Iterated Learning: The Exemplar-based Learning Approach

Joseph Charles Eddy, B. Sc.

MSc Evolution of Language and Cognition
The University of Edinburgh
2005

1 Introduction

The Iterated Learning Framework (Kirby 2002a, Kirby 2002b, Kirby and Hurford 2002) has been used to explore how human-type languages can emerge and evolve through a process of cultural transmission, given that agents have the necessary biological hardware to learn, use and process it. Different models have different built-in assumptions about the mechanics of language learning, use and processing, and can range from connectionist models (i.e. Tonkes 2002) to rule-induction models and other symbolic models (i.e. Kirby 2002a, Kirby 2002b, Kirby and Hurford 2002) to models in which agents are driven mostly by an attempt to find a compressible grammar (i.e. Teal and Taylor 1999, Brighton and Kirby 2001, Brighton 2002). The research described in this dissertation will deal mostly with an exemplar-based learning model described by Batali (2002), and will go into an in-depth examination of certain aspects of the model and attempt to answer two specific questions: 1) How does the lack of population turnover affect the behaviour of the model, if at all? 2) Does exemplar discouragement and pruning in the model implement a linguistic "bottleneck" with effects on the model similar to those described by Kirby and Hurford (2002, Kirby 2002b). I will also go into some comparisons with other models of the emergence of compositional language, specifically with the symbolic rule-induction model described by Kirby (2002b) and with findings discussed by K. Smith (2002), as well as other interesting things that came up in the course of the research, especially having to do with the effect of the types, complexity, and distribution of meanings given to the agents to discuss on the behaviour of the model.

The layout of the dissertation is as follows: First, a background section on Iterated Learning, in which iterated learning is defined, a layout of the components of the iterated learning framework is provided, a classification of different learning strategies into categories is discussed, some studies will be presented that look into how syntax can arise from a process of iterated learning, the role of meanings in the iterated learning model is discussed, some questions that the iterated learning model tries to answer will be laid out, and a brief look will be taken at some models that examine the interaction between learning and biological evolution. After the background section follows a section that looks more deeply into the exemplar-based learning model described by Batali (2002), what is unique about it, and what questions I attempt to answer with my research. Specifically, I will briefly describe Batali's model, discuss the lack of population turnover in his model, propose the existence of a "bottleneck" in the model, based on Batali's exemplar discouragement/pruning scheme, look into the "flat" semantic representation used, and compare it to the semantic representation used by Kirby's (2002b) symbolic rule-induction model, and finally discuss the history and merits of exemplar-based learning models, in general.

Following this discussion of Batali's model is a results section, in which I discuss differences between my implementation of the model, and the implementation originally put forward by Batali (2002), the results of a basic simulation using the model, and more detailed results addressing the specific questions my research was intended to address, as well as other interesting results that came out during the research. Finally, a general conclusion and discussion section is supplied at the end.

2 Background – Iterated Learning

The origins and evolution of human language can be seen as one of the most difficult problems in science. This is because it arises from, and continues to be influenced by, the interaction of multiple complex systems, each of which evolves adaptively, and each of which influences the adaptive evolution of the other. Human language evolves over time through different mechanisms including biological adaptation of human phylogeny over the course of many generations over thousands of years, individual learning during the ontogenetic lifespan of individuals, and cultural evolution from generation to generation, which lies on a time scale somewhere between those two (called glossogenetic: Hurford 1990). Kirby (2002a) details the interaction between these systems, and how they can influence each other, as summarised in figure 1.

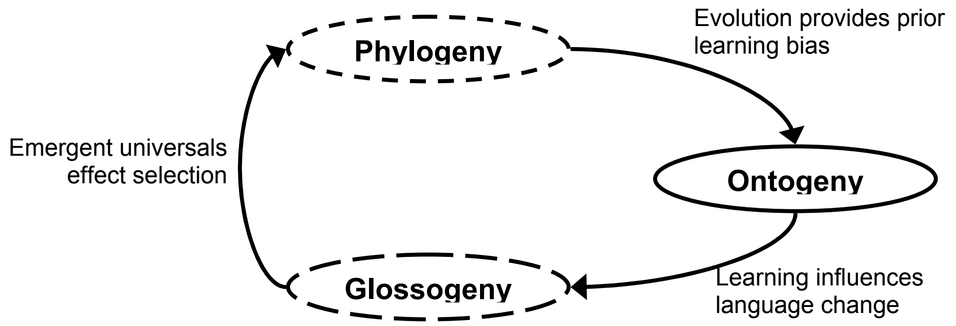


Figure 1: "Language is the result of an interaction between three complex adaptive systems that operate on different timescales: the timescale of biological evolution (phylogeny), the timescale of individual learning (ontogeny), and the timescale of language change (glossogeny)" (reproduced from Kirby 2002a, page 190).

What is represented by the bubbles in figure 1 and the interactions between them can be described as follows (Kirby 2002a): humans learn languages by observing the language use of others (ontogeny), languages change through this process of cultural learning (glossogeny), which changes the environment, and along with it the selection pressures that guide human biological evolution (phylogeny). This, in turn, can cause changes in the learning mechanisms that humans bring the problem of language learning, which again changes the way in which humans learn languages from their observations in a cultural context…which brings us back to where we began, resulting in a very drawn-out historical process of interactions between dynamically complex systems that is not completely unrelated (an can be seen as an extension of) Hurford's Diachronic Spiral (Hurford 1987).

Kirby and Hurford (2002) introduced a model whereby the evolution of cultural entities (such as human language) can be seen as (at least weakly) analogous to the biological evolution of organisms (though see discussion in Kirby, Smith and Brighton 2004 for a discussion on the dangers of taking this analogy too far). In this model, it is maintained that human languages, or certain aspects of human languages, can be seen as organisms living in an environment that is made up of human brains, utterances, and written work, and that these languages have differing fitness with respect to their environment, which causes them to evolve over time, changing in an adaptive fashion as languages and language constructs that are maladjusted to us, the language users, die off over time.

This model (called the Iterated Learning Model), as described by Kirby (2002a) is focused on a historical process whereby information in transmitted across generations through a repeating cycle in which people learn from the performance of others (especially their elders), and later exhibit behaviour based on that learning to create performance data for the next generation. Specifically in relation to language, this model focuses on a process whereby observed linguistic utterances of an older (teacher) generation shape the internal state of a younger (learner) generation, which then eventually become the teacher generation and create linguistic utterances of their own, based on what they had learned from their teachers.

It should be noted here that there is a difference between the historical process of the cultural evolution of language described by the Iterated Learning Model, and the historical process of language change studied by historical linguistics. The Iterated Learning Model looks at *qualitative* changes in the state of a language (i.e. from holistic to compositional), whereas historical language change studies movement of a language within one of these language states.

2.1 Components of the Iterated Learning Framework

The iterated learning framework consists of four major components (Brighton 2002, Brighton & Kirby 2001, Kirby 2001, Kirby & Hurford 2002):

1) A meaning space
2) A signal space
3) One or more learning agents
4) One or more adult agents

The structure of these components in any given incarnation of the iterated learning framework can vary widely: the meaning space can consist of some set of atomic meaning parts that can be re-combined to create more complex meanings (i.e. Kirby 2002b, Batali 2002), or it could consist of a more continuous set of meanings, such as a continuous colour space (i.e. Belpaeme 2001); the signal space can consist of strings built up from discrete characters (i.e. Kirby 2002b), or it could consist of a more continuous range of acoustic signals, in which case the agents might have to negotiate a shared set of signal categories, instead of or in addition to negotiating a mapping between these signals and their corresponding meanings in the semantic space (i.e. De Boer 1999).

The population of learning agents can be completely distinct from the population of adult agents with new agents always learning from the linguistic behaviour of adult agents, with no horizontal transmission (i.e. Kirby 2002b, Oliphant 1997, figure 2). The populations of learning and adult agents could be one and the same set of agents (i.e. Batali 2002, which he refers to as a model of grammar "negotiation" between an unchanging set of language-using agents). It would even be possible to build a model that lied somewhere between these two, in which all agents play dual roles of teacher and learner, but at some point agents may die out and be replaced by new agents starting out with no linguistic knowledge.

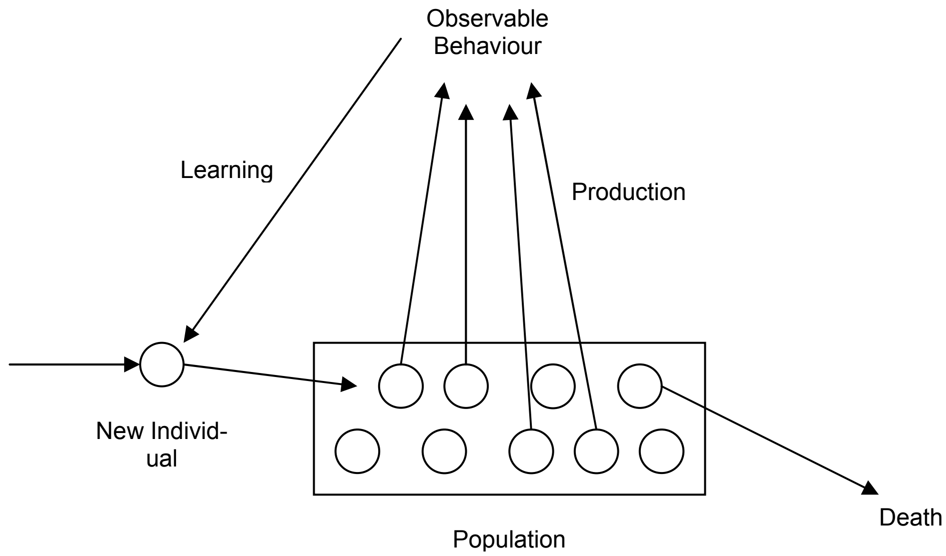


Figure 2: A cycle in which old agents continually die off and are replaced by new agents. The new agents learn from the communicative behaviour of the existing population (reproduced from Oliphant 1997, page 59).

2.2 Classification of Learning Categories

Another variable parameter in any iterated learning model is that of the observational learning strategies employed by the individual agents being modelled, and the innate biases that are built into those strategies. Oliphant (1999) classifies the abilities of populations of agents employing these different observational learning strategies into three categories:

1) Acquisition – the population is able to acquire an optimal system of communication
2) Maintenance – the population is able to acquire an optimal system of communication, even in the presence of noise
3) Construction – the population is able to construct an optimal system of communication when it begins with an initial random system

As Kirby (2002a) points out, it is essential to take iterated learning into account when studying the emergence of linguistic systems, since a learning strategy that allows an individual agent to succeed in learning a system of communication may not allow a population of agents employing such a learning strategy to construct the system of communication in the first place.

The importance of taking the process of iterated learning in a population into account when trying to explain the structures of natural languages is shown by Tonkes (2002), in his doctoral thesis. He shows that the adaptation of languages to the biases of general-purpose connectionist learners, who are not necessarily innately geared toward learning systems of communication, can result in the appearance that the agents have evolved into a population of especially "language-savvy" learners. He also shows that the iterated process of language production and acquisition can be used to help explain the *emergence* of optimally learnable languages, without positing especially language-specific constraints on the learning strategies of the individual agents.

K. Smith (2002) has shown that there is a crucial property that learning strategies of agents in a population must have in order to fall into the "constructor" category, that is, to be able to construct an optimal system of communication from randomness in the context of iterated cultural transmission: there must be a bias toward creating one-to-one mappings between signals and meanings. The model he uses consists of a population of agents that use Hebbian-like (Hebb 1949) learning networks to associate signals with meanings. During any given learning experience, the nodes representing the meaning as well as the nodes representing the corresponding signal are activated. Any pair of nodes will either be configured so that both of the nodes are active (*a*), both of the nodes are inactive (*b*), the meaning node is active while the signal node is inactive (*c*), or the signal is active while the meaning node is inactive (*d*).

A range of learning strategies were examined, in which for any of these cases the weight between the nodes was increased by one, decreased by one, or left unaltered.

The learning strategies that resulted in the ability to construct an optimal system of communication from randomness all shared two similar attributes: weights of connections between nodes with configuration (*a*) have a greater change (in the positive direction) than weights of connections between nodes with configuration (*c*), and weights of connections between nodes with configuration (*b*) have a greater change (in the positive direction) than weights of connections between nodes with configuration (*d*). In linguistic terms, this means that there must be a bias against synonymy as well as a bias against homonymy. In addition, communication systems that conform most closely to these biases (systems that are free of synonyms and homonyms) are more likely to be acquired by agents that exhibit them, with the result that communication systems with synonymy and homonymy will eventually be weeded out through a process of iterated learning, leaving the population with an optimal and unambiguous system of communication.

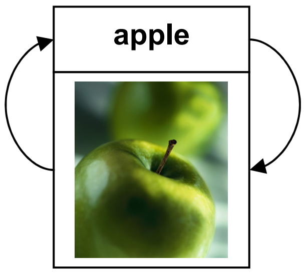


Figure 3: Mental representation of the mapping between meaning and signal as described by Saussure (1959).

The one-to-one mapping between meanings and signals that result from a process of iterated cultural transmission in a population of constructor agents is known as the Saussurean Sign (Saussure, 1959, figure 3). Oliphant and Batali (1997, also Hurford 1989) discuss how agents could learn a linguistic system that uses and maintains these mappings, and explain it in terms of what they call the 'Obverter' property. Under the obverter property, when an agent needs to express a meaning, it chooses the signal that, if it was heard by this agent, would most likely be interpreted as that meaning. Oliphant and Batali prove that populations of agents employing obverter learning strategies in an iterated learning context will always result in and optimal system of communication.

Batali (1998, reproduced with partial success by Goroll, 1999), studied the emergence of structured systems of communication in a population of neural-network based learning agents. He used a simple recurrent network that were fed sequences of characters one at a time at the input layer, and then treated the output layer after the last character was processed as the interpreted "meaning" of the string. To express a meaning, the agents basically chose the shortest string that would result in the closest match on their own output to the meaning to be expressed. Thus there is a sense of using an obverter strategy to express meanings, as well as a definite built-in bias toward shorter utterances. This model shows how slight statistical advantages of specific signal-meaning mappings can lead to structured agreement over time.

2.3 Syntax from Iterated Learning

One of the more interesting results to come out of studies of iterated learning is the emergence of syntax in the systems of communication of populations of agents in an iterated learning context. An important aspect of these communication systems is compositionality. A language is compositional if the meaning of an utterance is a function of the meanings of its constituent parts and the way they are put together (Kirby 2002a, Montague 1970). This is in contrast to holistic languages, in which there is an utterance for every meaning, and whose utterances are not mapped to meanings based on the meanings of their constituent parts and the way they are put together. Some examples of compositional English phrases and some equivalent holistic expressions are shown in the table below (from Kirby 2002a, page 203):

Compositional	Holistic
walked	went
I greet you	hi
died	bought the farm
I thought I saw a pussy cat	*bark* (vervet alarm call, *a la* Cheney and Seyfarth 1990)

Table 1: Examples of compositional English phrase and some equivalent holistic expressions.

Compositionality orders the mappings between meanings and signals in such a way that meanings that are nearer to each other in semantic space are expressed by signals that are more similar, this is in contrast to holistic systems of communications in which the closeness of meanings in semantic space bears little or no relation to the similarity of the strings used to express them (figure 4).

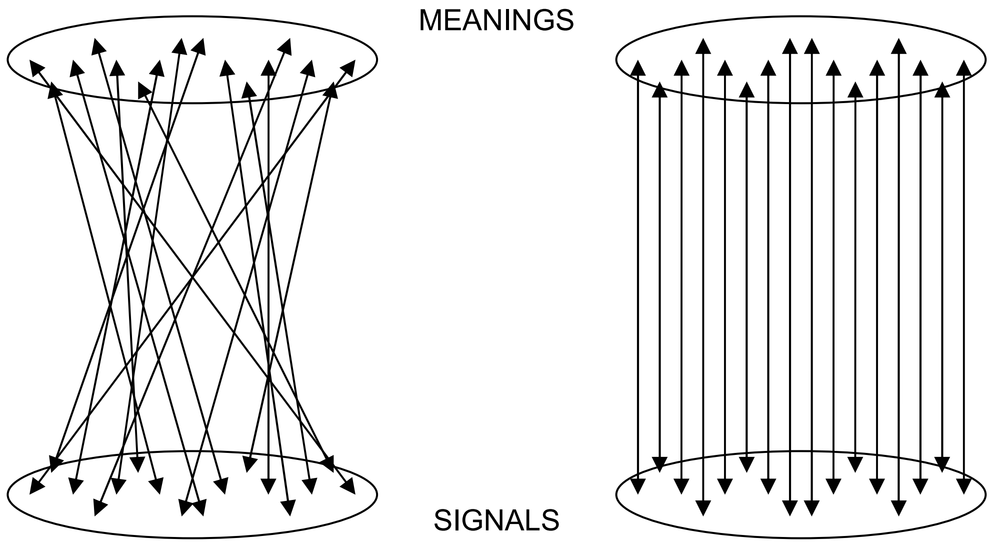


Figure 4: On the left, a holistic language with an unstructured mapping between meanings and signals, on the right, a compositional language in which the mapping is more structured (reproduced from Kirby 2002a, page 203).

It seems quite surprising, actually, how few compositional communication systems can be found in nature, given the number of communication systems that are used, and the efficiency and expressiveness that compositionality can impart to one. It seems that the only two that have been discovered so far are those used by humans, and the dance language employed by bees to communicate the location of food sources to the hive (von Frisch 1974).

Kirby (2000) describes a symbolic simulation in which agents represent linguistic knowledge internally in terms of context-free grammars. In this simulation, agents learn by observation alone, using the observed behaviour of other agents to guide their own and inventing random behaviour when necessary. The agents assign arbitrary numeric category labels to the parts on the right-hand side of their grammar rules. In the results of his simulation, Kirby shows that in the context of iterated learning, compositional languages eventually emerge, and the agents eventually reach some kind of agreement on how the constituent string-meaning pairs should be categorized, giving rise to what could be interpreted as a kind of semantic categorization.

Kirby and Hurford (2002, Kirby 2002b) use a symbolic simulation (quite similar to the one described in Kirby 2000) to explore the emergence of a recursive compositional communication system in a population of agents in the context of the iterated learning framework. The symbolic model they describe uses character strings as signals and a simple predicate logic variant for meanings. Agents in the simulation can store string-meaning pairs that they observe, and use them to determine their own production when prompted by the environment to express meanings as strings. They also have the ability to notice when two or more of the string-meaning pairs they have observed have simultaneous overlap in their string and meaning parts, and use that information to abstract more general rules to guide their production.

At the beginning of a simulation run, agents have entries in their grammars for every meaning they can express, and similarity between two meanings has little or no correlation with similarity of the strings used to express them. For example, a set of strings that an agent might use to express a set of corresponding meanings might be as follows (Kirby and Hurford 2002, pages 130-131, meanings are given as English glosses):

ldg
"Mary admires John"

xkq
"Mary loves John"

gj
"Mary admires Gavin"

axk
"John admired Gavin"

gb
"John knows that Mary knows that John admires Gavin"

After many generations, however, the mappings between meanings and the strings used to express them become much more structured, with similar meanings being expressed by similar strings. The set of utterances for the same meanings expressed above are given below, with spaces added to aid understandability for the reader:

gjh	f	tej	m
John		Mary	admires

"Mary admired John"

gjh	f	tej	wp
John		Mary	loves

"Mary loves John"

gjqp	f	tej	m
Gavin		Mary	admires

"Mary admires Gavin"

gjqp	f	h	m
Gavin		John	admires

"John admires Gavin"

i h	u	i	tej	u	gj	qp	f h	m
John	knows		Mary	knows		Gavin	John	admires

"John knows that Mary knows that John admires Gavin"

An important result that emerged from the model described by Kirby and Hurford was that the types of languages that emerged from the simulations depended largely on the number of observational episodes that agents in any given generation had as learners before they became the adult generation. They termed this restriction on the number of meanings expressed during learning for a given generation the learning 'bottleneck', also referred to as a "semantic bottleneck" as opposed to a "production bottleneck" as defined below (both definitions are from Hurford 2002, page 306).

Semantic Bottleneck – Individual speakers are prompted with only a (random) subset of meanings, so that the data given to an acquirer lacks examples of the expression of some meanings.

Production Bottleneck – Individual speakers' production mechanisms are designed to produce only a subset of the utterances that are possible for given meanings as defined by their grammars.

The languages that emerged from the simulations were qualitatively different, given different sizes of bottleneck, and they grouped these bottleneck size values into three different categories, based on the difference of the resultant emerging systems of communication.

If the bottleneck is too small, and the learning agents don't have enough learning episodes before becoming adult speakers, the result is an unstable, inexpressive language. Most meanings are expressed by holistic, unanalysed signals and the language changes rapidly from generation to generation without ever converging on a stable shared language. If the bottleneck is too large, and the learning agents see a large proportion of the possible meaning space before becoming speaking adults, the agents will eventually converge on a fairly stable, expressive language, but the eventual convergence on a stable shared language takes a relatively long period time.

For an intermediate size of bottleneck, though, the languages that emerge are completely stable, maximally expressive, as well as compositional. In addition to this, these languages emerge much more quickly than the fairly stable ones that emerge in simulations with large bottlenecks. It appears that this is the case because the intermediate bottleneck size puts a pressure on the *language* to adapt in such a way that they become compositional, whereas a large bottleneck removes that pressure. If agents can rote-learn a holistic string expression for every possible meaning in the meaning space, then there is no pressure for signal/meaning pairs to be analysed and broken down into constituent parts.

Batali (2002) uses an exemplar-based learning model to explore the emergence of compositional grammars in a population of agents. In his model, agents do not explicitly encode linguistic knowledge into grammatical rules, but instead directly store information from observational experience in the form of exemplars, and then use that information when prompted by the environment to produce their own utterances for other agents to observe. This model is especially interesting because of the fact that the agents never explicitly form grammatical rules, and also because the semantic representations are set up in such a way that their structure will not influence the ways agents find to express them. This model will be looked at in more detail below.

Teal and Taylor (1999) look at this language adaptation in terms of compression. They point out that the kind of generalisation across observed utterances described above is a way for agents to compress their knowledge of a language into a description that takes up less space (known as minimum description length, or MDL). They show that languages whose longer utterances are made by recombining constituents from a set of smaller utterances (what they describe as more "smooth") are more compressible, and that in an iterated learning context, languages evolve to become more compressible. A significant difference between the simulation described by Teal and Taylor and most of the others mentioned in this section is that their results deal only with signals and do not take meanings or semantics into consideration, in their model structure emerges in signals without being driven by structure in the meaning space.

Brighton and Kirby (2001, as well as Brighton 2002) also examine language adaptation in terms of MDL. This model attempts to deal with criticisms of models such as ones previously used by Kirby (2000) and Batali (2002), which argue that the learning algorithms used by the agents are too strongly biased towards compositionality, making the outcomes of the simulations inevitable. They argue that the minimum description length principle gives them a theoretical justification for the generalisations made by the agents in their simulation. Their findings uphold previous findings that compositional languages are more compressible than holistic ones, as well as more stable, and that iterated learning, in the presence of large, non-holistic meaning spaces and small transmission bottlenecks results in language adapting over time to become more compositional and more stable.

Zuidema (2001) discusses the mathematical model of Nowak, Komarova and Niyogi (2001) that explores the evolution of grammar in human language, and possible A-life extensions to the model to fit it into the iterated learning framework. He shows that the lower bound of the Q value they determine for the number of training samples that agents in a population need in order to achieve a coherent system of communication can be lowered further if iterated learning is taken into account. He also shows that some classical learnability problems, such as those discussed by Gold (1967), can be overcome by iterated learning.

Kirby (2001) explores the role that iterated learning can play in helping to understand the emergence and evolution of other features of language. Specifically, he looks at the how a non-uniform distribution in the space of meanings that are expressed by adult agents to learner agents from generation to generation can result in the emergence of regularity and irregularity, in the iterated learning framework. He shows that competing pressures on induction and production at an individual level can lead to compositionality under iterated learning, and features of regularity and irregularity when a non-uniformly distributed meaning space is taken into account.

Some research has also suggested that the iterated learning model can help explain the subjacency principle, which seems to be a universal feature of human language, without resorting to it being an innately specified feature, as it had been previously argued to be (i.e. Newmeyer 1991). Christiansen and Ellefson (2002) used a combination of Artificial Language Learning and neural network learning tasks to support the theory that subjacency evolved in order to satisfy non-linguistic constraints on human memory and comprehension, as opposed to a more classical view that it is an arbitrary rule specified as part of innate human Universal Grammar. They suggest, based on their results that the subjacency principle evolved through a process of language change, based on pressure from general human learning and cognition, and perhaps not through biological evolution of human Universal Grammar.

2.4 The Role of Meanings in the Iterated Learning Model

Most of the models discussed above use some kind of representation of meanings and semantic space, and the languages that emerge from them map strings or utterances onto those meanings. Teal and Taylor’s (1999) model doesn’t, and neither does Zuidema’s (2003, though he did build in a minimum expressiveness constraint which he describes in section 5 of his paper, which keeps his agents' grammars from completely collapsing under the compression pressure; this constraint can be interpreted as taking over the role that a meaning space might otherwise fulfill), but the results they find are quite similar to the results that arise from the models in which meanings are represented. For example, the grammars in Teal and Taylor's (1999) simulations became simplified over time under iterated learning in a way that is comparable to the simplification of grammars in Kirby's (2000) model. Also, in Zuidema's (2003) model, the learnability of languages increased and the number of rules needed to encode the language decreased while expressivity remained fairly constant, which seems to support similar results of increased learnability of the language evolving alongside a simplification of the encoding of the language in the other iterated learning models described above. This leads to the question: if the results of iterated learning are demonstrable using "semantic-free" models of language evolution, then why model meanings at all?

While it is true that basic syntax can be studied to some extent without taking account for meanings, and the two meaning-free models mentioned above can help to explain the emergence of certain syntactic regularities, we are also interested in the semantic side of things, specifically compositionality, which uses syntax to build a structured relationship between utterances and their meanings, as well as how the meanings space can be defined and partitioned through the process of repeated communication.

Different models represent meanings in different ways, and depending on those representations, can have different effects on the workings of those models. For example, the meanings in Tonkes's (2002) model are represented as a point on the one-dimensional line between (and including) 0.0 and 1.0 (i.e., 0.8125), whereas in Kirby's (2002b) model, they have a predicate logic-like representation, with a possible element of recursion (i.e., loves(John, Mary) or knows(Victor, loves(John, Mary))). In Kirby's model, the meanings represent actions that can be taken, as well as things that can take action or have action taken on them, and in the simulation, they are used to help guide the agents to make consistent generalisations over string-meaning pairs, and their recursive structure ends up being represented in the recursively-structured syntax of the agents' grammar. In Tonkes's model, the structure of the meanings space is represented by how near two points are in one-dimensional space, and the resulting languages were structured so that strings that were more similar mapped to meanings that were closer to each other in the meaning space. In both cases, the strings of the resulting languages matched closely with the structure of the meaning spaces, in Kirby's model, the agents learned a grammar with recursively-structured syntax that mapped directly onto the recursive structure of the meanings, whereas in Tonkes's model, the strings can be seen as a discrete representation of the continuously-valued meaning space, similar to how binary values are used by computers to represent continuously-valued floating-point numbers.

Batali (2002) uses a semantic representation that is similar to Kirby's in that meanings are given a predicate-like representation, and complex meanings are built up from smaller, atomic, constituent parts, but Batali's representation does not have built-in recursive structure (at least not in the sense that Kirby's model does), though recursive structure does arguably emerge in the languages negotiated by the agents in his simulation. A more in-depth discussion of how the difference between these two semantic representations is relevant will follow further below, along with a more detailed account of Batali's model.

Models that start with an *a priori* specified set of meanings have been criticised as unrealistic for two reasons. Firstly, they leave us with the question of where the meanings came from in the first place, and secondly, it may not be realistic that agents are really able to directly observe the meanings to be paired with utterances during language learning.

Vogt (2003) and A. Smith (2003) both attempt to circumvent the first criticism by using models in which meanings co-evolve with utterances within the iterated learning framework. In Vogt's model, agents are presented with a visual space (the environment) that has objects of different size, shape, and colour, arranged in two-dimensional space, and agents are able to partition the semantic space based on these features of the objects. It has been shown, using this model that for agents to converge upon a maximally efficient shared language, their set of meaning representations must eventually converge within some acceptable threshold, and the conditions under which this occurs have been explored. A. Smith's model shows how successful communication systems can emerge when agents are not allowed to transfer meanings to each other at all, through a process whereby the agents develop a highly synchronized set of internal conceptions, when agents can bring an intelligent strategy for creating meanings to the task.

As for the second criticism, the question of whether it is realistic to allow both meanings and utterances to be transferred during learning, it has been discussed in the literature (i.e. K. Smith 2002) and there has been some preliminary work toward models that don't make the assumption that learners can observe both meanings and utterances (i.e. A. Smith 2003), and it seems clear that the assumption has to be made for models in which the agents are not situated in an environment that they can somehow sense and create their own interpretations of, and that agents will need some similar internal representations of the external world in order for syntactic language to emerge.

2.5 Questions that the Iterated Learning Model tries to Answer

In the following few paragraphs, I will discuss some of the questions that the Iterated Learning Model attempts to answer. Much of the discussion will be drawn from Kirby, Smith and Brighton (2004), which I will reference below simply as KSB.

What kinds of prior learning bias are necessary for construction and maintenance of human-like languages? KSB treat innate Universal Grammar as a specification of the initial state of a language-learning child, along with the machinery that that child will use during language learning. It is a specification of what the child brings to the language learning task (the child's *prior bias*); independent of the data they will be exposed to, and is closely related to the Chomskyan Language Acquisition Device (Chomsky 1965, 1980). The Iterated Learning Model attempts to provide a framework within which different theories of Universal Grammar might be tested, to determine what kind of innate biases are necessary at an individual level in order for human-like languages to emerge in and be maintained by a population of language learners. As mentioned above, K. Smith (2002) has made one of the most thorough investigations of this question using the iterated learning framework to date.

How language-specific must out innate biases be? KSB point out that we as language learners must have *some* kind of innately-specified biases that we bring to the language-learning task. An example is the bias toward generalisation. The ability to learn language relies on a tendency to generalise over data from many examples. They make the distinction, though, between an innate bias toward generalisation that is *general-purpose*, and one that is specific to language learning. This is related to the discussion of FLN (faculty of language in the narrow sense) vs. FLB (faculty of language in the broad sense) by Hauser, Chomsky and Fitch (2002). Iterated Learning Model experiments can help us determine what innate biases, if any, must be specific to language, rather than for general-purpose learning.

How can we explain the process that links Universal Grammar to Language Universals? KSB note that an explanation of Language Universals directly in terms of innate Universal Grammar is missing something. Because it is not immediately clear that embedding innate biases that seem to fit observed language universals in individuals would result in the emergence of those universals over time in a population, there must also be an explanation of the process by which the biases coded in innate Universal Grammar give rise to the universals traits that we observe in languages today. The Iterated Learning Model attempts to provide just such an explanation. By embedding agents with different types of innate biases into a generational model with cultural transmission, we can get a better grasp on what role those biases play in shaping the languages that emerge, and how they influence those languages.

How does language learning interact with language usage in a cultural environment? As KSB point out on page 593, "Human language is an obvious example of a behaviour that is transmitted through iterated learning." Humans learn language based on the language usage of their parents, teachers, and peers, and their language is used as learning data by their children, students, and peers. It seems sensible to seek an explanation of how these two behaviours (language learning and language usage) interact within the framework of iterated learning. The Iterated Learning Model sheds light on this interaction, and the results of many of the studies discussed above show that it is non-trivial.

What are the impacts of the mechanisms that translate back and forth between mental representations and utterances over time? Language change differs from biological evolution in the way that information is transmitted (figure 5). As discussed by KSB, in the case of biological evolution, information is transmitted through genes, the genes are carried by the organisms they help determine, and when it comes time for the genetic information to be transmitted, the replication is more or less directly from gene to gene. In the case of language change, though, information is not replicated directly from one mind to another, instead it undergoes a process of transformation in which it is converted back and forth between what Chomsky (1986) refers to as I-language (the internal representation of language, such as is stored in the brain or in an internal grammar) and E-language (the external utterances produced when people use a language). Through this process, some linguistic form is translated from internal linguistic knowledge into an external utterance by the speaker, and then translated back from the utterance into linguistic knowledge internal to the listener. The Iterated Learning Model can help us understand how this continuous translation back and forth between mental representations and utterances impacts on the languages being transmitted over time.

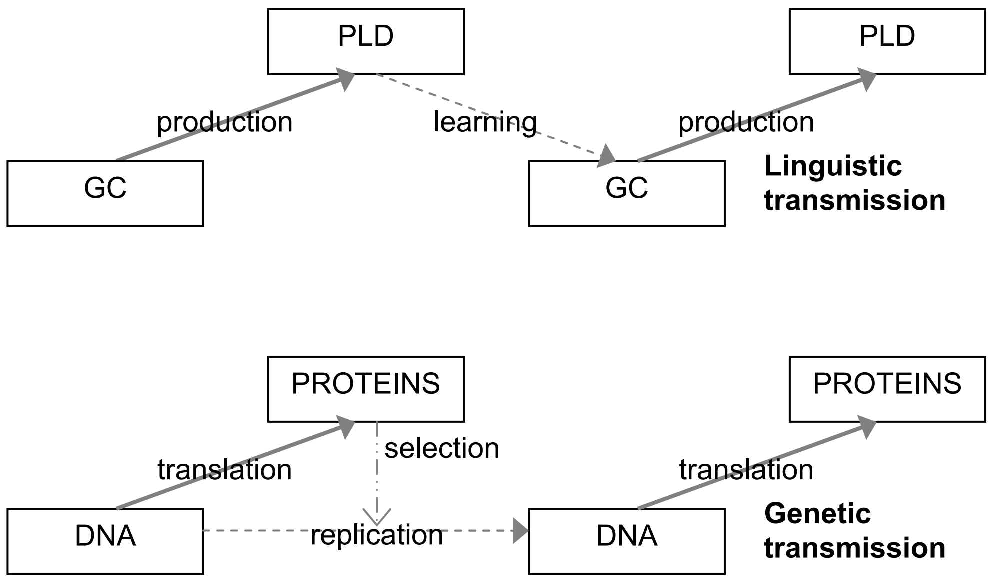


Figure 5: Similarities and differences between linguistic transmission and genetic transmission. In linguistic transmission, grammatical competence (GC) leads to the production of primary linguistic data (PLD), which is, in turn, used as learning input to shape the grammatical competence of the next generation. In genetic transmission, DNA is translated to create proteins, selection acts on the proteins that are created in order to determine what DNA is replicated to create the DNA of the next generation (reproduced from Kirby, Smith and Brighton 2004, page 602).

Can language universals arise "for free" from a process of repeated cultural transmission? The Iterated Learning Model can be used to show how a process of repeated cultural transmission can result in the emergence of language universals, eliminating the need to invoke extra innate language-specific machinery as an explanation, which often seems an unsatisfying 'hand-waving' way of explaining the universals (though by no means does this prove that such innate language-specific machinery *doesn't* exist, it puts pressure on researchers that would use it as an explanation to provide more convincing evidence of its existence). An example cited by KSB is that of Jäger (2003), who demonstrated the emergence of the case hierarchy of Aissen (2003), as a result of iterated learning in a functional Optimality Theory-based case-system variation model.

How can we explain dysfunctional yet stable aspects of language? The languages that arise in Iterated Learning Model simulations are much better for communication, from a functional viewpoint, than the initially random systems that the simulations are usually seeded with. As KSB point out, this result is a by-product of languages adapting to the problem of transmission. There are some language universals that are difficult to use, such as centre-embedded relative clauses, but persist in languages throughout the world. The process of Iterated Learning might shed some light on why apparently dysfunctional language constructs would persist from generation to generation, and perhaps why some more functional alternative does not crowd them out, but so far little research has been directed at this specific question.

How can recursive communication systems come to be? Simulations by Kirby (2002b, described above, and in more detail by KSB), and by Batali (2002, described above briefly and in more detail below) have both looked into how recursive communication systems can evolve through a process of iterated learning. The kind of recursive compositionality that emerged in these two models is a fundamental property of human language that allows a potentially infinite number of meanings to be expressed by utterances built up from a finite set of discrete constituents. One thing that is striking about the results of these experiments is that the recursive compositionality arose as a result of adaptation of the language to the learners, without any recourse to biological evolution and without any notion of biological fitness of the agents.

2.6 Modeling the Interaction between Learning and (Biological) Evolution

Though this dissertation will deal mainly with the emergence and subsequent evolution of *languages* within the iterated learning framework, it should be mentioned that there have been attempts to model the interaction between the learning strategies employed by agents in a population, and the biological evolution of those learning strategies over time. An example is a simulation studied by Batali (1994), in which recurrent networks were trained on a next-character prediction task from $a^n b^n$ languages, and the initial weights of the networks were allowed to evolve over many generations, using agents' success at the next-character prediction task as a fitness measure. The result of the simulation was the evolution of agents with initial connection weights that allowed them to learn any language in the $a^n b^n$ class.

This simulation shows how agents can evolve to learn a functional language, but doesn't address the emergence of such a language in the first place. Later work by Kirby and Hurford (1997) attempted to simulate the interaction between learning, cultural transmission, and biological evolution. The results of their simulations varied widely depending on the parameters used to control the interaction between the three systems, highlighting the fact that even in very simplified models the interactions between the different complex dynamical systems at work in language evolution cannot be ignored, if the entire process is to be understood.

In his PhD thesis, K. Smith (2003) introduced a model whereby the initial leaning biases of a population of agents (K. Smith 2002, discussed above) are allowed to evolve over time, and communicative accuracy of agents was used as a fitness function. He found that if, through the process of genetic drift, enough agents with constructor biases are introduced into the population of agents, cultural evolution of the language will begin having an effect, which in turn increases the fitness of agents with constructor biases. The result is that once a threshold proportion of agents with constructor biases is reached, the proportion will then steadily increase until it reaches a plateau level where constructor agents don't have a large advantage over agents with a maintainer bias, then drift will allow it to wander within a small distance.

3 The Exemplar-Based Learning Model

My research is based directly on an exemplar-based learning model developed by Batali (2002), and explores certain aspects and implications of his model in more detail through a replication and some extensions of his model's framework. In the sections that follow, I will give a brief description of Batali's model (though see the original paper for more detail), what are some of the major differences between this model and other iterated learning models, and what specific questions I attempted to address in my research.

3.1 (Brief) Description of the Model

Batali's (2002) model simulates the process of negotiation of a recursive grammar. Agents in the simulation are given the machinery necessary to communicate with each other and to negotiate a system of communication, but they don't start out with any shared system of communication. Agents start out with no communication system and begin the simulation "making up" random utterances in order to express meanings that the environment prompts them to express to each other. Through a process of iterated learning from each other and using the utterances that have heard from other agents, a shared, recursive system of communication is shown to emerge.

Agents store internal representations of learning observations in what Batali calls exemplars, which are made up of phrases that represent a structured (or unstructured) mapping between meanings and strings. Meanings are represented as sets of formulas that are linked by arguments, for example the meaning of the English sentence "A cat chased a dog" could be represented as {(cat 1) (chased 1 2) (dog 2)}. Phrases are structured as binary trees, in which a string-meaning pair can be broken down into a left and right sub-tree, and those sub-trees can likewise be broken down. Arguments are assigned to formulas through the use of argument maps, which may map one or more arguments of a sub-tree to different arguments (figure 6).

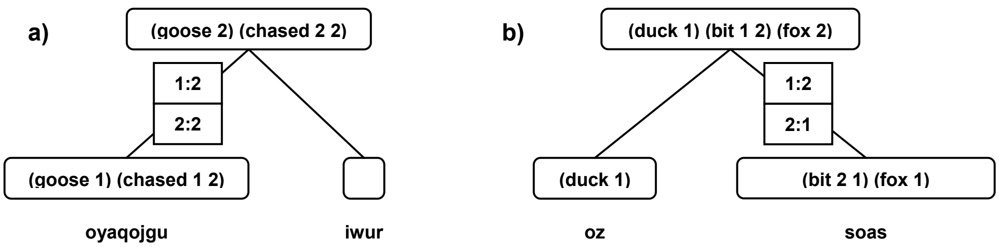


Figure 6: a) an argument map is used in a phrase to create the meaning {(goose 2) (chased 2 2)} from the meaning {(goose 1) (chased 1 2)}, effectively changing the meaning from "a goose chased something" to "a goose chased itself". b) an argument map is used to change the order of the right sub-tree's arguments, without changing the meaning of the right sub-tree itself.

A system of costs is in place to guide the agents' preferred expressions for meanings and interpretations of strings, in which agents prefer to use the lowest-cost exemplars for learning, expression, and interpretation. The cost of an exemplar stored in any agent's exemplar set is determined as a function of how often it has been successfully used during learning experiences and how often the agent has made learning observations that were inconsistent with it. When an agent is prompted by the environment to express a meaning or to interpret a string, it builds the lowest-cost phrase possible that matches the string or meaning that it was given to interpret or express, and uses the meaning or string (respectively) of the resulting phrase to guide it's interpretation or expression. The cost of one of these phrases is some function of the costs of the exemplars used to construct them, and the method that was used to put the exemplars together. Similarly, when an agent is given a string-meaning pair to learn, it constructs the lowest-cost phrase that is consistent with both the string and the meaning and stores it in its exemplar set with an initial cost of 1.0.

An agent can create phrases from its stored exemplar sets in four ways: 1) it can use a stored exemplar directly, if it is consistent with the required string and/or meaning, in which case the cost of the solution phrase is equal to the cost of the exemplar used; 2) it can create a new phrase with no meaning, or with a random string, in which case the cost of the solution phrase is equal to the sum of the number of formulas in the meaning and the number of characters in the string; 3) a sub-tree of an exemplar may be replaced by another exemplar (but only if the replacing exemplar has the same arguments as the sub-tree that is being replaced, a restriction that Batali refers to as the Replacement Condition, see figure 7a for an example), in which case the cost of the solution phrase is equal to the sum of the costs of the exemplar whose sub-tree was replaced, the exemplar that replaced it, and the replacement cost of 0.1; 4) two exemplars may be joined together in a specified order, with an argument map applied to either of the two (see figure 7b for an example), in which case the cost of the solution phrase is equal to the sum of the costs of the two exemplars that were joined together, and the new structure cost of 1.5.

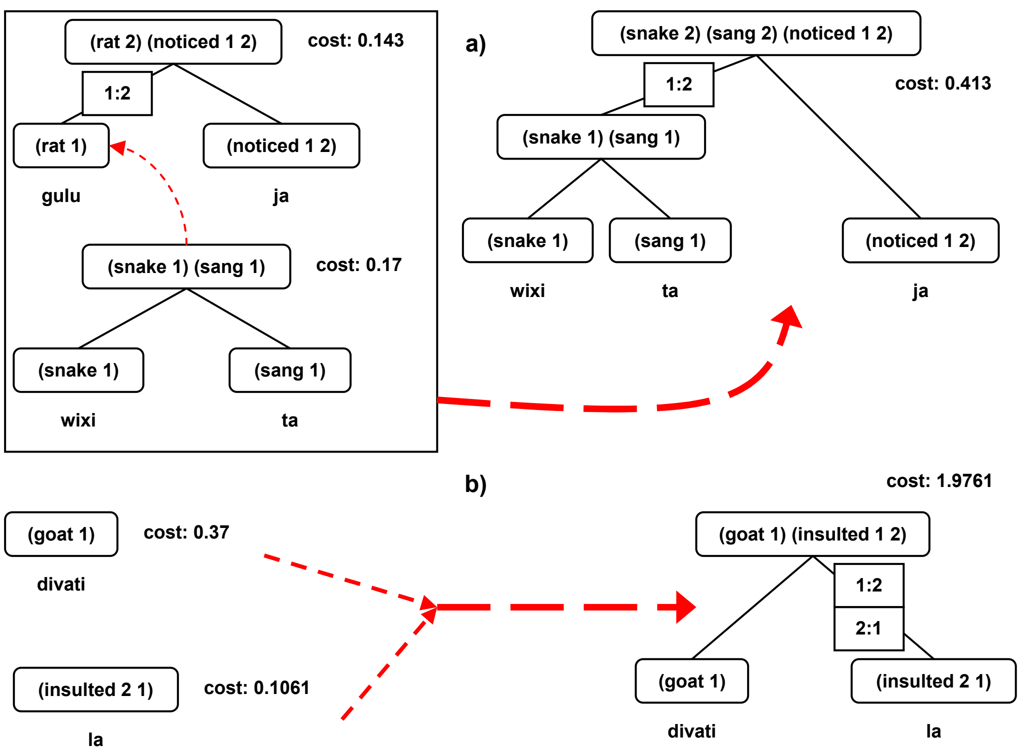


Figure 7: a) a sub-tree of an exemplar is replaced by another exemplar (that satisfies the Replacement Condition), the cost of the solution phrase is the sum of the cost of

the two exemplars being used, plus the replacement cost of 0.1. b) two exemplars are joined together in a specified order using an argument map to switch the arguments of the two-place predicate, the cost of the solution phrase is the sum of the cost of the two exemplars being used, plus the new structure cost of 1.5.

In the simulation runs that Batali describes, ten agents were used and there was no population turnover. Any exemplar that has not been used by the agent in one of the 200 previous speaking, interpreting, or learning episodes is deleted from the agent's exemplar set, as it is unlikely that it is being used actively, and to satisfy computer memory constraints. My research (as described below) shows that this pruning of little-used exemplars is not only useful for keeping computer memory usage within tolerable bounds, but that it is a key component in the emergence of structure in the languages negotiated by the agents.

3.2 No Population Turnover

This model differs from most other iterated learning models in that is has no population turnover. Tonkes (2002) calls this model and others like it (i.e. Batali 1998) *negotiation models* in order to set them apart from iterated learning models that use population turnover.

The fact that this model has no population turnover is significant because many other iterated learning models (especially those studied by Simon Kirby and others in the Language Evolution and Computation section of the University of Edinburgh Linguistics department, i.e. Kirby 2000, Kirby 2002b, Brighton and Kirby 2001, Kirby, Smith and Brighton 2002, Kirby and Hurford 2002) have shown the emergence of successfully stable, compositional languages relies on the languages being transmitted across a "bottleneck". This bottleneck has been implemented in these models by having a generation of learners learn a language or languages based on observations of utterances created by a generation of speakers expressing some subset of all the meanings expressible by their language(s).

One of the questions I had about Batali's model is whether or not the lack of population turnover has a significant effect on its predictions. In order to determine this, I added a parameter to the model that, when switched on, had two effects on the agents: 1) exemplar pruning was disabled; 2) every five hundred rounds an agent was selected at random, and its memory blanked. The blanking of one agent's memory in step (2) is equivalent to removing an agent from the population and replacing it with a new agent with no linguistic knowledge (see figure 2 above). Activating the population turnover parameter also changed the communicative accuracy tests slightly in that the oldest two agents in the population were always selected to communicate with each other for the tests. Results of this test are discussed below.

3.3 Possible Bottleneck: Exemplar Discouragement/Pruning

I propose that a bottleneck similar to the one discussed by Kirby and Hurford (2002, Kirby 2002b) is implemented in Batali's model, but that it takes a different form than the semantic bottleneck that they describe (Hurford 2002 attributes a "production bottleneck" to this model, which is similar to the one I am proposing though not exactly the same). My proposal is that a semantic bottleneck is implemented in Batali's (2002) model in the form of the process of discouragement of exemplars (i.e. increasing their cost) and pruning exemplars from an agent's exemplar set if they haven't been used during the last two hundred rounds that the agent participated in.

The reason for this is that if a simulation run has a large number of possible meanings that can be expressed, it is inevitable that meanings will go unexpressed for spaces of time long enough that Exemplars representing holistic form/meaning pairs for those meanings will be pruned out of the simulation. But, if an exemplar can be used consistently to form a useful compositional rule, its cost will remain low. This process of encouraging exemplars that embody useful compositional rules and discouraging ones that represent holistic meaning/form pairs will create a pressure for the language of the agents to become more compositional over time.

This proposal was tested by comparing simulation runs that have the standard exemplar discouragement/pruning process with ones that either prune more exemplars (say any that have not been used in the past ninety or so rounds, instead of the past two hundred) and ones that have no pruning, or not nearly as much (say pruning exemplars that haven't been used for two thousand rounds instead of two hundred), and seeing if the differences in the emerging languages are similar to the differences in languages that emerge under different bottleneck conditions in other iterated learning models (say that described in Kirby and Hurford 2002). The results from these runs are discussed below, along with a comparison with the results of the runs with population turnover. My prediction was that if exemplars are pruned too quickly, there will never be enough exemplars in a given agent's exemplar set to make the generalisations necessary to support a shared compositional language, and that if pruning is completely removed, or more exemplars are allowed to build up in the agents' exemplar sets, the result will be many exemplars representing holistic meaning/form mappings, and that a compositional language will not emerge, or will take a much longer time to emerge.

3.4 "Flat" Semantic Representation

Another difference between Batali's (2002) model, and other symbolic iterated learning models[1] is that the semantic representation used is "flat" in a sense that the semantic representations of the other models aren't. By this I mean that the kind of syntactic structure that emerges in the model is not "built into" the structure of the semantic representation used by the model. This is a point in its favour, as the syntactic structures that emerge in other models that use more hierarchically structured semantic representations can arguably be the result of a simple mapping of the semantic structure onto the agents' utterances.

An important example of this in the case of recursion is illustrated by the symbolic model described by Kirby and Hurford (2002). This model uses predicate logic-style semantic structures which can be embedded recursively, to create arbitrarily complex meanings with recursive structure, such as *knows(john, thinks(mary, loves(gavin, sue)))*. In this case, the emergence of a language with recursive structure may not be all the surprising, since any direct lexical mapping between names for predicates or the objects acted on by predicates in the semantic representation will necessarily have recursive structure.

In the case of Batali's (2002) semantic representation, though, there is no explicit embedding of formulas: any recursive relational embeddings happen only implicitly. For example, the meaning *{(snake 1) (goose 2) (sang 2) (noticed 2 1) (bit 1 3) (moose 3) (danced 3)}* contains a recursive relation, in that the *snake* is both the thing

[1]It's not clear that making the same comparison with the connectionist iterated learning models discussed above would make sense, because the languages in these models aren't recursive. It is possible, though, that such a model might be implemented in the future that uses recurrent networks in a different way than in the models discussed above.

that was *noticed* by the *goose*, as well as the thing that *bit* the *moose*. A string that expresses this meaning may or may not be recursively structured, and since meanings are treated as unordered formula sets, there is not way for the "order" of the formulas in the meaning to influence the structure of the resulting string.

3.5 Exemplar-based Learning Model

Batali's (2002) model is also unique in that it uses an exemplar-based learning model, instead of a neural-network based connectionist model, or a symbolic rule-induction model (such as the ones described in Kirby and Hurford 2002). This is important both because it allows us to test the differences that exemplar-based learning models make, as opposed to the predictions made by rule-induction models, and also because it might provide a more realistic representation of language learning than other symbolic simulation approaches.

Exemplar-based models of learning have been used since the late 1970s (Medin and Shaffer 1978), originally to explain how contextual information was used in the process of classification, and later as support for a functionalist framework (MacWhinney and Bates 1987). Also, Barsalou (1989) showed that category representation through abstraction could be explained through the use of exemplar memory. Aha, Kibler and Albert (1991) discussed various algorithms for implementing instance-based learning, Hammond (1990) developed a framework for using exemplar-based learning for planning tasks, and Bod (1998) makes an argument for exemplar-based language learning over explicit grammar induction.

More recently, Alison Wray (2005) proposed that exemplar-based learning models in which language users' intuitions are very strongly tied to their experienced input may be a better match for mental reality than models that explicitly search for systematicity in the input, such as the symbolic rule-induction models discussed above. She discusses compelling evidence in real-life language use that seems to match predictions that would be made by an exemplar-based model of language learning, especially when it comes to preferred interpretations of utterances as well as preferred utterances for expressing meanings that have been previously encountered.

Wray (2002, 2005) describes a mechanism of *needs only analysis* in which input is only broken down by an individual as much as is necessary to create or extract meaning.

> That is, there is no gratuitous analysis of form beyond the point where form-meaning mapping is sufficient for the present comprehension event, or for the construction of the presently required output. Over a period of time, an accumulation of event-specific comprehension and production requirements will lead to the identification of many small, recombinable items, plus rules for their recombination. However, large units that never require such reduction will remain intact, and this will result in a mixed inventory of small and large items, as determined by the patterns in the input (Wray 2005, page 154).

Supporting evidence for *needs only analysis* given by Wray (2005) are listed below:

- The existence of preferred interpretations for complex utterances over possible logical and grammatical interpretations (i.e. "don't count your chickens" and "the thing is").
- The existence of a preferred choice of interpretation between two logical and grammatical interpretations (i.e. native English speakers would not interpret "tear along the dotted line" as an instruction to run down a dotted divider line in the middle of a road, without substantial contextual clarification).
- The existence of preferred choices of expression, due to the relative ease of re-trieving the preferred unit directly, rather than constructing a new one (i.e. na-tive English speakers would not instruct someone to separate a piece of paper along a perforation by saying "rip along the marked pathway").
- The existence of unanalysed multi-word idioms (such as "by and large") that are never changed at the single-word level to create novel meanings (for ex-ample, there is no "by but large" or "by and small").
- The existence of partially analysed multi-word idioms (such as "from now on", "from then on", "from Tuesday on" and "from that moment on") in which some part of the phrase may be replaced to create novel meanings, but other parts of the phrase remain idiomatically fixed (for example, there is no "till now on" or "from now off").
- Findings by Bergen (2001) that native Esperanto-speaking children introduced irregularities into what was previously a perfectly regular system, which are in line with a *needs only analysis* model of learning, but not by a model in which systematic input automatically leads to generalisation.

The process described by Wray closely matches the process of learning undergone by the agents in Batali's model. Whereas in symbolic rule-induction models (i.e. Kirby 2002b) matches each learning experience against the agent's entire internalised grammar in order to generalise rules, and then throws away the old rules that are superseded by the new ones, Batali's model uses the most preferred (lowest cost) exemplars that can be used to express a given meaning or interpret a given string. This process may lead to the creation of new exemplars, but only the new exemplars that are created and the exemplars used to create them are affected by the process, there is no exhaustive search of the agent's grammar and no induction of "rules" that aren't tied to specific learning experience. Through this process, token exemplars are created, and a set of exemplars controlling how those tokens are to be recombined to create more complex meanings eventually becomes preferred (much like Wray's "recombinable items"). However, it is completely possible for large unanalysed exemplars to persist in an agent's exemplar set, if it can often be reliably used for expression, interpretation and learning. The persistence of such exemplars would be due to patterns in the agent's input, an outcome that is also described by Wray.

4 Results

In this section I describe the implementation and results of the experiments mentioned in the previous section, as well as some interesting results that came out during the course of conducting them.

4.1 Implementation Differences

One difference between my implementation[2] and the model described by Batali is that in my implementation agents are never prompted by the environment to discuss meanings involving reflexive predicates, such as *{(cow 1) (bit 1 1)}*. Though my implementation handles this correctly in some instances, there can be cases in which it wouldn't be handled correctly, due to differences in the way I determine if one formula is *equivalent to* another vs. if one formula set *contains* another, in order to avoid

[2] The source code for the implementation is available at http://www.jceddy.com/research/language_evolution/ along with discussions of continuing work.

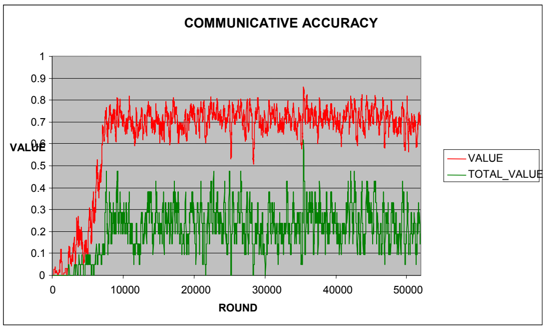
incorrect behaviour, I decided to remove reflexive predicates from the simulations completely.

Also, in most of the simulation runs discussed in this section, I used smaller potential meaning spaces than those used by Batali in the simulation runs he discusses, due to constraints on time and computing resources. The size of the potential meaning spaces was reduced in three ways: 1) Only two participants were allowed in any meaning that agents were prompted to express (Batali allowed up to three, personal communication); 2) In addition to constraint (1), the number of formulas allowed in a meaning that agents may be prompted to express was limited to five (Batali allowed up to seven, as described in Batali 2002, page 122); 3) meanings were drawn from a pool of 17 property predicates and 10 relation predicates for all simulation runs (Batali's were drawn from a pool of 22 property predicates and 10 relation predicates, as described in Batali 2002, page 122). This doesn't seem to affect the qualitative results, but due to some results that seemed inconsistent to Batali's discussed results, one simulation run has restriction 2 removed (I also ran the simulation with both restrictions 1 and 2 removed, but it has not finished running in time for submission of this dissertation, any significant differences in results will be discussed in future work). Differences in results due to differences in these restrictions will be discussed.

4.2 Basic Simulation Results

The basic simulation run, that others will be compared to, was run with the constraints described above. A discussion of the results of that run follows, as well as a description of the graphical representation of the data that will be used in the discussion of other runs.

The two graphs in figure 8 show the communicative accuracy of the agents during the simulation, as well of the average amount of different kinds of exemplars in the agents' exemplar sets. The red line in the first graph is the measure of communicative accuracy used by Batali, and based on the "precision" and "recall" values used by Kent *et al.* (1955). Communicative accuracy for a communicative episode is given by the formula $0.5*(c/s + c/r)$ where s is the number of elements in the sender's formula set, r is the number of elements in the receiver's formula set, and c is the number of formulae common to both sets. The green line in the first graph is a measure of communicative accuracy that is more like the measure used in most of the iterated learning models discussed in previous sections, where total communicative accuracy for a given utterance results in a value of 1, and anything less than total accuracy results in a value of 0 (Batali does not discuss a similar measurement). This alternate measure is given in order to illustrate the difference between merely understanding the words of an utterances and understanding the differences in meanings implied by how the words are put together. The red line in the second graph shows the average number of exemplars in the agents' exemplar sets, and the green line shows the average number of token exemplars in the agents' exemplar sets.



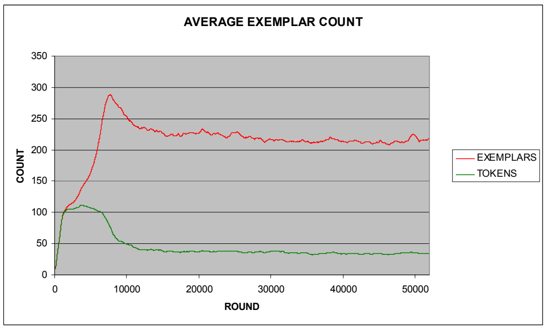


Figure 8: The communicative accuracy of the agents in the simulation rises sharply for the first 8000 rounds or so, then levels off, hovering around 0.7 for the remainder of the simulation. At the very beginning of the simulation, almost all of the exemplars in the agents' exemplar sets are tokens, until the count reaches just above 100, then the number of complex exemplars rises sharply while the number of token exemplars remains constant for a time. At about round 8000, the number of token exemplars drops as the agents negotiate an agreed-upon set of singleton tokens, and the number of complex exemplars drops and levels off. There are some complex exemplars that are agreed upon by the agents, which have low costs and remain in the simulation, other complex exemplars are created as needed and eventually pruned.

The two graphs in figure 9 show the average costs of exemplars in the agents' exemplar sets, as well as an illustration of how some types of communicative agreement is reached by the population. The red line on the first graph shows the average cost of complex exemplars in the agents' exemplar sets, the green line shows the average cost of token exemplars in the agents' exemplar sets. In the second graph, the red line shows the average singleton token ratio of the population. A token exemplar is any exemplar that maps a string directly to a meaning, with no structure linking any part of the string to any part of the meaning, whereas a singleton token has the added property that the exemplar's meaning can not be broken down any further. As the singleton token ratio goes up, it means that complex meanings are more often being expressed by complex exemplars that contain structure in the mapping from string to meaning. The green line in the second graph shows the rate at which agents either invent new token phrases for communication, or put two tokens together in a specified order to create a new complex exemplar. A low value for this measurement means that complex phrases are more often being expressed by combining existing exemplars through a process of sub-phrase replacement.

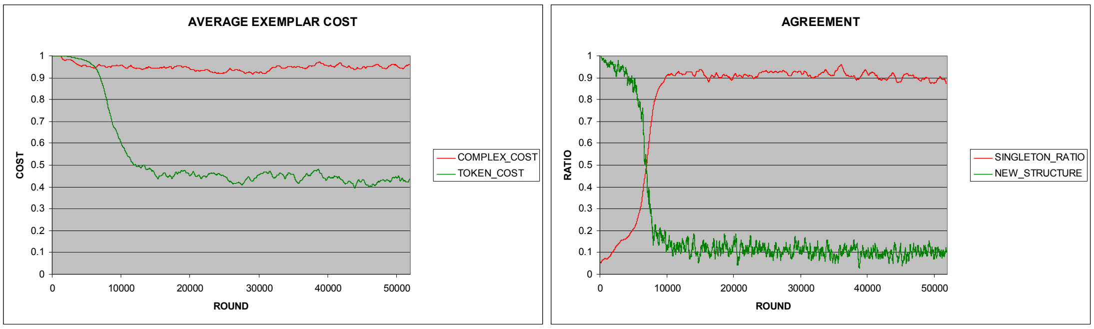


Figure 9: The cost of complex exemplars remains near 1 during the entire simulation, the cost of token exemplars declines slowly near the beginning, then drops sharply once the agents have agreed upon a shared set of singleton tokens. The average cost of token exemplars levels off just above 0.4, when the cost of agreed-upon singleton tokens nears 0. The remaining cost is due to non-singleton tokens that are continually being created and deleted by the system. The ratio of singleton tokens to non-singleton tokens rises for the first 10,000 rounds or so, and then levels off. At the same time, the rate at which new structures are created by agents during communicative rounds drops. Once the population has an agreed-upon set of singleton tokens, and each agent has a set of lower-cost complex exemplars to use for communication, the agents seldom need to invent new random string-meaning mappings, or arbitrarily decide on orderings for novel phrase structures.

The graph in figure 10 illustrates a measure of compositionality in the agents' production (this measurement is also one that I have introduced, Batali doesn't attempt to measure the compositionality of his agents' language use in a quantifiable way). The compositionality measure is determined by dividing the number of formulas at the simplest level of a phrase structure by the number of meanings in its formula set.

For example, the compositionality measure of a singleton token phrase will always be 1, since there will always be a single formula at the simplest level of the phrase structure and one meaning in the formula set, but the compositionality measure of a phrase with five formulas in its formula set could range from 0.2 (for a 5-formula token phrase) to 1.0 (where at the simplest level every formula in the phrase's formula set is mapped to a string).

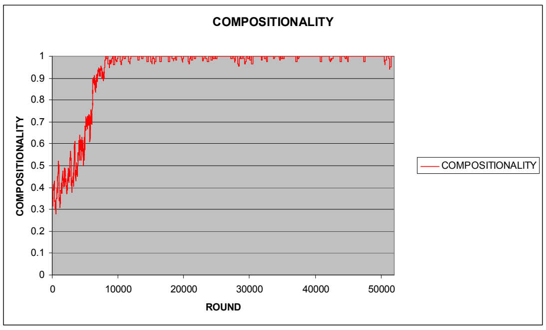


Figure 10: The compositionality measure rises quickly to 1, meaning that the agents are building up complex phrases from simpler components, instead of using holistic string-meaning mappings.

In a test of the agents' communicative ability at the end of this simulation, it was found that most of the time agents were not able to unambiguously communicate different meanings in which the only difference between the meanings was a different distribution of arguments. An example is that a speaking agent might express both the meanings *{(cat 1) (chased 1 2) (rat 2)}* and *{(cat 1) (chased 2 1) (rat 2)}* using the same string, and the receiving agent might interpret that string as *{(cat 1) (chased 1 2) (rat 2)}* or even *{(cat 1) (rat 1) (chased 1 2)}*, resulting in ambiguity in both production and reception. This looks qualitatively similar to a "protolanguage" described by Bickerton (1990). I discuss this phenomenon in more detail in the Meaning Frequency Discussion section, along with some ideas about what conditions are needed in the model to allow this kind of ambiguity to be overcome.

4.3 Population Turnover

In the simulation runs with population turnover (an extension to Batali's model), exemplar pruning was disabled. The eventual deletion of an agent takes the place of the exemplar pruning that occurs in Batali's (2002) original model; unused exemplars are deleted along with the agent in whose exemplar set they are stored. Due to limitations on computer memory, the pool of meanings that built up the complex meanings agents communicated about was limited. Meanings were built up from a pool of five property predicates and three relational predicates. In the simulation described below, once every 500 rounds an agent was selected at random, and its memory was blanked, effectively removing an agent from the population and adding a new "child" agent with no linguistic knowledge. Figure 11 below illustrates the behaviour of the generational model.

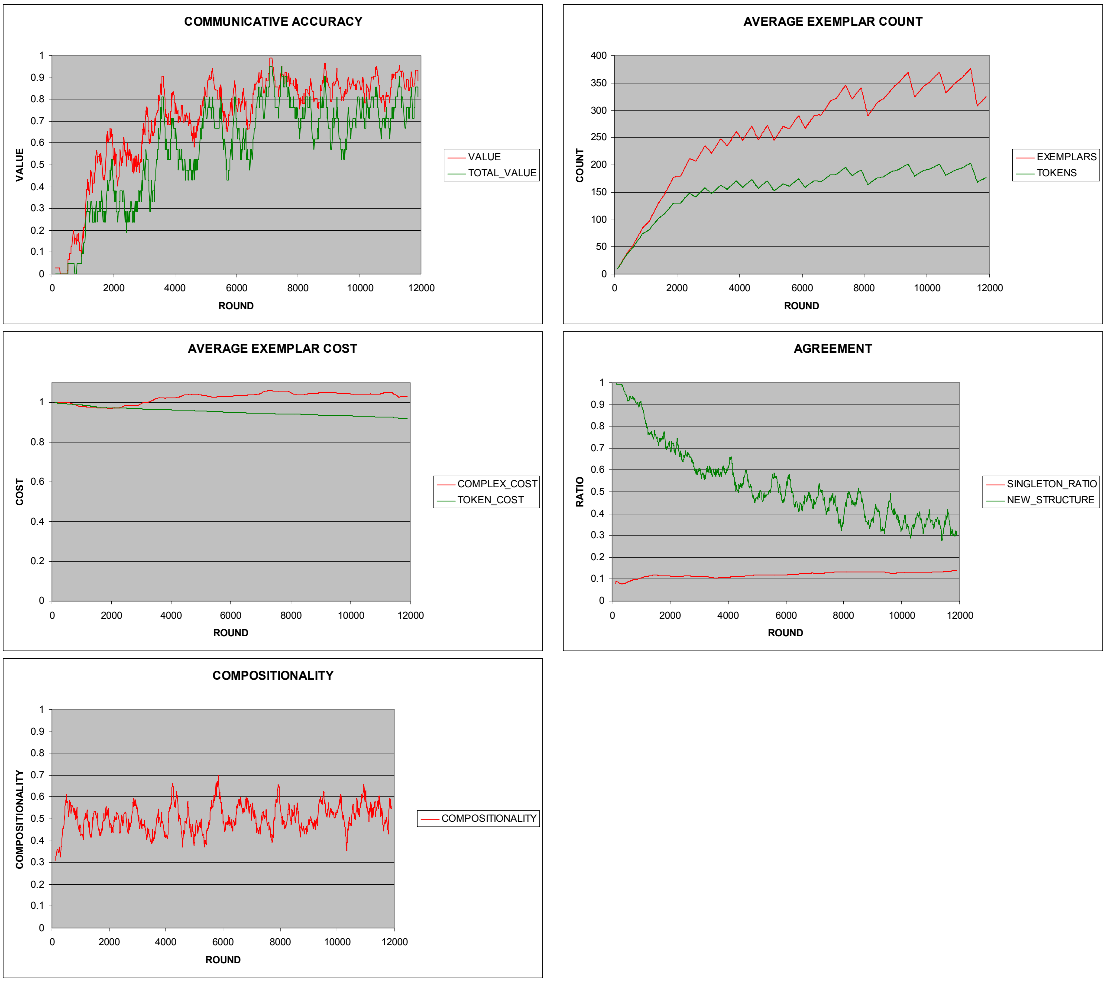


Figure 11: A simulation run with population turnover. Communicative accuracy approaches 1, the average exemplar count rises logarithmically, eventually leveling off, with discontinuities every 500 rounds when an agent is replaced. Average exemplar cost remains high, with the average complex exemplar cost remaining above 1. Singleton token ratio remains low. New structure rate continually decreases, with larger discontinuities every 500 rounds. The compositionality measure remains low, even though the system does exhibit compositional behaviour (as discussed in the main text).

As seen from the graphs above, the communicative accuracy approaches 1 quickly in this simulation. This cannot be directly compared to the communicative accuracy behaviour of the basic simulation results, because such a smaller meaning space was used, but it is comparable to the behaviour of other runs with smaller meaning spaces described below. The two most interesting phenomena that are apparent from the graphs in figure 11 are the complex exemplar cost measures and the new structure measures. At the beginning of the simulation, many holistic string-meaning mappings are created, but also some singleton token exemplars are created, which are then used in construction of some of the new phrases used to express complex meanings. As agents communicate, often these new complex exemplars are inconsistent with existing holistic expressions, and are discouraged. As the simulation progresses, though, negotiation is taking place, and some agreement is being reached on expressions for different meanings. As this agreement becomes more and more entrenched, new agents are more likely to observe the agreed-upon phrases during learning, meaning that they need to create new structure less often, hence the steady decline of the new structure rate measurement.

Taking a detailed look at how meanings were expressed in the system was interesting. Even though the compositionality measure was low (agents were using many phrases without explicit internal structure when expressing meanings or interpreting strings), a more objective observer that didn't have access to the agents' internal state might be persuaded otherwise. For example, at the end of the simulation run, one agent in the population expressed the meaning *{(duck 1) (slapped 1 2) (goose 2)}* using the string "anefezif". The phrase that was used to express the meaning contained internal structure that mapped the string "an" to the meaning *{(goose 1)}*, the string "ef" to the meaning *{(duck 1)}* and the string "ezif" to the meaning *{(slapped 2 1)}* which were put together to form the required meaning. The receiving agent correctly interpreted the string "anefezif" to mean *{(duck 1) (goose 2) (slapped 1 2)}*, but the phrase it used to interpret the string was a token exemplar that mapped the string directly to the meaning with no explicit internal structure (this is similar to structure maintained in rote-learned meaning-string pairs in Hurford, 2000). This phenomenon arises because once older agents have negotiated consistent compositional phrases for expressing meanings, they will use their complex, compositional exemplars to teach the meaning-string mappings to the younger agents. If the younger agents do not yet possess exemplars that can be combined to match the learning observations, they will create token exemplars containing what is internally stored as a holistic mapping from string to meaning, but might be viewed by an objective observer as containing compositional structure. There is a chance of similar phrases appearing in the original model which lacks population turnover, but it is *much* less likely, given that in the original model the survival rate of exemplars depends partly on agents being able to recombine them with others, allowing them to be used more often to express different meanings: exemplars with complex structure can be used to express meanings that correspond exactly to one they express, meanings that contain the one they express, as well as meanings that share something (but not everything) in common with the one they express; token exemplars cannot be used to do the latter, which very much reduces their survival rate in the original model.

4.4 Size of Bottleneck

(Note that the simulations in this section do not use population turnover.) First, we will look at what happened when exemplars that had not been used for the past ninety-five communicative episodes were pruned (a number of simulations were run with an exemplar "time to live" of about this value, with similar results, it seems that there is a cutoff around a time to live of 100 rounds; simulations where exemplars were pruned after remaining unused for 100 rounds behaved in much the same way as the basic simulation). Unsurprisingly, we see that the agents fail to negotiate a system of communication that is even remotely reliable (figure 12). The communicative accuracy of the agents remains below 30% for the entire course of the simulation, the agents fail to reach agreement on a small set of shared singleton tokens, with the token count hovering around fifty for most of the simulation, of which about 20% are singleton tokens, exemplar cost remains high (above 0.8 for both complex and token exemplars) for the duration of the simulation, new token exemplars are constantly being created and randomly re-combined as needed, and the agents' utterances remain for the most part non-compositional.

The reason for this is that no exemplar is ever reinforced to the point that it will be consistently used for constructing phrases before it is pruned. During the course of the simulation some singleton tokens were reinforced often enough for their costs to drop below 0.5, but they were consistently lost from the simulation before becoming much cheaper.

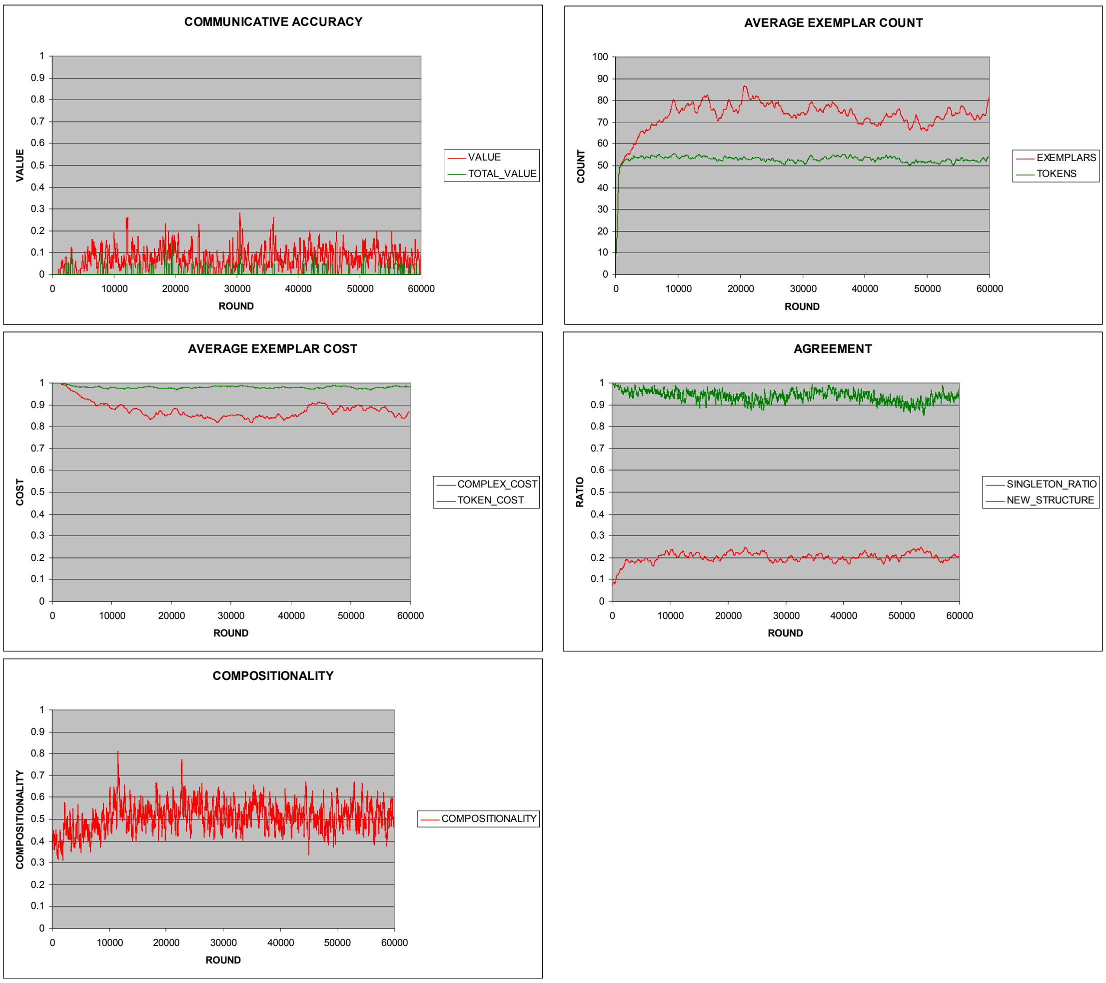


Figure 12: Graphical representation of the behaviour of a simulation in which exemplars were pruned after not having been used in the agent's past ninety-five communicative episodes. Communicative accuracy remains below 30% for the duration of the simulation, exemplar sets average a little more than fifty token exemplars, about 20% of which are singletons. Exemplar costs stay relatively high for the entire simulation, and the compositionality measure remains low. The high new structure rate shows that agents are continually creating new token exemplars and putting them together randomly as needed.

Figures 13 and 14, below, depict a typical communicative episode between two of the agents in the simulation.

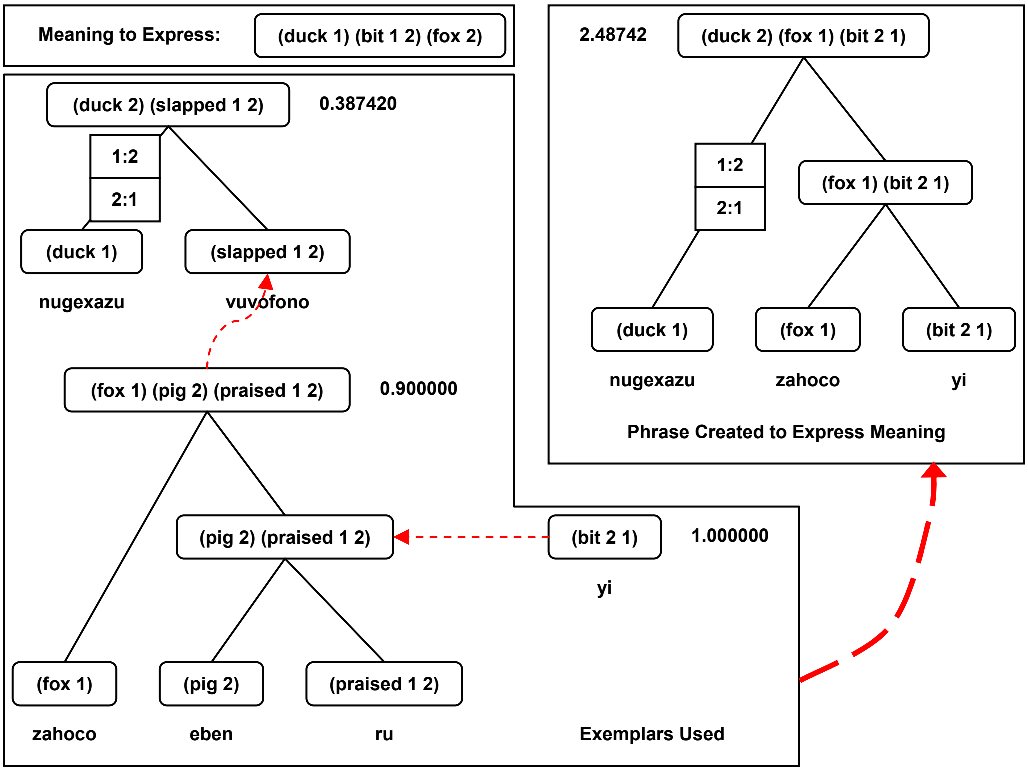


Figure 13: The sending agent is prompted by the environment to communicate the meaning *{(duck 1) (bit 1 2) (fox 2)}* to another agent, and puts three exemplars together to do it using the string "nugexazuzahocoyi", showing the partial compositionality of the system.

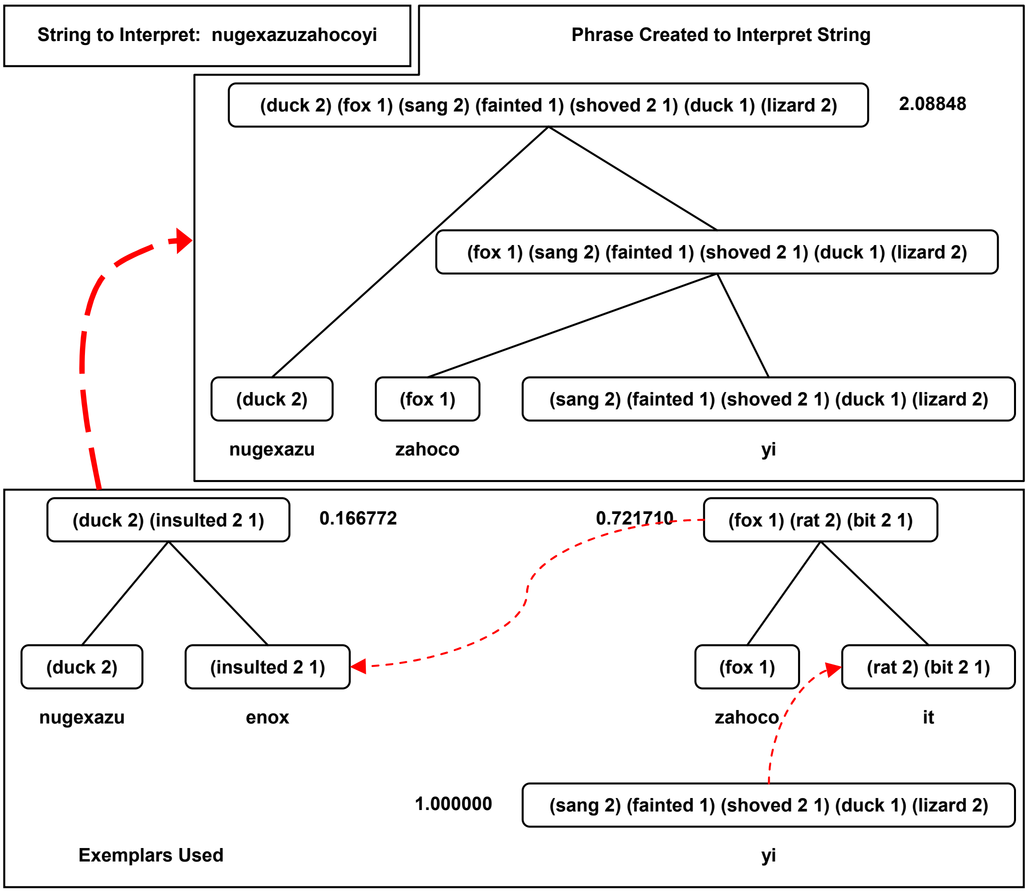


Figure 14: The receiving agent gets the string "nugexazuzahocoyi" to interpret, and uses three existing exemplars to do it, interpreting the meaning as *{(duck 2) (fox 1) (sang 2) (fainted 1) (shoved 2 1) (duck 1) (lizard 2)}*, demonstrating again partial compositionality along with unanalysed holistic expressions for some complex meanings in the system.

This communicative episode demonstrates the failure of the agents to negotiate a shared compositional system, and also illustrates how the learning biases of the agents drive the system towards a partial compositionality combined with large holistic chunks. The sending agent's exemplar that has the meaning *{(duck 2) (slapped 1 2)}* has been used consistently in many communicative episodes by the agent, and its resulting low cost allows it to be used to express meanings and interpret strings, but the other exemplars used have not. The receiver agent also has one low-cost exemplar (the one with meaning *{(duck 2) (insulted 2 1)}*) which has been used consistently by the agent in many communicative episodes, and can be used to build new phrases to express meanings and interpret strings, but must resort to combining it with higher-cost more holistic exemplars in order to interpret the string.

In this example, the low-cost exemplars being used by the agents have been able to be used consistently in many recent rounds, but that is due mostly to random variation in the meanings that the agents are being prompted to communicate about. Low-cost exemplars such as these occur often during the course of the simulation, but are usually eventually lost due to under-representation of the meanings covered by the exemplars in the randomly-generated meanings discussed by the agents. They are then replaced by newly-created phrases (often large holistic ones like the sender's exemplar with the string "yi" in the example above). This cycle of exemplar loss and re-creation results in a very unstable system of communication with newly created holistic chunks being constantly created.

Now we look at the case where exemplars are never pruned. The absence of exemplar pruning does create a problem when it comes to computer memory limitations, so the only way I was able to successfully run simulations in which pruning was completely disabled was to reduce the space of meanings that it was possible for agents to talk about. In the simulation described below, meanings were chosen from a pool consisting of five property predicates, and three relational predicates. Figure 15 shows a graphical representation of the behaviour of the model when exemplar pruning is removed completely.

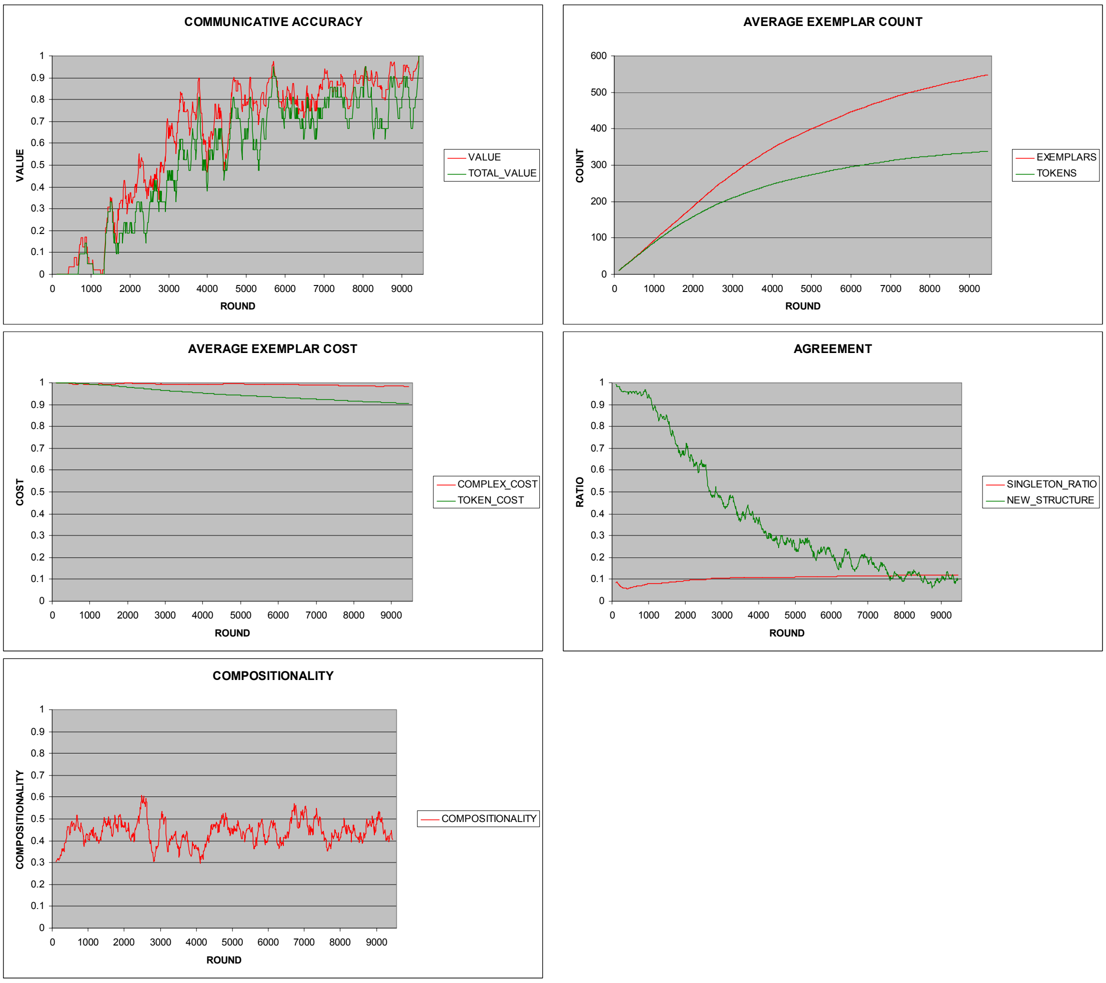


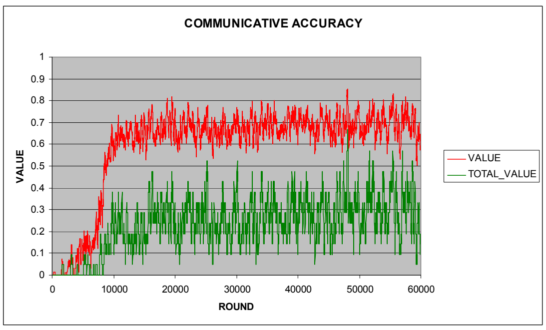
Figure 15: Communicative accuracy (by both measures) rises quickly to 1, as the agents negotiate agreed-upon ways of expressing most of the possible meanings in the simulation. The average exemplar count grows logarithmically, as agents build collections of consistent exemplars that cover most of the meaning space. The average token exemplar count rises more slowly than the average total exemplar count, implying that there is some structure in the system. The average exemplar cost decreases slowly. Since singleton tokens are not being used often for the expression of complex meanings, their cost does not plummet quickly as it does in the basic simulation. The new structure rate quickly falls toward zero, since once an agent has an exemplar covering a meaning, it is never lost. The average singleton token ratio rises slowly, as there is much less pressure to create complex exemplars to cover complex meanings. The compositionality measure hovers around 0.45; compositionality has not emerged in the system.

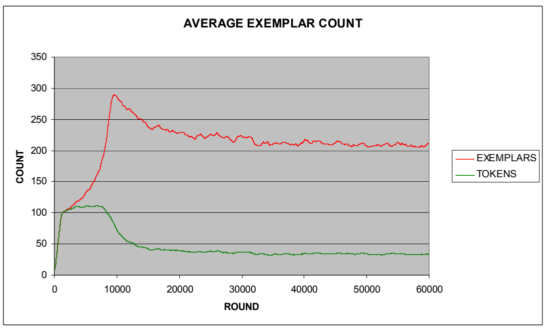
In the simulation runs where exemplar pruning was disabled completely, the languages that arose remained mostly holistic, though some compositionality did arise. For instance, in the simulation run described above, the population had agreed upon a mapping between the meaning *{(pig 1) (insulted 1 2) (goose 2)}* and the string "koqo". To express the meaning *{(pig 1) (insulted 2 1) (goose 2)}*, the string "om" was used, and agents had unanalysed mappings in their exemplar sets for both, with relatively low associated costs (mostly less than 0.5). Most three-predicate meanings followed this pattern, but the meaning *{(goose 1) (duck 2) (insulted 1 2)}* was expressed using the string "teyoza", which was built up from an exemplar mapping the meaning *{(goose 2)}* to the string "teyo" and the meaning *{(duck 1) (tickled 2 1)}* to the string "za". Some agents in the simulation had an exemplar that reflects this partial structure, but most of them had a token exemplar mapping the three-predicate meaning directly to the string "teyoza", illustrating the fact that agents did not break it down if they didn't need to.

The farther we get into the simulation, the more chance there is for structure to emerge. If an agent has a set of exemplars that can be put together to cover a meaning, but no exemplar that covers it alone, it is more likely to combine the existing exemplars to create a new one than to invent a new token with a random string. Since exemplars are never deleted, complex meanings that are encountered early on in the simulation are likely to be represented by unanalysed token exemplars throughout the entire simulation, whereas complex meanings that show up later in the simulation are more likely to be expressed using more structured phrases.

4.5 Meaning Complexity Effects

Figure 16 below shows the results of a simulation run in which restriction (2) was removed: meanings that agents were prompted to communicate with each other could have up to seven formulas.





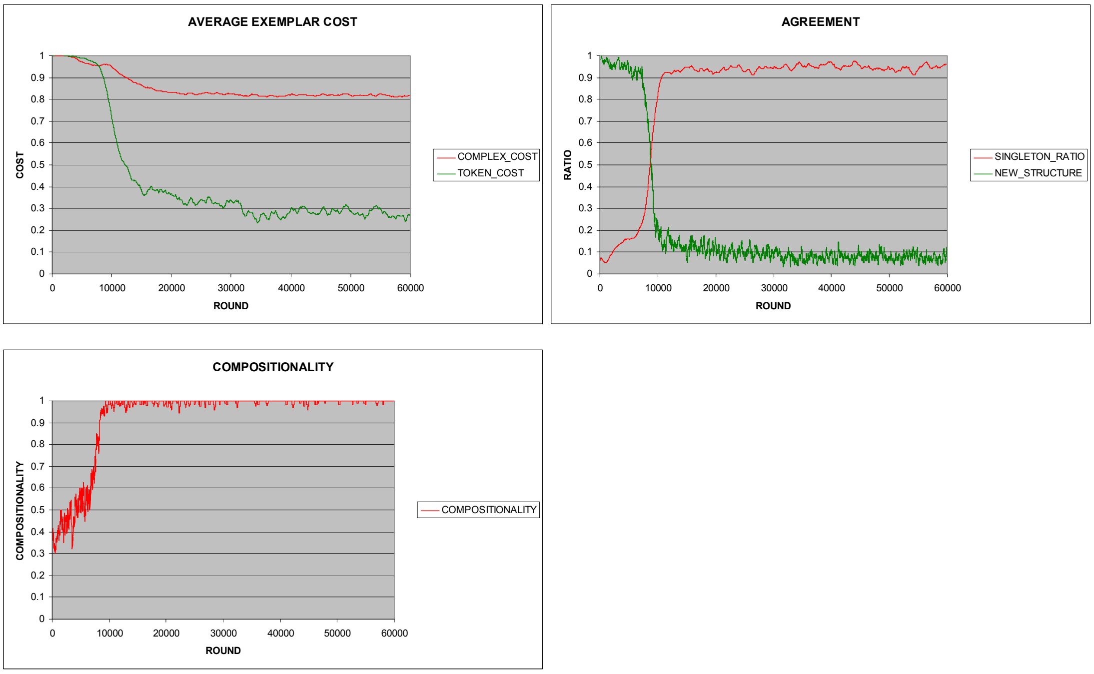


Figure 16: Restriction (2) is relaxed: communicative accuracy gets relatively high, the average exemplar count shows the effects of the meaning bottleneck, and the eventual negotiation of a shared set of singleton token exemplars, the new structure rate falls dramatically in the rounds leading up to round 10,000, and the compositionality measure get close to and stays near or at 1 at about the 10,000th round.

As we can see, the qualitative behaviour of the system with a greater range of meaning configurations is similar to that of the basic simulation. When we look in more detail at some of the communications between agents in the simulation illustrated in figures 17 through 20, we see more similar behaviour. Figures 17 and 18 illustrate two communicative episodes in which the only difference in the meanings to be communicated is that the relational formula has its arguments switched, figures 19 and 20 show a communicative episode where accurate communication was not achieved.

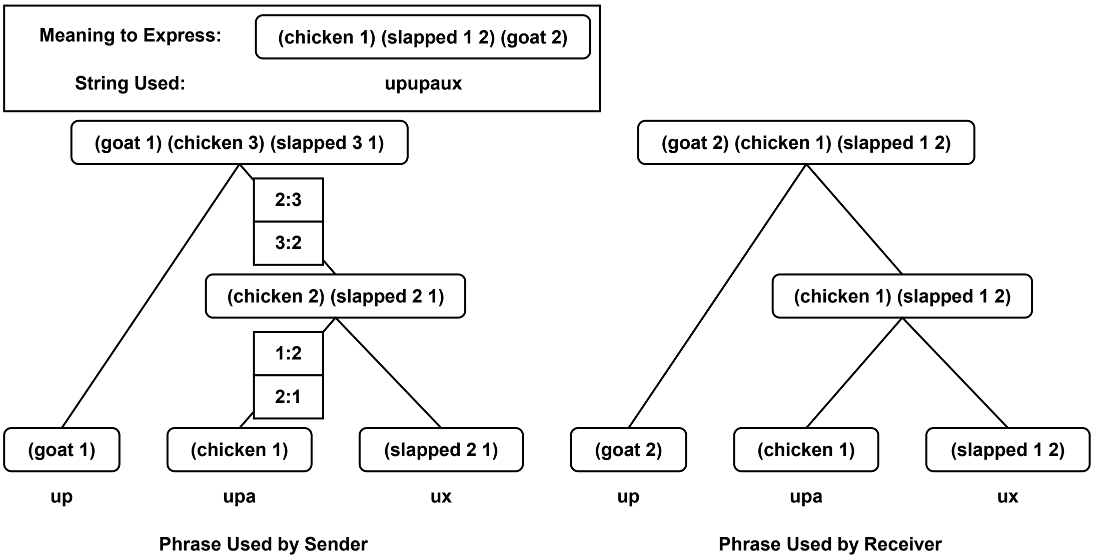


Figure 17: The sender builds a phrase to express the meaning that is has been prompted to express, and passes it to the receiver, who interprets it correctly.

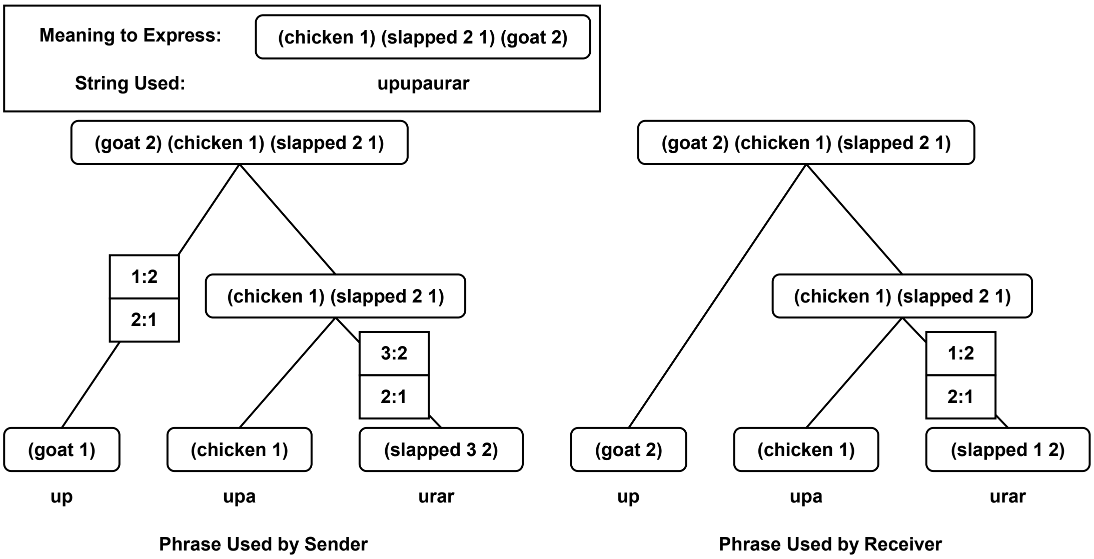


Figure 18: The sender agent is prompted to express a meaning that is identical to the one in figure 12, except that the arguments in the relational formula are switched. The agents accurately communicate this meaning as well, using a different word for the relational formula than in the previous communicative episode.

In this case, the sending agent's preferred expression for the "slapped" relational formula on its own is given by a token exemplar that maps the meaning *{(slapped 3 2)}* to the string "urar", but there are a few low-cost complex exemplars in which a sub-phrase maps the meaning *{(slapped 2 1)}* to the string "ux". The receiver agent's exemplar set shows the same phenomenon, except that in its case, the strings "urar" and "ux" are both mapped to the meaning *{(slapped 2 1)}*. These exemplars can co-exist in the model and achieve low costs because the exemplars that use the former formula are consistent with the exemplars that use the latter; the phrases used by the sender in figures 19 and 20 are an example of this.

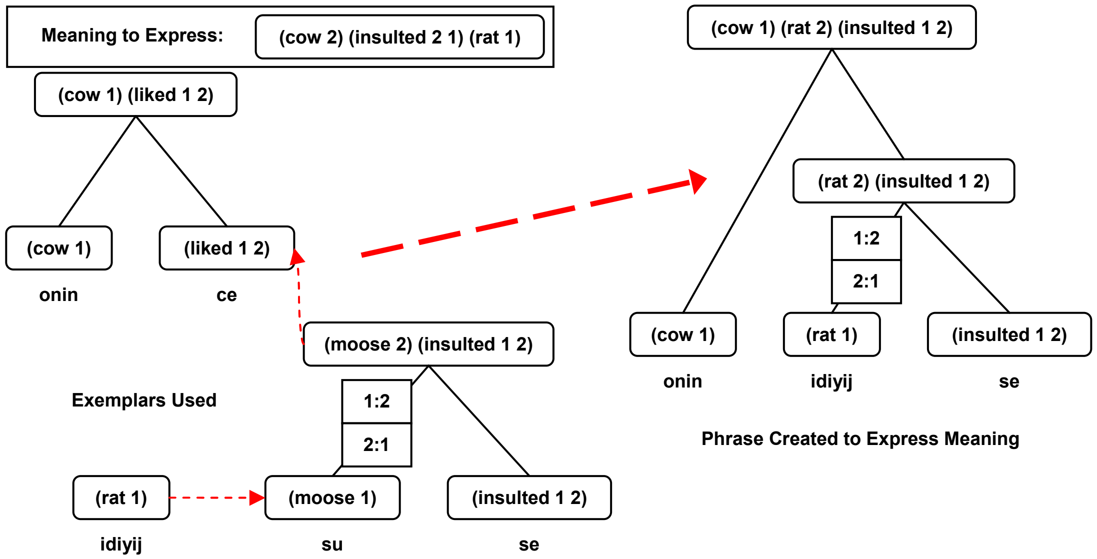


Figure 19: Three exemplars are used to create a phrase that maps the meaning *{(cow 2) (insulted 2 1) (rat 1)}* to the string "oninidiyijse".

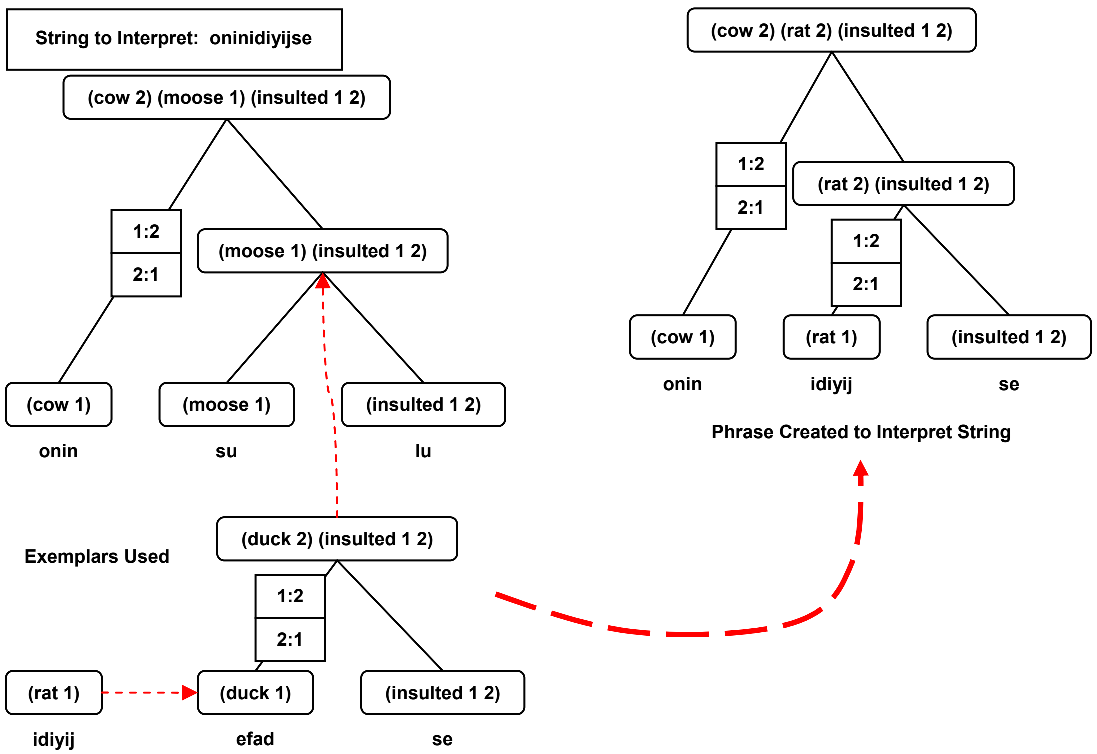


Figure 20: The receiver agent interprets the sting "oninidiyijse" as meaning *{(cow 2) (rat 2) (insulted 1 2)}*, illustrating remaining ambiguity in the system.

When I first did this simulation run, I thought that there may have been an effect of the complexity of the meanings the agents are exposed to during learning on the complexity of the exemplars that the agent's negotiate non-ambiguous phrases for. This is because a more complex meaning with seven predicates allowed is more likely to use a larger number of exemplars to express during any given round, meaning that a larger number of exemplars remain in the agents' exemplar sets, giving the agents more of a chance to negotiate shared non-ambiguous phrases for dealing with the constituent meanings that make up the more complex meanings.

The results of the run, on the whole, seem to mean that this isn't strictly the case, at least not the way I thought it would be, but the fact that the agents in the more complex simulation *did* negotiate an unambiguous method for dealing phrases using the "slapped" relational predicate, and that this negotiation, in part, was due to the usage of three variables instead of two, I wondered if allowing more participants in the meanings discussed would have more of an effect on successful negotiation of an unambiguous grammar than meaning complexity. To test this, I ran two much smaller simulations: one in which a small subset of meanings consisting only of two participants and a single relational predicate using the variables 1 and 2 were allowed, and then one in which the variables 1, 2 and 3 were randomly chosen for the different argument positions.

The results of the smaller simulations did not support the hypothesis. The only effect of introducing more variables into the simulation was to lower the communicative accuracy that was eventually achieved by the agents. Due to the Replacement Condition, the addition of more variables effectively increased the number of exemplars needed to effectively cover what amounted to precisely the same set of meanings. The precise conditions that lead to the behaviour illustrated in figures 19 and 20, as opposed to that in figures 17 and 18, have not yet been completely understood, and will have to be more closely examined in future work.

4.6 Findings

Four main findings come out of the results described in this section:

- *Population Turnover* – Population turnover can replace exemplar discouragement/pruning in Batali's model. Under population turnover systems of communication emerge that have compositional behaviour, but rote-learning is more important and apparent; agents don't break down meanings to create structured phrases because there is no need to.
- *Small Bottleneck* – If the production bottleneck on the agents is below some threshold, exemplars are continually lost before they can be used to negotiate a shared set of consistent meaning-string mappings in the population, resulting in low communicative accuracy for the entire course of the simulation.
- *Large Bottleneck* – If the production bottleneck on the agents is removed completely, the emerging system of communication remains largely holistic, since holistic meaning-string mappings invented and used for learning early on in the simulation are never lost.
- *Meaning Complexity* – Allowing more complexity in meanings has an effect on the behaviour of the simulation, especially in that it seems to allow more flexibility in the strategies agents end up using to disambiguate meaning-string mappings. Exactly what differing conditions lead to this, though, are not completely clear.

5 Conclusions and Discussion

Batali's (2002) model is different from the symbolic model explored by Kirby (2002b) in many ways, and though there are differences in the behaviour of the two models (i.e. the agents in Kirby's model never forget generalisations they have made, whereas the agents in Batali's model the persistence of generalisations in any agent's exemplar set is dependent on meanings that utilise the generalisation being communicated with some frequency), there are also some similarities. One similarity is the loss of holistic mappings from strings to complex meanings in mature agents. In Kirby's model, those holistic mappings are lost due to being thrown away outright when a generalisation that covers it is made, in Batali's model, they are "forgotten" since they are unlikely to be usable for successful communication every two-hundred rounds.

In the following sub-sections, I will discuss the results pertaining to specific questions that my research attempted to address, as well as other issues I discovered during the course of the research.

5.1 Population Turnover

The introduction of population turnover into the model, along with the removal of exemplar pruning, shows that exemplar pruning does the job in this model that population turnover does in other iterated learning models. The direct pruning of exemplars, as described by Batali (2002), serves the same purpose as population turnover, causing inconsistent exemplars to be removed from the system, and creating a pressure toward compositionality that causes the language to change over time. It is interesting in this model, though, that population turnover allows the agents to converge on a system of communication that is compositional to an objective observer, but that the agents don't necessarily store the structure explicitly in their internal representations. In linguistic terms, this matches the *needs only analysis* model in which parts of language may remain idiomatic if the language user has no need to analyse it further.

5.2 Size of Bottleneck

The discouragement and pruning of exemplars creates a bottleneck effect similar to the one described by Kirby (2002b). If exemplars are pruned too quickly, generalisations are not often made and the communication systems used by the agents remain unstable and for the most part non-compositional. If exemplars are seldom pruned, holistic expressions will by chance be reinforced much more often and will remain in the system. Although agents in those systems reach a high level of communicative accuracy by the end of the simulation run, the systems of communication that emerge are much less compositional than those that emerge under ideal conditions of discouragement and pruning.

Exemplar encouragement and discouragement create a bias toward compositionality in the model, as discussed by K. Smith (2003). Exemplar discouragement amounts to a pressure against homonymy, penalising exemplars that can be used to cheaply interpret strings in incompatible ways. Exemplar encouragement rewards consistent exemplars, and exemplars that can be used consistently in learning, expression and interpretation, and lowers their cost, which in turn makes them more likely to be used in compositional utterances in future communicative episodes. This is only part of the story, though. Encouragement and discouragement, along with the frequency with which different meanings arise, determine which exemplars get pruned, but the pruning process itself creates a bottleneck, by forcing out exemplars that holistically map strings to complex meanings, if those meanings do not arise often in communication. If these non-compositional exemplars are not pruned out, they can often be the cheapest way of expressing the meaning they cover, and if they are used often enough in the simulation, will become the preferred phrases to use. This is clear in the cost scheme: if an exemplar that covers a five-formula meaning achieves a cost lower than 0.2, than the agent will never choose to combine more than two exemplars to express that meaning, because the cost will necessarily be at least 0.2.

5.3 Meaning Complexity Effects

The types of meanings discussed by the agents in this model, their frequency and distribution, has quite a large effect on the behaviours of the communication systems that eventually emerge. It seems plausible that the use of more complex meanings in a population would lead to the negotiation of strategies for communicating those meanings unambiguously, but the results of my simulations support the conclusion that having more different ways of expressing things leads to more accurate communication. That is, the introduction of synonymy into the model actually helps it achieve higher communication accuracy, and a more compositional system. This is at odds with the results of work done by K. Smith (2002), and is something that will need to be resolved in the future.

5.4 Recursion

Structural recursion is achieved in the model through the computationally recursive process of combining existing exemplars to create new ones to cover novel meanings and strings (it should be noted that the structural recursion that arises in Batali's 2002 model is strictly in reference to recursive relations like "the dog chased the cat the bit the monkey" which is different than the structural recursion that arises in Kirby's 2002b model, which is in reference to recursively-embedded predicates such as "John knows that Susie loves Mike"). Chomsky (1980) describes the structurally recursive nature of many natural language constructs, describing their construction in terms of recursive rules in which phrases of one type can contain constituents of the same type. In this model a similar concept of "constituent type" does not exist, but operations used in exemplar construction do result in a kind of type separation due to the arguments used by a given phrase. Since phrases can only be used to replace sub-phrases in a tree if the replacing phrase covers the same variables as the sub-phrase being replaced, the model does seem to distinguish between "1-phrases", "2-phrases", and "1+2-phrases", and a phrase of any given type might be made up of sub-phrases of any other given type, which can be replaced without changing the type of the phrase as a whole. Though this is much looser than Chomsky's rules of construction, it can also result in systems that have many of the same structural properties.

Hauser, Chomsky and Fitch (2002, Hauser 2001) posit computational recursion as a uniquely human language-specific adaptation, perhaps the only component of what they term the faculty of language in the narrow sense, or FLN. Since this model only deals with a language-specific domain, it has nothing to say about whether or not computational recursion is a language-specific adaptation, but the compositional behaviour of the model is wholly dependent on the computationally recursive process of replacing sub-phrases with other phrases. If this model and others like it are psychologically realistic depictions of how humans make generalisations through learning, then some kind of computational recursion is necessarily available in the human brain.

6 Future Work

The research done for this dissertation has opened the door toward exciting future research. I would like to continue improving my implementation of the model in order to be able to loosen some of the restrictions I made in the current implementation due to issues of time and computational resources. Some aspects of the model that I would like to look at in more detail in future work are geographical distribution, horizontal vs. vertical transmission, meaning frequency effects, population effects, other bottleneck effects, an introduction of an overt anti-synonymy bias, what conditions result in ambiguous versus unambiguous exemplars for expression and interpretation, and a possible extension to the model to allow negotiation of syntactic/semantic categories along with lexical mappings. There are many other ideas I'd like to explore in using this model's framework, too many to list here, but these are some of the most interesting.

6.1 Geographical Distribution

In the original model, agents are chosen at random to communicate with each other as speakers and hearers, or teachers and learners. I would like to expand the model to look at the effects of geographical distribution, where the probability of two agents being selected for a communication round depends on where they reside in some kind of geographical space.

6.2 Horizontal vs. Vertical Transmission

Batali's model uses strictly horizontal transmission between agents in a fixed population. I would like to expand the model to study the effects of changing to a vertical transmission system on the model, and also to allow "tuning" between different levels of vertical and horizontal transmission.

6.3 Meaning Frequency Effects

The effects of the frequency of complex meanings on the outcomes of simulations using this model are discussed above, except for differences between the runs described in this dissertation and a run whose meaning types and frequencies match more closely to those describe by Batali, which I would like discuss in future work. I would also like to look into the effects of making specific meanings more or less frequent in the model than others, especially when it comes to issues of regularity and irregularity in the emergent communication systems. Hopefully, the results of such a study could be compared with those found by Kirby (2001).

6.4 Population Effects

It was observed, though not yet explored in detail, that this model might demonstrate a population memory effect. That is, having more agents around might qualitatively affect the behaviour of the simulations by allowing more exemplars to be stored in the environment of the simulation through distribution of exemplars between the agents. I would like to look at this in more detail and see if such a population effect can be found, and if so, conditions are necessary for it to arise.

6.5 Other Bottleneck Effects

In this dissertation I look at the effect of discouraging and pruning of exemplars as a bottleneck effect on learning and production in the model. One thing that I would like to explore in the future is how the model would behave if exemplar encouragement and discouragement were not there to help determine which exemplars are pruned during any given round. That is, give exemplars a constant cost, with associated costs for creating and re-combining exemplars during learning, expression and interpretation, and prune exemplars based solely on their usage in the system. Though this would remove a direct bias against homonymy from the system, and reduce the pressure toward compositionality, it still seems that the costs associated with invention and the different kinds of recombination (it would still be cheaper by 0.5, for example, to combine two exemplars with costs of 1.0 to express a two-predicate meaning than to create a new exemplar with two formulas and a two-character string, and two-character strings are the shortest that can be randomly "invented" in this model) could still drive the agents toward a compositional system of communication.

6.6 Anti-Synonymy Bias

K. Smith (2002) describes the need for biases against homonymy and also against synonymy in agents of a population in order for it to be viable as a population of constructors. The model described in this dissertation has an overt bias against homonymy which is implemented by the discouragement of exemplars who would interpret a string used in a learning round differently than as the meaning that it is provided in the learning input, but it has no overt bias against synonymy; in fact, the introduction of synonymy into the system seems to aid in the construction of a successful system of communication. I would like to add a component to the learning algorithm that similarly penalises exemplars that would express the meaning used in a learning round using a different string than the one provided in the learning input, and see if this allows the agents to negotiate a robust system without resorting to synonymy, and would also like to work at resolving my findings with those of K. Smith.

6.7 Conditions for Different Behaviour

As discussed in the results section, the difference between conditions under which agents negotiate exemplars that can be used to unambiguously express and interpret a meaning and conditions under which agents fail to negotiate such a set of exemplars, is not yet well understood. I would like to pinpoint these conditions at some point in the future, as that will be a key to understanding why the model behaves as it does.

6.8 Negotiation of Syntactic/Semantic Categories

I have been thinking of ways to extend the semantic representation of the existing model in order to allow more richly defined information to be stored, and to allow agents some leeway in negotiating categories and/or semantic types for the different meanings to be expressed, based on how they are used. It is not well thought-out as of yet, but is definitely exciting work for the future.

References:

Aha, D. W., Kibler, D., and Albert, M. K. (1991). Instance-based learning algorithms. *Machine Learning,* 6:37-66.

Aissen, J. (2003). Differential object marking: iconicity vs. economy. *Natural Language and Linguistic Theory,* 21:435-483.

Barsalou, L. W. (1989). On the indistinguishability of exemplar memory and abstraction in category representation. In T. K. Srull and R. S. Wyer (Eds.), *Advances in Social Cognition,* (pp. 61-88). Hillsdale, NJ, USA: Lawrence Erlbaum Associates.

Batali, J. (1994). Innate biases and critical periods: Combining evolution and learning in the acquisition of syntax. In R. Brooks and P. Maes (Eds.), *Artificial Life 4: Proceedings of the Fourth International Workshop on the Synthesis and Simulation of Living Systems*, (pp. 160-171). Redwood City, CA, USA: Addison-Wesley.

Batali, J. (1998). Computational simulations of the emergence of grammar. In J. R. Hurford, M Studdert-Kennedy and C. Knight (Eds.), *Approached to the evolution of language: Social and cognitive bases* (pp. 405-426). Cambridge, UK: Cambridge University Press.

Batali, J. (2002). The negotiation and acquisition of recursive grammars as a result of competition among exemplars. In E. Briscoe (Ed.), *Linguistic Evolution through Language Acquisition: Formal and Computational Models* (pp. 111-172). Cambridge, UK: Cambridge University Press.

Belpaeme, T. (2001). Simulating the Formation of Color Categories. In B. Nebel (Ed.), *Proceedings of the International Joint Conference on Artificial Intelligence 2001(IJCAI'01)* (pp. 393-398). Seattle, WA, USA: Morgan Kauffman.

Bergen, B. K. (2001). Nativization processes in L1 Esperanto. *Journal of Child Language*, 28:575-595.

Bickerton, D. (1990). *Language and species*. Chicago, IL, USA: University of Chicago Press.

Bod, R. (1998). *Beyond Grammar: An Experience-Based Theory of Language.* Chicago, IL, USA: CSLI Publications, Center for the Study of Language and Information, University of Chicago Press.

Brighton, H. (2002). Compositional syntax from cultural transmission. *Artificial Life,* 8:25-54.

Brighton, H. and Kirby, S. (2001). The survival of the smallest: Stability conditions for the cultural evolution of compositional language. In J. Kelemen and P. Sosik (Eds.), *Advances in Artificial Life (Proceedings of the 6th European Conference on Artificial Life)* (pp. 592-601). Heidelberg: Springer-Verlag.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA, USA: MIT Press.

Chomsky, N. (1980). *Rules and Representations.* New York, NY, USA: Columbia University Press.

Chomsky, N. (1986). *Knowledge of Language.* Praeger.

Christiansen, M. H. and Ellefson, M. R. (2002). Linguistic Adaptation Without Linguistic Constraints: The Role of Sequential Learning in Language Evolution. In A. Wray (Ed.), *Transitions to Language* (pp. 335-358). Oxford, UK: Oxford University Press.

De Boer, B. (1999). *Self-Organisation in Vowel Systems.* Ph.D. thesis, Vrije Universiteit Brussel AI-lab.

Gold, E. M. (1967). Language identification in the limit. *Information and Control,* 10:447-474.

Goroll, N. (1999). *(The deep blue) Nile: Neuronal influences on language evolution.* Master's thesis, University of Edinburgh.

Hammond, K. (1990). Case-based planning: A framework for planning from experience. *Cognitive Science,* 14(3): 385-443.

Hauser, M. D. (2001). What's so special about speech? In E. Dupoux (Ed.), *Language, Brain, and Cognitive Development: Essays in Honor of Jacques Mehler* (pp. 417-434). Cambridge, MA, USA: MIT Press.

Hauser, M. D., Chomsky, N. and Fitch, W. T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science,* 298:1569-1579.

Hebb, D. (1949). *The Organization of Behavior: A Neuropsychological Theory.* New York, NY, USA: Wiley.

Hurford, J. R. (1987). *Language and Number: The Emergence of a Cognitive System.* New York, NY, USA: Basil Blackwell.

Hurford, J. R. (1989). Biological evolution of the saussurean sign as a component of the language acquisition device. *Lingua,* 77:187-222.

Hurford, J. R. (1990). Nativist and functional explanations in language acquisition. In I. M. Roca, (Ed.), *Logical Issues in Language Acquisition* (pp. 85-136). Foris, Dordrecht.

Hurford, J. R. (2000). Social transmission favours linguistic generalization. In C. Knight, J. R. Hurford and M. Studdert-Kennedy (Eds.), *The Evolutionary Emergence of Language: Social Function and the Origins of Linguistic Form* (pp. 324-352). Cambridge, UK: Cambridge University Press.

Hurford, J. R. (2002). Expression/induction models of language evolution: dimensions and issues. In E. Briscoe (Ed.), *Linguistic Evolution through Language Acquisi-*

tion: Formal and Computational Models (pp. 301-344). Cambridge, UK: Cambridge University Press.

Jäger, G. (2003). Simulating language change with functional OT. In S. Kirby (Ed.), *Proceedings of Language Evolution and Computation Workshop/Course at ESSLLI* (pp. 52-61). Vienna.

Kent, A., Berry, M., Leuhrs, F. U. and Perry, J. W. (1955). Machine literature searching viii: Operational criteria for designing information retrieval systems. *American Documentation*, 6(2):93-101.

Kirby, S. (2000). Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners. In C. Knight, M. Studdert-Kennedy and J. R. Hurford (Eds.), *The evolutionary emergence of language: Social function and the origins of linguistic form* (pp. 303-323). Cambridge, UK: Cambridge University Press.

Kirby, S. (2001). Spontaneous evolution of linguistic structure: An iterated learning model of the emergence of regularity and irregularity. *IEEE Journal of Evolutionary Computation,* 5(2):102-110.

Kirby, S. (2002a). Natural Language from Artificial Life. *Artificial Life,* 8(2):185-215.

Kirby, S. (2002b). Learning, bottlenecks and the evolution of recursive syntax. In E. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational models* (pp. 173-203). Cambridge, UK: Cambridge University Press.

Kirby, S. and Hurford J. R. (1997). Learning, culture and evolution in the origin of linguistic constraints. In P. Husbands and I. Harvey (Eds.), *Fourth European Conference on Artificial Life* (pp. 493-502). Cambridge, MA, USA: MIT Press.

Kirby, S. and Hurford J. R. (2002). The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi and D. Parisi (Eds.), *Simulating the evolution of language* (pp. 121-148). London, UK: Springer.

Kirby, S., Smith, K. and Brighton, H. (2004). From UG to Universals: Linguistic adaptation through iterated learning. *Studies in Language,* 28(3):587-607.

MacWhinney, B. and Bates, E. (1987). Functionalism and the competition model. In B. MacWhinney and E. Bates (Eds.), *The Crosslinguistic Study of Sentence Processing* (pp. 3-73). Cambridge UK: Cambridge University Press.

Medin, D. L. and Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review,* 85:207-238.

Montague R. (1970). English as a formal language. In B. Visentini, *et al.* (Eds.), *Linguaggi nella Società e nella Tecnica* (pp. 189-223). Milan: Edizioni di Comunità.

Newmeyer, F. J. (1991). Functional explanation in linguistics and the origin of language. *Language and Communication,* 11:3-28.

Nowak, M. A., Komarova, N. L. and Niyogi, P. (2001). Evolution of universal grammar. *Science,* 291:114-117.

Oliphant, M. (1997). *Formal Approaches to Innate and Learned Communication: Laying the Foundation for Language.* PhD thesis, Department of Cognitive Science, University of California, San Diego.

Oliphant, M. (1999). The learning barrier: Moving from innate to learned systems of communication. *Adaptive Behavior,* 7:371-384.

Oliphant, M. and Batali, J. (1997). Learning and the emergence of coordinated communication. *Center for Research on Language Newsletter*, 11(1).

Saussure, F. de. (1959). *Course in general linguistics* (translated by Wade Baskin). New York, NY, USA: McGraw Hill.

Smith, A. D. M. (2003). Intelligent Meaning Creation in a Clumpy World Helps Communication. *Artificial Life,* 9(2): 175-190.

Smith, K. (2002). The cultural evolution of communication in a population of neural networks. *Connection Science*, 14(1):65-84.

Smith, K. (2003). *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.

Teal, T. and Taylor, C. (1999). Compression and adaptation. In D. Floreano, J. D. Nicoud and F. Mondada (Eds.), *Advances in artificial life* (pp. 709-719). (No. 1674 in Lecture Notes in Computer Science). Berlin: Springer.

Tonkes, B. (2002). *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.

Vogt, P. (2003). Iterated Learning and Grounding: From Holistic to Compositional Languages. In S. Kirby (Ed.), *Proceedings of Language Evolution and Computation Workshop/Course at ESSLLI* (pp. 76-86). Vienna.

von Frisch, K. (1974). Decoding the language of the bee. *Science,* 185:663-668.

Wray, A. (2002). *Formulaic language and the lexicon.* Cambridge, UK: Cambridge University Press.

Wray, A. (2005). 'Needs only analysis' in linguistic ontogeny and phylogeny. *Proceedings of the Second International Symposium on the Emergence and Evolution of Linguistic Communication, 2005,* 151-160.

Zuidema, W. H. (2001). Emergent syntax: The unremitting value of computational modeling for understanding the origins of complex language. In J. Kelemen & P. Sosik (Eds.), *Advances in Artificial Life* (pp. 641-644). (No. 2159 in Lecture Notes in Artificial Intelligence). Berlin: Springer.

Zuidema, W. H. (2003). How the poverty of the stimulus solves the poverty of the stimulus. In Suzanna Becker, Sebastian Thrun and Klaus Obermayer (Eds.), *Advances in Neural Information Processing Systems 12 (Proceedings of NIPS'02)*. Cambridge, MA, USA: MIT Press.