $(2\,1)$ has three possible components, the fully-specified $(2\,1)$ and the partially specified $(2\,*)$ and $(*\,1)$. These components can be grouped together into ordered sets, which constitute an analysis of a meaning. For example, there are three possible analyses of the meaning $(2\,1)$ — the one-component analysis $\{(2\,1)\}$, and two two component analyses which differ in order, $\{(2\,*)\,,(*\,1)\}$ and $\{(*\,1)\,,(2\,*)\}$. Similarly, components of

[2] We refer the reader to Smith (2002a) for a more thorough description of this model.

signals can be grouped together to form an analysis of a signal. This representational scheme allows the networks to exploit the structure of meanings and signals. However, they are not forced to do so.

Learners observe meaning-signal pairs. During a single learning episode a learner will store a $\langle m, s \rangle$ pair in its network. The nodes in $\mathcal{N}_M$ corresponding to all possible components of the meaning m have their activations set to 1, while all other nodes in $\mathcal{N}_M$ have their activations set to 0. Similarly, the nodes in $\mathcal{N}_S$ corresponding to the possible components of s have their activations set to 1. Connection weights in $\mathcal{W}$ are then adjusted according to the rule:

$$\Delta W_{xy} = \begin{cases} +1 & \text{iff } a_x = a_y = 1 \\ -1 & \text{iff } a_x \neq a_y \\ 0 & \text{otherwise} \end{cases}$$

where W_{xy} gives the weight of the connection between nodes x and y and a_x gives the activation of node x. The learning procedure is illustrated in Fig. 4 (a).

In order to produce an utterance, agents are prompted with a meaning m and required to produce a signal s. All possible analyses of m are considered in turn with all possible analyses of every $s \in \mathcal{S}$. Each meaning analysis-signal analysis pair is evaluated according to:

$$g\left(\langle m, s \rangle\right) = \sum_{i=1}^{C} \omega\left(c_{mi}\right) \cdot W_{c_{mi}c_{si}}$$

where the sum is over the C components of the analysis, c_{mi} is the ith component of m and and $\omega\left(x\right)$ is a weighting function which gives the non-wildcard proportion of x. This process is illustrated in Figure 4 (b). The meaning analysis-signal analysis pair with the highest g is returned as the network's utterance.

Environment structure

In the mathematical model outlined above, the environment consisted of a set of objects labelled with meanings drawn at random from the space of possible meanings. In the computational model we can relax this assumption, and investigate how non-random assignment of meanings to objects impacts on linguistic evolution. As before, an environment consists of a set of objects labelled with meanings drawn from $\mathcal{M}$. The number of objects in the environment gives the *density* of that environment — environments with few objects will be termed low-density, whereas environments with a large number of objects will be termed high-density. When meanings are assigned to objects at random we will say the environment is *unstructured.* When meanings are assigned to objects in such a way as to minimise the average inter-meaning hamming distance we will say the environment is *structured.* Sample low- and high-density environments are shown in Fig. 5. Note the new usage of the term "structured" — while in the mathematical model we were concerned with structure in the meaning space, given by F and V, we are now concerned with the degree of structure in the environment. Different levels of environment structure are possible within a meaning space of a particular structure.

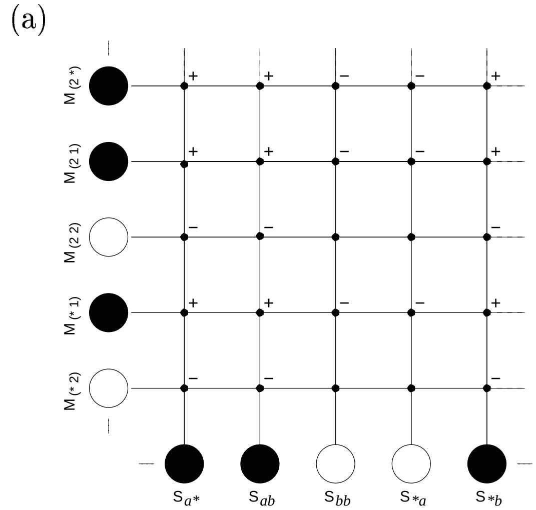

Figure 4: Nodes with an activation of 1 are represented by large filled circles. Small filled circles represent weighted connections. (a) Storage of the meaning-signal pair $\langle (2\ 1), ab \rangle$. Nodes representing components of $(2\ 1)$ and ab have their activations set to 1. Connection weights are then either incremented $(+)$, decremented $(-)$ or left unchanged. (b) Retrieval of three possible analyses of $\langle (2\ 1), ab \rangle$. The relevant connection weights are highlighted in grey. g for the one-component analysis $\langle \{(2\ 1)\}, \{ab\} \rangle$ depends of the weight of connection i. g for the two-component analysis $\langle \{(2\ *), (*\ 1)\}, \{a*, *b\} \rangle$ depends on the weighted sum of two connections, marked as ii. The g for the alternative two-component analysis $\langle \{(2\ *), (*\ 1)\}, \{*b, a*\} \rangle$ is given by the weighted sum of the two connections marked iii.

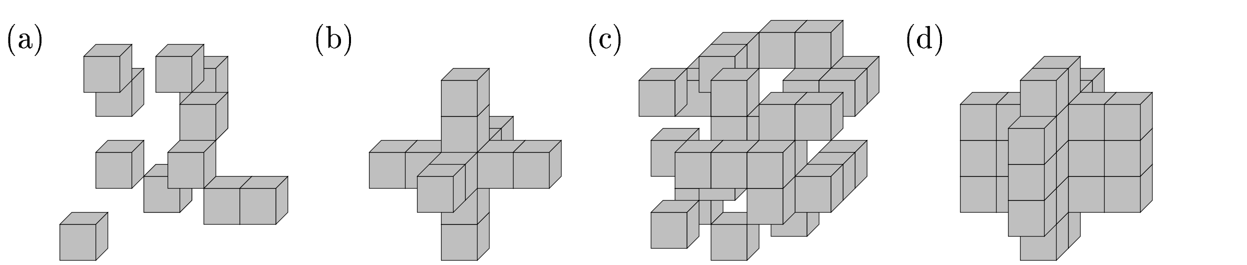

Figure 5: We will present results for the case where $F = 3$ and $V = 5$. This defines a three-dimensional meaning space. We highlight the meanings selected from that space with grey. (a) is a low-density, unstructured environment. (b) is a low-density, structured environment. (c) and (d) are unstructured and structured high-density environments.

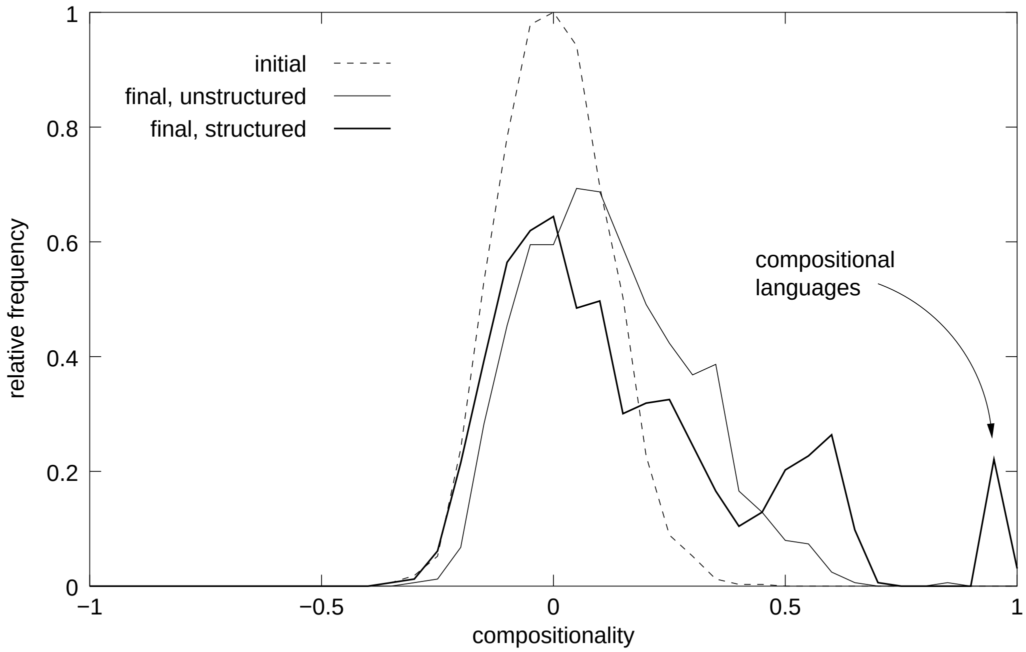

Figure 6: The relative frequency of initial and final systems of varying degrees of compositionality, when there is no bottleneck on cultural transmission. The results shown here are for the low-density environments given in Fig. 5. The initial languages are generally holistic. Some final languages exhibit increased levels of compositionality. Highly compositional languages are infrequent.

The impact of environment structure and the bottleneck

The network model of a language learner/producer is plugged into the ILM framework. We will manipulate three factors — the presence or absence of a bottleneck, the density of the environment and the degree of structure in the environment.

Fig. 6 plots the frequency by compositionality of initial and final systems in 1000 runs of the ILM, in the case where there is no bottleneck on cultural transmission. The initial agent has the maximum-entropy hypothesis — all meaning-signal pairs are equally probable. The learner at each generation is exposed to the complete language of the previous generation — the adult is required to produce utterances for every object in the environment. Each run was allowed to proceed to a stable state. Our measure of compositionality is simply the degree of correlation between the distance between pairs of meanings and the distance between the corresponding pairs of signals. Perfectly compositional languages have a compositionality of 1, whereas holistic languages have a compositionality of approximately 0.

Two main results are apparent from Fig. 6.

1. The majority of the final, stable systems are holistic.

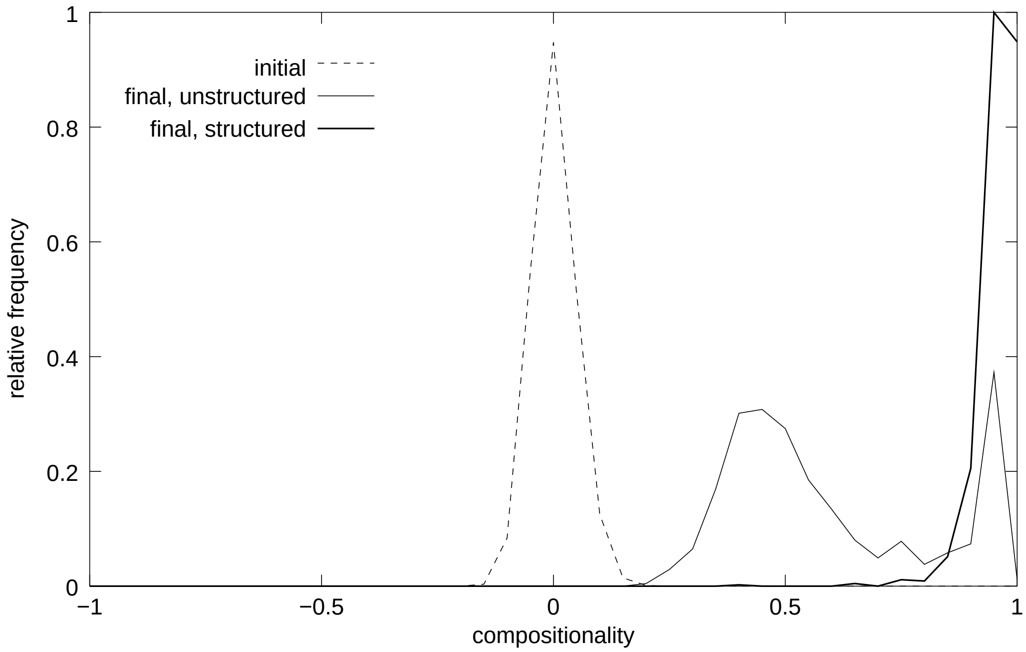

Figure 7: Frequency by compositionality when there is a bottleneck on cultural transmission. The results shown here are for the high-density environments given in Fig. 5. The initial languages are holistic. The final languages are compositional, with highly compositional languages occurring frequently.

2. Highly compositional systems occur infrequently, and only when the environment is structured.

In the absence of a bottleneck on cultural transmission, the compositionality of the final systems is sensitive to initial conditions. The initial system may exhibit, purely by chance, a slight tendency to express a given feature value with a certain substring. This compositional tendency can spread, over iterated learning events, to other parts of the system, which can in turn have further knock-on consequences. The potential for spread of compositional tendencies is greatest in structured environments — in such environments, distinct meanings are more likely to share feature values than in unstructured environments. However, this spread of compositionality is unlikely to lead to a perfectly compositional language.

Fig. 7 plots the frequency by compositionality of initial and final systems in 1000 runs of the ILM, in the case where there is a bottleneck on cultural transmission ($b = 0.4$). Learners will therefore only see a subset of the language of the previous generation. Whereas in the no-bottleneck condition each run proceeded to a stable state, in the bottleneck condition runs were stopped after 50 generations. There is no such thing as a truly stable state when there is a bottleneck on cultural transmission. For example, if all R utterances an individual observes refer to the same object then any structure in the language of the previous generation will be lost. However, the final states here were as close as possible to stable. Allowing the runs to continue for several hundred more generations results in a very similar distribution of languages.

Two main results are apparent from Fig. 7.

1. When there is a bottleneck on cultural transmission highly compositional systems are frequent.

2. Highly compositional systems are more frequent when the environment is structured.

As discussed with reference to the mathematical model, only highly compositional systems are stable through a bottleneck. The results from the computational model bear this out — over time, language adapts to the pressure to be generalisable, until the language becomes highly compositional, highly generalisable and highly stable. Highly compositional languages evolve most frequently when the environment is structured, because in a structured environment the advantage of compositionality is at a maximum — each meaning shares feature values with several other meanings, and a language mapping these feature values to a signal substring is highly generalisable.

Conclusions

Language can be viewed as a consequence of an innate language organ. If we take this view, we can form an evolutionary account which explains linguistic structure as a biological adaptation to social function — language is socially useful, and the language organ yields a fitness payoff. However, we have presented an alternative approach. We focus on the the cultural transmission of language. We can then form an account which explains much of linguistic structure as a cultural adaptation, by language, to pressures arising during repeated production and acquisition of language.

We have presented the ILM as a framework for studying the cultural evolution of language. We have focussed here on the cultural evolution of compositionality. Compositional language emerges when there is a bottleneck on cultural transmission — compositionality is an adaptation by language which allows it to slip through the transmission bottleneck. The advantage of compositionality is at a maximum when language learners perceive the world as structured — if objects are perceived as structured entities and the objects in the environment relate to one another in structured ways then a generalisable, compositional language is highly adaptive.

Of course, biological evolution still has a role to play in explaining the evolution of language. The ILM is ideal for investigating the cultural evolution of language on a fixed biological substrate, and identifying the cultural consequences of a particular innate endowment. The origins of that endowment then need to be explained, and natural selection for a socially-useful language might play some role here. We might indeed then find, as suggested by Deacon, that "the brain has co-evolved with respect to language, but languages have done most of the adapting" (Deacon 1997:122). The poverty of the stimulus faced by language learners forces language to adapt to be learnable. The transmission bottleneck forces language to be generalisable, and compositional structure is language's adaptation to this problem.

References

Batali, J. 2002. The negotiation and acquisition of recursive grammars as a result of competition among exemplars. In *Linguistic Evolution through Language Acquisition: Formal and Computational Models*, ed. by E. Briscoe. Cambridge: Cambridge University Press.

Bloom, P. 2000. *How Children Learn the Meanings of Words*. Cambridge, MA: MIT Press.

Brighton, H. 2002. Compositional syntax from cultural transmission. *Artifical Life* 8.25–54.

Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

—— 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.

Deacon, T. 1997. *The Symbolic Species*. London: Penguin.

Dunbar, R. 1996. *Grooming, Gossip and the Evolution of Language*. London: Faber and Faber.

Hurford, J. R. 1990. Nativist and functional explanations in language acquisition. In *Logical Issues in Language Acquisition*, ed. by I. M. Roca, 85–136. Dordrecht: Foris.

Jackendoff, R. 2002. *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.

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

—— 2001. Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity. *IEEE Journal of Evolutionary Computation* 5.102–110.

—— 2002. Learning, bottlenecks and the evolution of recursive syntax. In *Linguistic Evolution through Language Acquisition: Formal and Computational Models*, ed. by E. Briscoe, 173–203. Cambridge: Cambridge University Press.

Krifka, M. 2001. Compositionality. In *The MIT Encyclopaedia of the Cognitive Sciences*, ed. by R. A. Wilson & F. Keil. Cambridge, MA: MIT Press.

Livingstone, D., & C. Fyfe. 1999. Modelling the evolution of linguistic diversity. In *Advances in Artificial Life: Proceedings of the* 5^{th} *European Conference on Artificial Life*, ed. by D. Floreano, J. Nicoud, & F. Mondada, 704–708. Berlin: Springer.

Pinker, S. 1994. *The Language Instinct*. Penguin.

——, & P. Bloom. 1990. Natural language and natural selection. *Behavioral and Brain Sciences* 13.707–784.

Smith, A. D. M. 2001. Establishing communication systems without explicit meaning transmission. In *Advances in Artificial Life: Proceedings of the* 6^{th} *European Conference on Artifical Life*, ed. by J. Kelemen & P. Sosik. Berlin: Springer-Verlag.

Smith, K. 2002a. Compositionality from culture: the role of environment structure and learning bias. Technical report, Theoretical and Applied Linguistics, University of Edinburgh.

—— 2002b. The cultural evolution of communication in a population of neural networks. *Connection Science* 14.1–21.

Steels, L. 1998. The origins of syntax in visually grounded robotic agents. *Artificial Intelligence* 103.133–156.

——, F. Kaplan, A. McIntyre, & J. Van Looveren. 2002. Crucial factors in the origins of word-meaning. In *The Transition to Language*, ed. by A. Wray, 252–271. Oxford: Oxford University Press.