The Evolution of Meaning-space Structure through Iterated Learning

Simon Kirby

Language Evolution and Computation Research Unit
University of Edinburgh
40, George Square, Edinburgh, EH8 9LL

Abstract. In order to persist, language must be transmitted from generation to generation through a repeated cycle of use and learning. This process of *iterated learning* has been explored extensively in recent years using computational and mathematical models. These models have shown how compositional syntax provides language with a stability advantage and that iterated learning can induce linguistic adaptation. This paper presents an extension to previous idealised models to allow linguistic agents flexibility and choice in how they construct the semantics of linguistic expressions. This extension allows us to examine the complete dynamics of mixed compositional and holistic languages, look at how semantics can evolve culturally, and how communicative contexts impact on the evolution of meaning structure.

1 Introduction

One of the most striking aspects of human linguistic communication is its extensive use of compositionality to convey meaning. When expressing a complex meaning, we tend to use signals whose structure reflects the structure of the meaning to some degree. This property is the foundation upon which the syntax of language is built. It is natural, therefore, that an evolutionary account of human language should contrast compositional communication with a non-compositional, holistic alternative where whole signals map onto whole meanings in an arbitrary, unstructured way. Indeed, Wray (1998) has argued that holistic communication (which is still in evidence in particular contexts today) can be seen as a living fossil of an earlier completely non-compositional protolanguage.

A compositional syntax has clear adaptive advantages — with it we are able to successfully communicate novel meanings (in the sense that we may never have witnessed signals for those meanings in the past). Despite this, research over the past decade has suggested that compositional syntax may have emerged not because of its utility to us, but rather because it ensures the successful transmission of language itself (see e.g. Kirby, 2000). It is suggested that the process of linguistic transmission, termed *iterated learning* (Kirby & Hurford, 2002), is itself an adaptive system that operates on a timescale intermediate between individual learning and biological evolution. Computational models of this process (e.g. Kirby, 2000; Batali, 1998) have demonstrated that syntactic systems can emerge out of random holistic ones without biological evolution, at least for particular assumptions about learning, production and so on.

Further evidence for the argument that iterated learning can explain features of syntax has been provided by idealised computational (Brighton & Kirby, 2001) and mathematical (Brighton, 2002) models of iterated learning in general showing that compositional languages have a stability advantage over holistic ones. These models compare two scenarios under a number of different parameters. They analyse completely holistic languages and completely compositional ones. The parameters that are varied relate to, on the one hand, the structure of the meaning space, and on the other, the number of training examples an individual is exposed to (also known as the *bottleneck* on linguistic transmission). The overall conclusion is that with highly structured meaning spaces and few training examples, compositional languages are more stable than holistic ones.

2 Problems

This foundational work on the cultural evolution of meaning-signal mappings through iterated learning, though important in demonstrating that language itself has significant adaptive dynamics, suffers from two significant drawbacks, which we will turn to below.

2.1 Stability analysis

Early models such as Batali (1998) and Kirby (2000) involved populations of individual computational agents. These agents were equipped with: explicit internal representations of their languages (e.g. grammars, connection weights etc.); a set of meanings (provided by some world model) about which they wished to communicate; mechanisms for expressing signals for meanings using their linguistic representations; and algorithms for learning their language by observing meaning-signal pairs (e.g. grammar induction, back-propagation etc.).

Typically, these simulations initialise the population with no language, or a random pairing of meanings and signals and then allow the linguistic system to evolve through repeated encounters between speaking agents and learning agents.

There has been much work in building simulation models within this general iterated learning framework (e.g. Batali, 1998; Kirby, 2000; Tonkes, 2001; Kirby & Hurford, 2002; Brighton, 2002; K. Smith, 2003; Zuidema, 2003). The great advantage of this kind of modelling is that it allows the experimenter to demonstrate possible *routes* by which language can evolve from one qualitative state, such as holistic coding, to another, such as compositionality. The models show how fundamental features of language can emerge in a population over time given reasonable assumptions about how linguistic behaviour may be transmitted.

The emergence of compositionality in particular has received a lot of attention. However, it is important to note that other fundamental linguistic universals may well be explicable within this general framework. The central message is that wherever there is iterated learning, there is potential for adaptation of the system being transmitted to maximise its own transmissibility.

Models such as these tend to have a large range of parameters, and it is therefore reasonable to want to know the relationship between the emergent property and the parameter space of the model. Once we understand this, we can eventually hope to uncover theoretical principals that may apply to iterated learning *in general* rather than the specific model in question.

As mentioned above, two key parameters in the emergence of compositionality are: meaning-space structure (i.e. the set of things agents communicate about); and learning bottleneck[1] size (i.e. the number of training examples agents are exposed to).

Computational simulations indicate that it is important that there is some kind of learning bottleneck for there to be any interesting linguistic evolution. To put it simply, only when training data is sparse will language evolve to be compositional.

This parameter is relatively straightforward to experiment with, but meaning-space structure is far more difficult, and most of the simulations of iterated learning simply chose some kind of system of meaning representation and stick with it for all simulations.

The work of Brighton & Kirby (2001) and Brighton (2002) was an attempt to get round this problem by exploring a large range of possible meaning-spaces and examining what impact they would have in an iterated learning model.

In those papers — as in this one — a highly idealised notion of "meanings" is employed: meanings are simply feature vectors. A meaning-space is defined by the number of features F it has and the number of different values V over which each feature can vary. So, to communicate about a world where objects were either squares, circles or triangles, and could be coloured green, blue or red, agents would need a meaning-space with at least $F = 2$ and $V = 3$. A red triangle would thus be represented with the *colour* feature taking the *red* value and the *shape* feature taking the *triangle* value.

A reasonable strategy for thoroughly exploring the role of meaning-space structure might be to run many iterated learning simulations, each with a different meaning space, and determine the trajectory of the linguistic system in each instance. This proves computationally costly, so Brighton and Kirby instead looked at what would happen to either a completely compositional language or a completely holistic one for each meaning-space.

Firstly using a computational model, and then using a mathematical generalisation of this model, they were able to calculate how stable either language type was for all meaning spaces. Simplifying somewhat, the overall result was that compositional languages have a stability advantage over holistic ones for larger meaning spaces, especially where the number of features is high.

This kind of simplification of the iterated learning process is very useful but leads to the first of our two problems. Whereas a standard iterated learning sim-

[1] See Hurford (2002) for discussion of why the term "bottleneck" is appropriate, and for an analysis of different types of bottleneck in language evolution.

ulation can demonstrate a trajectory, or route, from holism to compositionality, the Brighton and Kirby idealisation can only tell us about the relative stability of end-points of such a trajectory. In other words, we don't know whether there is a way to get to a stable compositional language from an unstable holistic one because we don't know anything about the languages in-between.

2.2 Fixed, monolithic meaning space

A second problem with much research into iterated learning so far has been its reliance on a pre-existing meaning space provided for and shared by all agents in the simulation.[2] The work described in the previous section makes strong claims about the likelihood of the emergence of compositional syntax given a particular prior space of meanings. But, where does this meaning space come from? It is assumed that biological evolution somehow endows the agents with a representational scheme prior to language, and if those representations are of sufficient complexity, a compositional system of expressing them will follow naturally.

Furthermore most, if not all, models assume that there is a single, monolithic system for representing meanings. Everything the agents in the simulations want to talk about can be expressed in the same format, be that a feature vector of particular dimensionality, a predicate-logic representation, or a point on a real-number line etc. Equally, there is assumed to be one and only one meaning for representing every "object" in the agents' world. (The term "object" is used here by convention to stand-in for any communicatively relevant situation. In other words, an "object" is anything that an agent may wish to convey to another agent through language.)

As with the study of the relative stability of "end-points" in language evolution, a monolithic, fixed and shared meaning-space is a sensible idealisation to make. Modellers hold one aspect of the object of study constant — meanings — and allow another aspect — signals — to evolve through iterated learning. Much has been learned through these idealisations, but equally it is important to explore what happens if we relax these assumptions.

3 A simple model

In this paper I will set out a simple extension to the model in Brighton (2002) which allows us to look at what happens when agents have flexible meaning representations for objects. It turns out that this extension also allows us to move beyond a simple stability analysis of end-points of iterated learning and give us, for the first time, a complete view of the dynamics of iterated learning.

[2] This is not true of the extensive work on symbol grounding carried out by, for example, Steels & Vogt, 1997; Steels, 1998; A.D.M. Smith, 2003; Vogt, 2003.

3.1 Meanings

Language can be viewed as a system for mapping between two interfaces (see, e.g., Chomsky, 1995). On the one hand, there is an articulatory/perceptual interface, which handles input and output of signals. On the other, there is a conceptual/intentional interface, which relates linguistic representations to the things we actually communicate about. It is primarily the latter of these two that we are concerned with here.

In the model, there is a predefined set of things about which the agents wish to communicate — we will call this the *environment*, E. The conceptual/intentional interface C consists of a number of *meaning spaces* $M_{\langle F,V \rangle} \in C$ onto which every object $o \in E$ in the environment is mapped. Each of these meaning spaces, in keeping with previous models is defined as a set of feature-vectors, such that each meaning space is defined by the number of features F it has (its *dimensionality*), and the number of values V each of these features can take (its *granularity*). (For simplicity we will assume that there are the same number of possible values each feature can take. So, in our earlier example in section 2.1, both the *shape* feature and the *colour* feature ranged over three possible values – *square*, *circle*, *triangle* and *green*, *blue*, *red* respectively.)

Throughout a simulation run, every object in the environment is paired with a particular point in every meaning space. For the simulation runs described here, this is set up completely randomly at the start of the run. Loosely speaking, we can think of this as giving an agent a number of different ways of conceiving an object. Note that each point in each meaning space can be mapped to zero, one or many objects in the environment. So, for example, there may be particular feature-vectors in particular meaning spaces that are *ambiguous* in that they map to more than one object in the environment.

The important point here is that agents are prompted to produce expressions for *objects in the environment* and not meanings themselves. Part of the task of the agent is to choose which of that object's meanings will be used to generate the linguistic expression. It is this that is the novel extension to previous work. Previously, only one meaning-space was available, so expressing an object and expressing a meaning were the same thing. Now that the latter is under the control of the agent the use of meanings can be learned and, ultimately, itself be subject to cultural evolution through iterated learning.

3.2 Learning

In this model I will follow Brighton (2002, 2003) in considering the task of learning a compositional system to be one of memorising signal elements that correspond to particular values on particular features. A single compositional utterance carries information about how to express each feature-value of the meaning expressed by that utterance.

If we consider just a single meaning space, then learning a perfect compositional system proceeds exactly as in Brighton (2002, 2003). The learner is exposed to a series of R meaning/signal pairs $(p_1, p_2, \ldots, p_R)$ each of which represents a point in the space $F \times V$. After this exposure, the learner is able to express at least as many meanings as are uniquely expressed in the training data. Note that this is likely to be less than R since meanings may be repeated.

Is this the best expressivity that the learner can expect to achieve after learning? Not if the learner is exposed to a compositional language. The learner may be able to express novel combinations of feature-values as long as each feature-value occurs somewhere in the training data.

Brighton (2003) gives the following simple approach to modelling the transmission of a compositional language. The first step is to construct a lookup table recording how each feature-value is to be expressed. This table, O, is an $F \times V$ matrix of signal elements. In fact, in this model the actual nature of those signal elements is irrelevant. This is based on the assumption that the learner can correctly generalise a compositional language from the minimum exposure. Brighton terms this the *assumption of optimal generalization.* (This idealises away from the task of decomposing the input signal into parts and identifying which parts of the signal correspond to which parts of the meaning. We should be aware that, in a more realistic scenario, more data is likely to be required and furthermore, segmentation errors are likely to occur.)

The benefit of this assumption is that we can simply treat each entry in the O matrix as a truth value:

$$O_{i,j} = \begin{cases} \textbf{true} & \text{if the } j\text{th value of the } i\text{th feature is observed} \\ \textbf{false} & \text{otherwise} \end{cases}$$

When the entry $O_{i,j}$ is true, this means that the sub-signal for the jth value of the ith feature has occurred at some point in the training data.

On receiving some meaning/signal pair $p = \langle m, s \rangle$ the matrix is updated so that each of the feature-values contained in m are logged in the O matrix. If $m = (v_1, v_2, \ldots, v_F)$, then:

$$O_{i,v_i} = \textbf{true} \text{ for } i = 1 \text{ to } F$$

So far, this is simply a restatement of Brighton's (2003) formalism. The novel feature here is just that there are multiple meaning-spaces, and therefore multiple O matrices to keep track of. To simplify matters for this paper, we will maintain the assumption that learners are given meaning-signal pairs. That is, learners are able to infer which point in which meaning-space a speaker is expressing. It is a topic of crucial and ongoing research, particularly by those researchers looking at symbol-grounding, to develop strategies to relax this assumption (e.g., Steels & Vogt, 1997; A.D.M. Smith, 2003).

So far, contra Brighton (2002, 2003), we have not looked at *holistic* languages. Holistic languages are those where meanings are unanalysed and each given distinct, idiosyncratic signals. Learners cannot, therefore, generalise beyond the data that they are given. However, we can simply equate a holistic language with a compositional language for a meaning-space with only one feature. The machinery described so far, is therefore sufficient to explore the difference between compositional and holistic language learning — we simply need to provide agents with the relevant meaning-spaces.

3.3 Language production

We have specified an **environment** containing **objects** each of which are labelled with **feature-vectors** drawn from each of a set of **meaning-spaces**. We have set out a model of learning whereby sets of **meaning-signal pairs** given to a learning agent are transformed into **O matrices**, one for each meaning-space.

In order to complete a model of iterated learning, it is necessary to provide agents not just with a way of learning, but also a way of producing behaviour for future generations of agents to learn from.

Clearly, a particular meaning $m = (v_1, v_2, \ldots, v_F)$ can be expressed by an agent if, and only if, that agent has a way of expressing each feature-value using the language it has learned so far. In other words, iff $O_{1,v_1} \wedge O_{2,v_2} \wedge \ldots \wedge O_{F,v_F}$.

It is important to note, however, that the agents in this model are not prompted to express a *meaning*. Rather, they attempt to produce expressions for *objects* in the environment. This means that an agent may have a choice of potential meaning spaces to employ when signalling about any one object. An object is expressible, therefore, if *any* of the meanings associated with that object are expressible. If more than one meaning is expressible by an agent, a choice must be made. For the first simulations described below, that choice is simply made at random.

The goal of language production in this model is to produce a meaning-signal pair. However, learning as described in the previous section actually makes no use of signals because of the assumption of optimal generalisation. This means we can ignore the signal part of the signal-meaning pair. When a learning agent observes the behaviour of a speaker, the simulation need only note the set of meanings used.

3.4 Simulation run

A simulation run consists of the following steps:

1. **Initialise environment.** Associate each object in the environment with a single random meaning in every meaning space.
2. **Initialise population.** In this simple model, the population consists of a single speaker, and a single learner. At the start of the simulation, the O matrices of the adult speaker are initialised with patterns of "true" and "false". The particular way in which they are filled depends on the experiment being run, and represents the initial language of the simulation. The learner's O matrices are filled uniformly with "false" because learners are born knowing no language.

3. **Production.** An object is picked randomly from the environment. A list of candidate meanings — one from each meaning space — is compiled for the object. The O matrices of the speaker are used to determine which, if any, of these candidates the speaker can express. One of these is picked at random.
4. **Learning.** If the speaker has been able to find an expressible meaning, the learner takes that meaning and updates its own O matrix for that meaning space.
5. **Repeat.** Steps 3 and 4 are repeated R times (this defines the size of the learning bottleneck).
6. **Population update.** The adult speaker is deleted, the learner becomes the new speaker, and a new learner is created (with O matrices filled with "false" entries).
7. **Repeat.** Steps 3 to 6 are repeated indefinitely.

The relevant simulation parameters are: size of bottleneck, R; number of objects in the environment, N; the make-up of the conceptual/intentional system, C (i.e. the particular $\langle F, V \rangle$ values for each $M_{\langle F,V \rangle}$); and the initial language (i.e. the O matrices for each meaning space in C).

4 Results

This simulation model can be used to explore the dynamics of iterated learning given multiple meaning-spaces. Because, as mentioned earlier, holistic languages are identical to compositional languages for 1-dimensional meaning-spaces, it can also be used to examine how compositional communication can arise out of a prior holistic protolanguage.

4.1 Meaning space stability

As many previous models have shown, compositional languages are more stable than holistic ones through iterated learning with a bottleneck. We can track *expressivity* of the agents' languages in a simulation over generations given an initial completely expressive language that is compositional, and compare that with a simulation initialised with a completely expressive language that is holistic. Expressivity is defined simply as the proportion of all the objects in the environment that an agent is able to find an expression for.

iteration	0	1	2	3	4	5	6	7	8
holistic	1	.45	.22	.13	.08	.02	.02	.02	0
comp.	1	1	1	1	1	1	1	1	1

Table 1. Expressivity over time for a simulation with $N = 100, R = 50, C = \{M_{\langle 8,2 \rangle}\}$ and a simulation with $N = 100, R = 50, C = \{M_{\langle 1,256 \rangle}\}$.

Unsurprisingly, the holistic language cannot survive in the presence of a bottleneck. The size of the bottleneck affects the rate of decay of expressivity in the holistic language (table 2). As in previous models, this demonstrates once again the crucial advantage a language gains from a compositional syntax.

iteration	0	50	100	150	200	250	300
R=100	1	0	0	0	0	0	0
R=200	1	.15	.1	.06	.06	.04	.02
R=300	1	.3	.21	.16	.16	.16	.12
R=400	1	.61	.43	.38	.34	.32	.31

Table 2. Rate of decay of expressivity in holistic meaning spaces varies with size of bottleneck.

4.2 Complete holistic/compositional dynamics

Recall that one of the motives for this extension to previous work is to move beyond simple stability analysis to see the complete dynamics of the move from holism to compositionality. To do this, we can simply run simulations with two meaning spaces instead of one, such as: $C = \{M_{\langle 8.2\rangle}, M_{\langle 1,256\rangle}\}$.

A particular point in the space of possible languages can be described in terms of the proportion of objects that can be expressed using the compositional language, $M_{\langle 8,2\rangle}$, and the proportion of objects that can be expressed using the holistic language, $M_{\langle 1,256\rangle}$.

The complete dynamics for all points in holistic/compositional space is visible in the top graph in figure 1. The arrows show the magnitude and direction of change after one iteration of the model for that particular combination of holistic versus compositional expressivity. There is a single attractor at (0,0). In other words, the inevitable end state is one where no objects are expressible either holistically or compositionally.

The reason for this is obvious: once a word is lost from the language, there is no way of getting it back. In fact, the agents rely on the expressivity of the language that is injected at the start of the simulation. To get round this, most iterated learning models allow agents to “invent” new expressions. To model this, a new parameter is added — the invention rate I. This gives the probability that, on failure to find any way of expressing an object, an agent will pick a meaning space at random and invent an expression for the relevant meaning in that space.

The bottom graph in figure 1 shows how an invention rate of $I = 0.1$ affects the dynamics of iterated learning. Now, the single attractor is the completely compositional language. This demonstrates that there is a clear route from all parts of the language space towards a completely compositional language, through intermediate mixed languages.

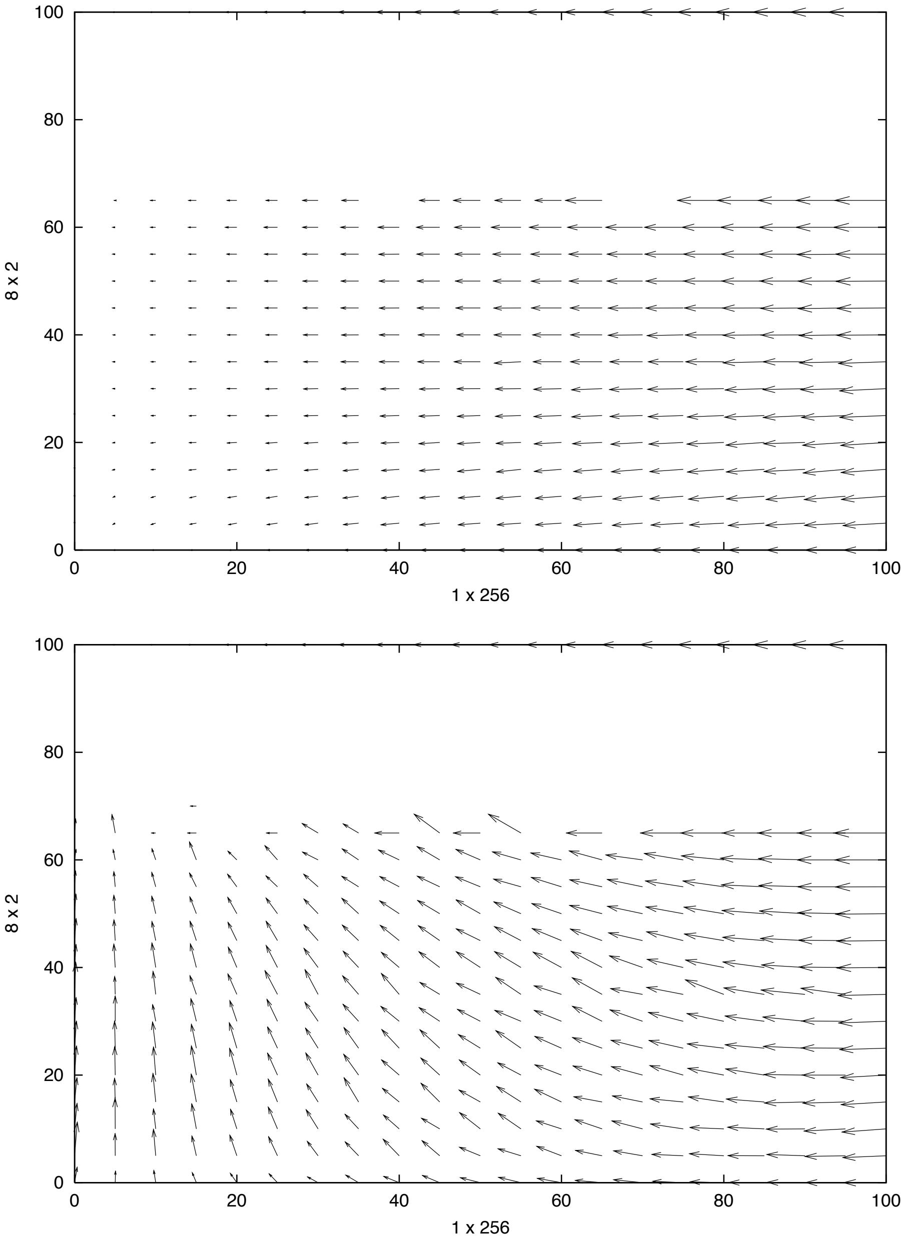

Fig. 1. Complete dynamics for languages that are partially holistic and partially compositional, without invention (top graph) and with invention (bottom graph). Each point represents a language with a particular combination of holistic and compositional signals. Each arrow shows the direction and magnitude of movement in this space after a single instance of learning, and represents the average of 100 simulation runs. (The gaps in the graph result from points in this space that cannot be constructed for an environment of 100 objects.)

As has been shown before, the size of bottleneck is a crucial determinant of whether compositionality will replace holism. If the size of the bottleneck is increased, holistic utterances no longer have such a disadvantage and the movement to the left-hand side of these plots is removed. It is the fact that language must pass through a learning bottleneck as it is transmitted from generation to generation that causes it to adapt and causes idiosyncratic non-compositional expressions to die out.

4.3 The evolution of meaning spaces

The second motivation for the current model was to see how iterated learning might result in adaptation of the meanings of expressions as well as the form of the expressions themselves. Previous models used a monolithic, fixed meaning space, but the current model allows for any number of meaning spaces to exist concurrently. An agent's learning experience (and hence, ultimately, its cultural inheritance) decide the structure of the meaning used to express an object in the environment.

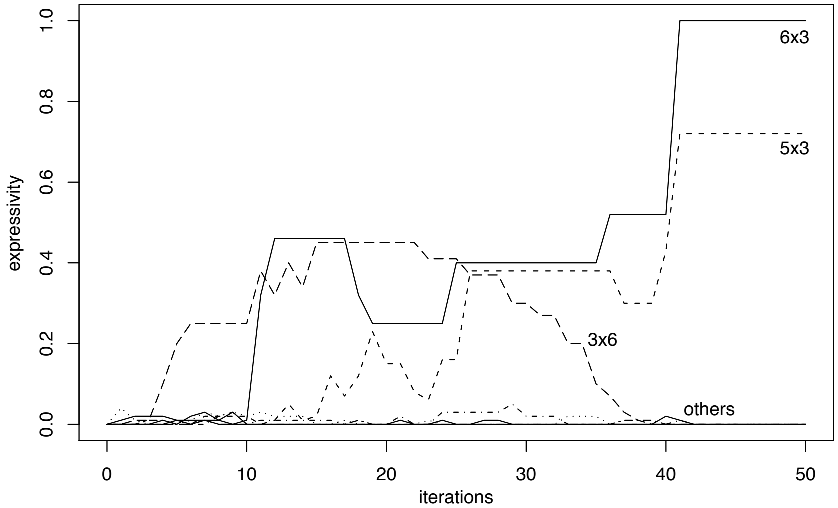

Fig. 2. Competition between different meaning-spaces through cultural evolution in a single simulation run where agents have conceptual systems capable of representing each object in many different ways (see main text for details).

The graph in figure 2 shows an example simulation run with $I = 0.1, N = 100, R = 50$ and the following conceptual system:

$$C = \{M_{\langle 1,256\rangle}, M_{\langle 2,16\rangle}, M_{\langle 3,6\rangle}, M_{\langle 4,4\rangle}, M_{\langle 5,3\rangle}, M_{\langle 6,3\rangle}, M_{\langle 7,3\rangle}, M_{\langle 8,2\rangle}\}$$

Table 3 shows the pattern of meaning space usage averaged over 100 simulations with these parameters measured at 50 generations. Despite being identical initially, agents end up using different systems of meaning for expressing objects in the environment in each simulation. In some runs, such as in figure 2, multiple meaning spaces remain partially expressive and stable. This means that agents may have different ways of expressing the same object. Real languages have different ways of carving up the world, and real speakers have different ways of expressing the same message. This simulation demonstrates a mechanism by which this can be acquired and can evolve culturally.

features	1	2	3	4	5	6	7	8
values	256	16	6	4	3	3	3	2
average expressivity	0	0	0	.11	.29	.15	.03	.45

Table 3. Expressivity of the various meaning spaces at the end of simulation runs where agents have a complex conceptual system allowing flexibility in the way objects are expressed (average of 100 simulations).

Are there any generalisations that can be made about the particular linguistic systems that emerge through this evolutionary process? A clear answer to this requires further research, but it may be that the meaning space adapts to structure in the environment. In the current model, the pairing between objects and points in meaning spaces is initialised randomly with uniform probability. A future version of the model will allow the experimenter to populate the environment with objects with non-uniform distribution in meaning space.

4.4 Uninformative meaning spaces and the role of context

In this model, there is a many-to-one mapping from objects in the environment onto meanings in any one meaning space. This means that the simulation can be set up in such a way that agents can produce expressions that are hugely ambiguous. Conceivably, a meaning space could be available that mapped all the objects in the environment onto one point. We can think of an agent using such a meaning space as expressing every object as “thing”.

What happens in the iterated learning model when these “uninformative” meaning spaces are included? An experiment was run with the following parameters:

$$I = 0.1, N = 100, R = 50,$$

$$C = \{M_{\langle 1,256\rangle}, M_{\langle 8,2\rangle}, M_{\langle 2,2\rangle}\}$$

In this situation, the agents end up expressing all 100 of the objects in the environment using the two-by-two meaning space. To put it another way, they use two word sentences with a vocabulary of four words. This kind of language is very stable since it requires very little data to learn.

This seems a rather implausible result. In reality, language is used to communicate rather than merely label objects. To simplify somewhat, in a particular situation, a speaker may attempt to draw a hearer's attention towards one of a range of possible objects in the current context.[3] If all the objects in the context map to the same meaning in the language, then no expression could be possible that would successfully direct the hearer's attention. Only if the context size was minimised could an uninformative meaning space hope to discriminate the intended object from the others, but in the limit this essentially renders communication irrelevant. If there is only one possible object to talk about, then the hearer will already know what it is.

Contexts can be added to the simulation model relatively easily. Speakers are given a target object and a number of other objects that form the context. When choosing a meaning space to use to convey the target, speakers will reject meanings that fail to discriminate the target from one or more of the objects in the context.

Repeating the previous simulation with a context of 5 objects leads to the domination of the informative eight-by-two meaning space over the uninformative two-by-two one. This result demonstrates once again how iterated learning can result in language adapting over a cultural timescale to the particular constraints placed on its transmission.

5 Conclusions

In this paper I have shown how previous models of iterated learning which used monolithic meaning spaces can be extended to deal with a more flexible notion of meaning. By allowing agents choice over the semantics of linguistic expressions, we can see how meanings as well as signals evolve culturally.

This extension has allowed us to expand on earlier analyses of the relative stability of completely compositional versus completely holistic languages to look at the complete dynamics of a space of languages that are partially compositional. In addition, we can look at far more complex systems with ambiguity of meaning, varying degrees and types of compositionality and semantic structure, and examine how communicative contexts affect the way language is transmitted.

There is much work to be done in this area — I consider this model to be a preliminary investigation only. Many possible extensions of the model could be worth pursuing. For example, the results suggest a puzzle: why aren't all languages binary? The binary meaning spaces seem to be highly stable in the model, but nothing like this exists in natural language. What is needed is a more realistic treatment of semantics and also considerations of signal complexity.

[3] Recall that "object" here is merely a term of convenience. We might wish to gloss this with "communicative intention".

Natural language semantics does not take the form of fixed-length vectors, and there are plausible pressures to keep signals short.

Another interesting direction would be to combine this kind of idealised model with the mechanisms for collaborative meaning construction and grounding developed by those working with robotics models (e.g., Steels & Vogt, 1997; Steels, 1998; Vogt, 2003; Cangelosi, 2004). In this manner, we may begin to be able to relate abstract notions of expressivity, learnability and stability with the particular features of natural language semantics grounded in the real world and embodied in human agents.

The overarching conclusion of this line of work is that iterated learning is a surprisingly powerful adaptive system. The fact that language can only persist if it is repeatedly passed through a transmission bottleneck — the actual utterances that form the learning experience of children — has profound implications for its structure. This point has been made clear before in relation to the syntax of language. The model in this paper shows that the semantics of language are also likely to have been shaped by iterated learning.

References

Batali, J. Computational simulations of the emergence of grammar. In Hurford, J. R., Studdert-Kennedy, M. and Knight C., editors, *Approaches to the Evolution of Language: Social and Cognitive Bases*. Cambridge: Cambridge University Press, 1998.

Brighton, H. Compositional Syntax from Cultural Transmission. *Artificial Life*, 8(1):25–54, 2002.

Brighton, H. *Simplicity as a Driving Force in Linguistic Evolution*. PhD thesis, Theoretical and Applied Linguistics, The University of Edinburgh, 2003.

Brighton, H. and Kirby, S . The Survival of the Smallest: Stability Conditions for the Cultural Evolution of Compositional Language. In J. Kelemen and P. Sosk, editors, *ECAL01*, pages 592–601. Springer-Verlag, 2001.

Cangelosi, A. The sensorimotor bases of linguistic structure: Experiments with grounded adaptive agents. In S. Schaal et al., editor, *SAB04*, pages 487–496. Los Angeles: Cambridge MA, MIT Press, 2004.

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

Hurford, J. Expression/induction models of language evolution: dimensions and issues. In Ted Briscoe, editor, *Linguistic Evolution through Language Acquisition: Formal and Computational Models*. Cambridge University Press, 2002.

Kirby, S. Syntax without Natural Selection: How compositionality emerges from vocabulary in a population of learners. In C. Knight, editor, *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form*, pages 303–323. Cambridge University Press, 2000.

Kirby, S. and Hurford, J. The Emergence of Linguistic Structure: An overview of the Iterated Learning Model. In Angelo Cangelosi and Domenico Parisi, editors, *Simulating the Evolution of Language*, pages 121–148. London: Springer Verlag, 2002.

Smith, A. D. M. Intelligent Meaning Creation in a Clumpy World Helps Communication. *Artificial Life*, 9(2):559–574, 2003.

Smith, K. *The Transmission of Language: models of biological and cultural evolution*. PhD thesis, Theoretical and Applied Linguistics, School of Philosophy, Psychology and Language Sciences, The University of Edinburgh, 2003.

Steels, L. The origins of syntax in visually grounded robotic agents. *Artificial Intelligence*, 103(1-2):133–156, 1998.

Steels, L. and Vogt, P. Grounding adaptive language games in robotic agents. In I. Harvey and P. Husbands, editors, *ECAL97*. Cambridge, MA: MIT Press, 1997.

Tonkes, B. *On the Origins of Linguistic Structure: Computational models of the evolution of language*. PhD thesis, School of Information Technology and Electrical Engineering, University of Queensland, Australia, 2001.

Vogt, P. Iterated Learning and Grounding: From Holistic to Compositional Languages. In Simon Kirby, editor, *Proceedings of Language Evolution and Computation Workshop/Course at ESSLLI*. Vienna, 2003.

Wray, A. Protolanguage as a holistic system for social interaction. *Language and Communication*, 18(1):47–67, 1998.

Zuidema, W. How the poverty of the stimulus solves the poverty of the stimulus. In Suzanna Becker, Sebastian Thrun, and Klaus Obermayer, editors, *Advances in Neural Information Processing Systems 15 (Proceedings of NIPS'02)*. Cambridge, MA: MIT Press., 2003