Mutual Exclusivity: Communicative Success Despite Conceptual Divergence

Andrew D. M. Smith

0.1 Introduction

Traditional explanatory accounts of the evolution of language frequently appeal to a "conventional neo-Darwinian process" (Pinker & Bloom 1990: 707), assuming that humans have evolved an innate, genetically-encoded language acquisition device, which specifies a formal coding of Universal Grammar (Chomsky 1965), and which evolved incrementally through a series of steps via natural selection (Jackendoff 2002). An alternative approach focuses instead on the evolution of linguistic structures themselves, as utterances used and understood by speakers and hearers (Christiansen 1994; Croft 2000). Under the latter approach, the continual cycle of expressing and re-interpreting these utterances (Hurford 2002) drives the cultural evolution of language. Other things being equal, languages which can be readily interpreted and expressed through this cycle are more likely to persist than those which cannot.

An explanation of the evolution of syntactic structure remains the holy grail of evolutionary linguists by researchers in both these traditions, because syntax has been seen as the defining characteristic which separates human language from animal communication systems, and in recent years, computational simulations have been used extensively to shed light on this issue. Kirby (2002), for example, shows that structured signals can develop from unstructured signals through the analysis of signal-meaning pairs and the subsequent generalization of rules based on the analysis; similar accounts are presented by Batali (2002), whose agents combine and modify phrases based on exemplars of signal-meaning mappings which they receive, and by Brighton (2002), who shows how the poverty of the stimulus is an important factor in the emergence of compositional syntax.

Despite these exciting findings, however, there are some problematic assumptions in models such as these. In particular, the emergence of syntactic structure in the signal space is a direct result of the signals' explicit association with pre-defined meanings (Nehaniv 2000), and of the explicit transfer of meaning in communication (Smith 2001). Furthermore, the models often rely on reinforcement learning to guide the learners, although error signals are widely rejected in language acquisition (Bloom 2000). I have, however, developed a model of meaning creation and communication which addresses these problems and have shown that communication can succeed through the inference of meaning (Smith 2001, 2003a, 2003b). Crucially, inferential communication allows the development of communication between individuals who do not necessarily share exactly the same internal representations of meaning. This flexibility then opens the possibility of a realistic evolutionary scenario, by allowing both for the necessary variation among individuals, and also for mutations which might enhance the inferential capabilities of one individual, while still allowing them to be communicatively consistent with the rest of the population.

In this paper, I extend my inferential model to explore the usefulness of one of the main psycholinguistic acquisition biases which has been proposed to explain how children learn the meaning of words without explicit meaning transfer, namely Markman (1989)'s *mutual exclusivity assumption*. The remainder of the paper is divided into four parts: In section 0.2, I describe the signal redundancy paradox which is contained in other models; this pre-determines the outcomes which are achieved and, to a large extent, undermines the strength of their conclusions. In section 0.3, I focus further on Quine (1960)'s famous problem of the indeterminacy of meaning, and on proposals made by psychologists and psycholinguists to explain how children manage to solve this problem when they acquire language, including, of course, the mutual exclusivity assumption. In section 0.4, I briefly describe my model of individual, independent meaning creation and negotiated communication which avoids these pitfalls and yet still allows successful communication. I show, crucially, that there is a strong relationship between levels of meaning co-ordination and communicative success. Finally, in section 0.5, mutual exclusivity is added to the model, and I show that, in contrast to expectations based on my earlier models, this can lead to high levels of communicative success despite agents having divergent conceptual structures.

0.2 The Signal Redundancy Paradox

Kirby (2002) and Batali (2002), among others, have shown how the simple ability to generalize can result in the emergence of a compositional, 'syntactic' communication system. In their simulations, agents initially create idiosyncratic rules to represent each different meaning they need to express, and each of these rules generates just one signal. Over time, coincidental matches occur between parts of signals and parts of meaning, and the agents create more general rules based on these matches; these rules use symbolic variables and can therefore generate more than one signal. Brighton (2002) shows that if there are pressures on agents which limit their exposure to the language, such as the poverty of the stimulus, then the agents are more likely to encounter general rules than idiosyncratic ones, and so these general rules are preferentially replicated over generations, leading to the eventual evolution of a fully compositional communication system, where the meaning of a signal is made up of a combination of the meanings of its parts and an algorithm for joining these together.

The successful emergence of syntax in such models, however, is completely dependent on the signals being explicitly coupled to meanings which have a pre-defined and complex structure. It is not coincidental that the emergent syntactic structure parallels this semantic structure exactly, as the semantic structure is effectively used as a template against which the syntactic structure is constructed.

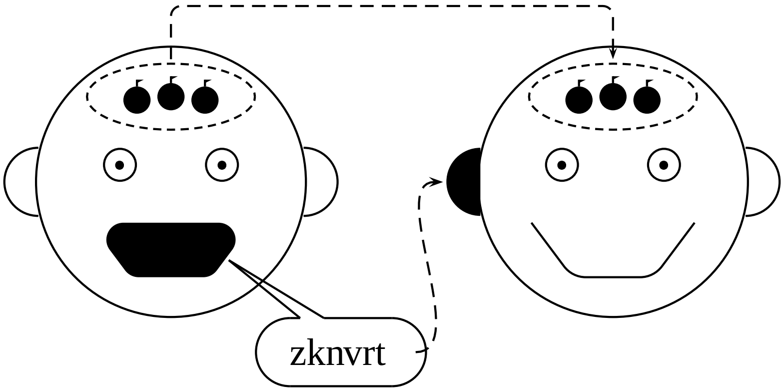

Figure 1: A communicative episode which consists of the explicit transfer of both a signal 'zknvrt' and a meaning 'three apples' from speaker to hearer.

0.2.1 Explicit Meaning Transfer

Figure 1 shows a schematic diagram of the linguistic transfer in such a communicative model, where the utterances are made up of pairs of signals and meanings. We can see that the speaker (on the left of figure 1) utters a signal "*zknvrt*", which is received by the hearer. Simultaneously, the meaning in the speaker's brain, represented in figure 1 by three symbols meant to resemble apples, is transferred directly to the hearer's brain. This explicit linkage of signal and meaning in the communication process means that it is a trivial task for the hearer to learn the association between them.

Models which make this idealization, therefore, ignore one of the most important and difficult problems facing researchers into the acquisition of language, namely Quine (1960)'s famous problem of the indeterminacy of meaning. Quine presented an imaginary anthropologist, who observes a speaker of an unfamiliar language uttering the word *"gavagai"* while pointing to a rabbit, and then shows that, logically, *"gavagai"* has an infinite number of possible meanings and, moreover, that the collection of further relevant information by the anthropologist will never reduce the number of possible hypotheses which will be consistent with the data; no matter how much evidence is collated, the meaning of *"gavagai"* can never be determined.

The consequences of the idealization of the learning process as shown in figure 1 are considerable, not least because if meanings are explicitly and accurately transferable by telepathy, then the signals are not actually being used to convey meaning, and their very role in the model must be called into question; if the agents can transfer meanings between each other, there can be no justification for them to waste time and energy worrying about learning a redundant additional system of signalling. This paradox, which I call the *signal redundancy paradox*, arises whenever meanings are explicitly transferred in communication:

- if the meanings are transferable, then the signals are redundant;

- but if the signals are removed, then to what extent does the model represent communication?

The most obvious way out of this paradox is to conclude that meanings *cannot* be explicitly transferred, but must be inferred from elsewhere.

0.2.2 Accessibility and Privacy

If there is no explicit meaning transfer, however, how does a hearer know which meaning to associate with a particular signal? The hearer must be able to infer a meaning from somewhere; the most obvious and general source for this is surely the environment in which the agent is placed. This in turn suggests that at least some of the meanings agents talk about are used to *refer* to objects and events which actually happen in the environment. In this way, the agents' meanings are grounded (Harnad 1990); without the possibility of inferring the signals' reference, real communication cannot emerge. Indeed, the existence of an external world from which meaning can be inferred is crucial to a realistic model of meaning, for without it, any 'meanings' are necessarily abstract and pre-defined. If the meanings do not identify anything in the world, or do not have reference, they can only be communicated through explicit transfer, which of course entails the signal redundancy paradox.

In order to avoid the signal redundancy paradox, therefore, there must be at least three levels of representation in the model, as shown in figure 2:

A: an external environment, which is public and accessible to all, which provides the motivation and source for meaning creation;

B: a private, agent-specific internal representation of meaning, which is not perceptible to others;

C: a set of signals, which can be transmitted between agents and is in principle public.

The mere existence of an external world, as for instance in Hutchins and Hazlehurst (1995)'s model of the emergence of a shared vocabulary, is not sufficient to avoid the paradox; if the agents' meanings are publicly accessible, either directly as in Hutchins and Hazlehurst's model where the external scenes *are* the meanings, or indirectly through an accessible mapping between the environment and the meanings, then the signals are again rendered unnecessary. For this reason, note in figure 2 that the mappings between A and B and between B and C fall to the right-hand side of the demarcation line between the public and private domains.

0.2.3 Inferential Communication

There are at least two possible explanations for how the agents come to have meanings which refer to things in their environment: either the meanings are innate, and have evolved through biological evolution; or they are created by the agents, as a result of their interactions with the environment. Innate meanings are not inherently implausible, but they seem to require either that the number of meanings useful to the

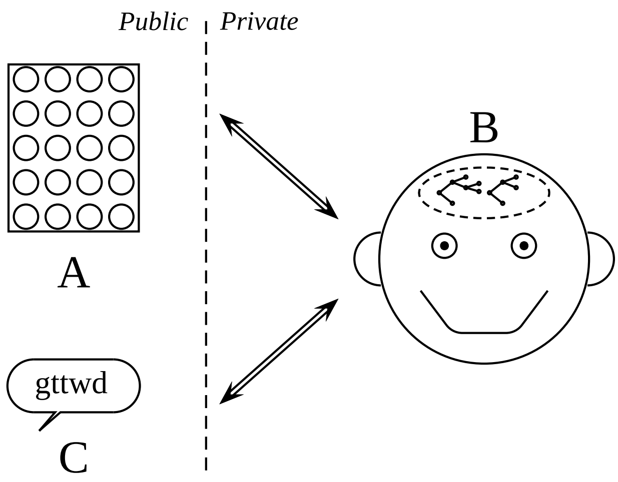

Figure 2: A model of communication which avoids the signal redundancy paradox must have three levels of representation for the agents: an external environment (A); an internal, private semantic representation, represented by the trees in the agent's brain (B); and public signals (C). The mappings between A and B, and between B and C, represented by the arrows, must also be private and inaccessible to other agents.

agents is small and fixed, or that the world in which the agents exist is very stable and unchanging, so that the evolved meanings which were useful to their ancestors are still of use to the current generation. In practice, then, it is more reasonable to assume that the agents have an innate quality space, as suggested by Quine (1960), within which they create meanings anew in each generation, based on empirical testing of their environment, which allows them to discover which distinctions are communicatively relevant.

Thinking of the communicative function of language as a simple coding system between signals and meanings, however, is problematic not just in terms of the communication model itself, but also in terms of the evolution of such a system. From this perspective, it is important to remember that language is necessarily both reciprocal and cultural. There is no advantage, therefore, in a mutant obtaining a language acquisition device if others do not have one. In addition, however, there is no advantage in *many* mutants having a language acquisition device, if there is no language existing in the community for them to acquire. As Origgi and Sperber (2000) point out, a mutation which allows individuals to infer the meanings of signals can not only provide an explanation for how language got started, through the accidental discovery of what another is referring to, but can also provide a plausible account of the progressive complexification of language. For instance, a mutation which promotes the construction of a more complex semantic representation does not, in an inferential model, cause catastrophic effects on communication due to the ensuing mismatch between the speaker's meaning and the hearer's meaning; instead, because communication is based on reference, individuals can have very different internal representations of meanings, and yet still communicate successfully, as I have shown through simulation experiments (Smith 2003b). Those without the enhanced semantic representation can still communicate with everyone in blissful ignorance, while the mutants might receive an advantage in more accurate or detailed inference of the meaning, and might, in time, develop new symbols to represent the patterns they find in this structure. Indeed, this process of structural development is most obviously attested in historical processes of language change, particularly in the case of grammaticalization (Hopper & Traugott 1993), where (more complex) grammatical markers such as case markers and complementizers are created from (less complex) lexical items over generations of inference, a process which has been explicitly described as “context-induced reinterpretation” (Heine & Kuteva 2002: 3).

The model I describe, then, departs from previous accounts which assume that language learning is merely equivalent to learning a mapping between signals and pre-defined meanings. Instead, I argue that it must include at least the construction of empirical meanings, learning which of these meanings are relevant, and learning a mapping between meanings and signals through the inference of meaning in context.

0.3 Overcoming Indeterminacy

Learning the meanings of words, of course, is utterly unremarkable to children, who effortlessly overcome Quine’s problem of indeterminacy; a common suggestion for how this happens is that they are explicitly taught by parents and teachers, by being given feedback on their use of words. Despite the intuitive appeal of this idea, it is actually very rarely observed in practice, and is by no means culturally universal. Lieven (1994), for instance, describes cultures in which parents do not even speak to their children in the initial stages of acquisition, much less provide them with either encouragement or discouragement about their use of words. Bloom (2000), furthermore, describes a study on mute children who clearly could not receive feedback on their own speech, and yet still developed language normally. In view of this, researchers have explored the existence of other representational constraints within the learners themselves which predispose them to disregard some of the theoretically possible meanings, thus reducing the size of the set of semantic hypotheses, thereby making the set finite and Quine’s problem soluble.

Macnamara (1972), for instance, argues that children naturally represent their environment in terms of the objects within it, and that, when learning words, they have a similar *object bias*, under which they automatically assume that a new word refers to a whole object, rather than particular parts or properties thereof. The object bias is indeed a very useful tool in explaining how children might bootstrap language acquisition, but it is not a sufficient explanatory tool for the larger problem, and so many additional biases or restrictions have also been proposed in order to account for more complex facets of word learning. Landau, Smith, and Jones (1988), for instance, discovered experimentally that children are more likely to categorize new objects in terms of their shape, rather than other perceptual features. Markman and Hutchinson (1984) have shown that children categorize objects taxonomically (grouping on the basis of type) rather than thematically (grouping on the basis of relationships between them) when they are learning words, but not otherwise. For instance, when word learning is not involved, a car and a car tyre can be grouped together thematically, but when the car is given a name, and the children asked to find another object which can called by the same name, they are much more likely to find the taxonomically related bicycle.

Interpretation biases, too, have often been proposed; in particular, many of these suggestions, by for instance Barrett (1986), Clark (1987) and Markman (1989), can be summarized by the proposal that "children should construct mutually exclusive extensions of the terms they acquire" (Merriman & Bowman 1989: 1). Although there are slight differences between these suggestions in terms of their theoretical and explanatory emphasis, in this paper I will consider them as related versions of an over-arching *mutual exclusivity assumption*. Merriman and Bowman (1989) analyse the implications behind mutual exclusivity, and propose three crucial ways in which the bias could affect the learning of new words; the most important of these, and the only one which does not rely on the explicit naming of objects, is through the *disambiguation of reference*. This phenomenon has been shown experimentally a number of times, particularly by Markman and Wachtel (1988), who describe experiments in which young children were presented with random pairs of objects, one of which is familiar to them, such as a banana or a spoon, and one of which is unfamiliar, such as a lemon wedge presser or a pair of tongs. The children, on being presented with both objects, were asked by the experimenters to "show me the *x*", where *x* was a randomly chosen nonsense syllable. Markman and Wachtel found that the children are much more likely to interpret *x* as referring to the tongs, rather than the banana; they hypothesize that this is because the children already understand a word which means BANANA, and they assume, under the mutual exclusivity bias, that "[w]hen children hear a novel term in the presence of a familiar and unfamiliar object, children are able to use mutual exclusivity to determine the referent of the novel term." (Markman & Wachtel 1988: 128). In section 0.5, I explore how mutual exclusivity can improve the levels of communicative success relative to the shared conceptual structure of agents in my model.

0.4 Details of the Model

0.4.1 Independent Meaning Creation

Before investigating the effects of mutual exclusivity, however, it is useful to give a brief description of my basic model of meaning creation and communication, which takes as its starting point the model initially described by Steels (1996). A simple model world is simulated, containing a number of objects, each of which can be described in terms of the values of their observable features. Feature values in the model are real numbers, pseudo-randomly generated in the range [0.0. . . 1.0]; the features themselves, however, are deliberately abstract, with neither specific nor pre-defined meanings, although for ease of understanding, they can of course be considered analogous to features in human language such as 'height', 'smell' or 'colour'. Simulated agents interact with the objects in the world using *sensory channels*; they have the same number of sensory channels as the objects have features, and there is a one-to-one mapping between them. Sensory channels are sensitive to the objects' feature values; specifically, they can detect whether a particular feature value falls between two bounds on a sensory channel. The process of meaning creation takes place through *refinement*, or the splitting of a channel's sensitivity range into two discrete segments of equal size. This results in the formation of two new categories, each sensitive to half the original range. Each category is itself a candidate for further refinement, so producing, over time, a hierarchical, dendritic structure, with the nodes on the tree representing categories, or *meanings* (Steels 1999). Such structures are shown schematically in the agent's private semantic representation in figure 2.

Interaction with the environment occurs through Steelsian discrimination games, which are made up of the following four constituent parts:

scene-setting: the agent is presented with a specific set of objects, called the *context*, one of which is chosen to be the *target* of discrimination.

categorization: the agent goes through all the objects in the context, returning for each an association with one or more of its existing semantic representations.

discrimination: the agent tries to find a distinctive category for the target. A category is distinctive if it is a valid representation of the target, and is not a valid representation of any other object in the context.

adaptation: the agent modifies its internal conceptual structure, by refining one of the sensory channels.

Adaptation of an agent's conceptual structure is triggered by failure in a discrimination game. Each agent has a *tree growth strategy* for choosing a channel for refinement, which is based on its cognitive biases and/or the details of the particular discrimination game which failed, as described in Smith (2003b). In a stable world, the agents will eventually always develop a conceptual structure which can succeed in describing every object in the world. Different agents, however, will create different conceptual structures which will each be able to distinguish objects in the world, and so it is useful to be able to measure the level of meaning similarity σ between two agents' conceptual structures (Smith 2003a).

0.4.2 Introspective Obverter

Having established that agents can create meanings which are helpful in describing the world around them, I simulate communication without explicit meaning transfer and without feedback by providing the agents with the ability to create simple signals and transmit them without error, and also with a mechanism for associating signals and meanings, an individual dynamic lexicon (Smith 2003a). In a communication episode, one agent (the speaker) is trying to convey a meaning to another agent (the hearer) by the use of a signal.

Preparatory to communication, a successful discrimination game provides the speaker with a distinctive meaning which has identified the target object from others in the context, and it is this meaning which the speaker tries to convey; it utters a signal to represent the meaning, either taking one from its lexicon, or, if none suitable exists, creating a new one at random. The hearer then tries to infer the meaning of the signal from the context in which it is heard, attempting to deduce which of the objects in the context was identified by the speaker. Successful communication is defined by *referent identity*, which occurs when the object identified by the speaker is the same object as that identified by the hearer. Note that it is not necessary that the agents use the same agent-internal meaning, only that both agents pick out the same object in the world. Importantly, neither speaker nor hearer is given any feedback on whether the communication episode is successful.

This communication model, therefore, relies neither on explicit meaning transfer meaning, nor on feedback guiding learning. The algorithms which determine the agents' behaviour, however, are crucial to its success, and are based on Oliphant and Batali (1997)'s strategy for achieving an accurate communication system in a population of agents, which they dub *obverter*. Essentially, the obverter strategy boils down to the speaker choosing signals which it knows the hearer will understand correctly, rather than choosing signals that it might prefer to say. Unfortunately, true obverter learning in the theoretical situation defined by Oliphant and Batali assumes that the speaker has access to the lexicons of the other members of the population, so that it can choose the optimal signal for each meaning. Such mind-reading is of course unrealistic, and returns us, more damagingly, to a telepathic world and the signal redundancy paradox. In order to maintain the benefits of obverter, whilst also avoiding any reliance on telepathy, I implement a modification to the obverter algorithm, in which I allow the agent to read only its own mind. The agent therefore uses introspection as a basis for decision making, choosing a signal which *it itself* would be most likely to understand if it heard the signal in this context.

Choosing a signal is relatively straightforward, but interpreting that signal is much more difficult; the hearer, to whom this task falls, knows neither the object to which the speaker is referring, nor the meaning which the speaker has in mind for the signal. The hearer creates a list of *possible meanings* or semantic hypotheses, containing every meaning in its conceptual structure which identifies *any one* of the objects in the context and distinguishes it from all the other objects in the context. The hearer has no reason to prefer any one of these possible meanings over another yet, so each of them is paired with the signal in the hearer's lexicon. Having done this for all the possible meanings, the hearer searches through its list of semantic hypotheses, and chooses the meaning m in which it has the highest confidence, which is, as Vogt and Coumans (2003) explain, the highest conditional probability that, given the current signal, the meaning m is expected. The object which the chosen meaning identifies is then compared to the original target object of the speaker's discrimination game to determine the success of the communicative episode.

0.4.3 Communicative Success

I have shown previously (Smith 2003b), that, in such a model, where the agents infer the meanings of words from the contexts in which they hear them, the percentage of successful communicative episodes, or the communicative success rate κ, is highly dependent on the level of conceptual similarity σ between the agents. I experiment with various cognitive biases, environmental factors and meaning creation strategies, to discover the circumstances under which high levels of conceptual similarity are most likely to occur, and show moreover that in a randomly-generated world, the agents cannot improve on creating meanings based on their cognitive biases, using a *probabilistic* tree growth strategy; high levels of conceptual similarity will always arise if the agents share similar values of these biases. In a structured, or clumpy world, on the other hand, then it is much better for the agents to use a more *intelligent*, ecologically rational (Gigerenzer & Todd 1999) tree growth strategy, which can exploit the information in the environmental structure to a much greater degree.

0.5 Mutual Exclusivity

Successful communication, therefore, can emerge without the need for innate meanings and without meanings being explicitly transferred between agents, if the agents use introspective obverter to choose signals. On the other hand, communicative success rates are highly correlated with levels of meaning similarity; the exact relationship varies according to the experimental conditions, but it is always a logarithmic curve with communicative success in general slightly higher than meaning similarity. In this section, I implement the mutual exclusivity bias in the model, to see what effects its inclusion has on the development both of coordinated meanings and successful communication. Two factors, in particular, are crucial in triggering the use of mutual exclusivity, and must be taken into account in developing the model, namely:

signal novelty: the utterance in question is novel, and unfamiliar to the learner;

disambiguation of reference through prior knowledge: the learner reduces the set of meanings under consideration by excluding all objects for which it already understands a word.

Under normal circumstances within my model, the hearer would, on hearing an unfamiliar word in context, build a set of all possible semantic hypotheses and use these to decipher the utterance, as described in section 0.4. Disambiguating the utterance's reference through prior knowledge, therefore, will allow this set of semantic hypotheses to be reduced; the agent works through the objects in the context, and excludes from consideration all those objects for which it already knows an appropriate word, namely a word which the agent would use, in this context, to describe the object, and which therefore represents a distinctive meaning which would distinguish this object from all the other objects in the current context. The agent is then left with a set of *unfamiliar* objects, and it assumes that the speaker must be referring to one of these. Its list of semantic hypotheses is therefore based only on these unfamiliar objects, from which the agent interprets the word as before, choosing the meaning in which it has the highest confidence probability.

In addition to this, however, Markman and Wachtel also hypothesize that mutual exclusivity can help the child to develop new meanings, when they cannot interpret an unfamiliar word, because "children would be left with a word for which they have not yet figured out a meaning. This should then motivate children to find a potential meaning for the novel term." (Markman & Wachtel 1988: 153). If no interpretation at all is possible, therefore, i.e. there are no appropriate meanings which distinguish any

Table 1: Meaning similarity σ and communicative success κ in a randomly-generated world.

Tree Growth Strategy	Meaning Similarity		Communicative Success	
	Mean ($\bar{\sigma}$)	CI	Mean ($\bar{\kappa}$)	CI
Probabilistic	0.53	(0.50 – 0.56)	0.70	(0.67 – 0.72)
Intelligent	0.59	(0.56 – 0.63)	0.73	(0.70 – 0.76)

of the unfamiliar objects from all the others in the context, then the agent searches through the unfamiliar objects in turn, trying to create a new, appropriate meaning which will be suitable to describe it in this context. It tests potential refinements on its sensory channels, until it finds a node which, once refined, will distinguish this object from all the other objects in the context, and then creates this new meaning, associating it with the unfamiliar signal.

The hearer's meaning creation process is now very different from the speaker's, both in the mechanism by which it is triggered and in the algorithm through which it is implemented; meaning creation in the hearer now occurs as a result of encountering an unfamiliar word and is a direct attempt to find a relevant interpretation of this word, but in the speaker occurs as a result of failure to discriminate a target object. This implementation of the mutual exclusivity bias differs from my earlier implementation of the principle of contrast (Smith 2003b); although both sets of simulations use the same framework of meaning creation and communication, in the earlier simulations, the agent did not divide the context into two sets of familiar and unfamiliar words before interpretation, so the list of semantic hypotheses was not reduced, and the meaning creation process was triggered very infrequently.

0.5.1 Experimental Results

In the results reported here, the agents in the model each had five sensory channels with cognitive biases distributed uniformly, and the objects in the world were generated randomly. Each simulation consists of 5000 discrimination and communication episodes, and was run 50 separate times, after which the levels of meaning similarity σ and communicative success κ were then averaged, and expressed together with 95% confidence intervals (CI).

Table 1 shows that in this randomly-generated world, for both the probabilistic and intelligent tree growth strategies, the levels of communicative success are indeed slightly higher than those of meaning similarity, as we would expect. On the other hand, we can also see that, in contrast to the experiments in Smith (2003b), there is no significant difference between the tree growth strategies, as their confidence intervals overlap; the large differences I found previously in levels of meaning similarity between these tree growth strategies are almost completely neutralized if the hearer uses mutual exclusivity to guide its interpretation and meaning creation.

Table 2, on the other hand, shows similar experiments in a simulated clumpy

Table 2: Meaning similarity σ and communicative success κ in a clumpy, structured world.

Tree Growth Strategy	Meaning Similarity		Communicative Success	
	Mean ($\bar{\sigma}$)	CI	Mean ($\bar{\kappa}$)	CI
Probabilistic	0.35	(0.33 – 0.37)	0.81	(0.79 – 0.83)
Intelligent	0.92	(0.88 – 0.95)	0.90	(0.88 – 0.92)

world, where the objects are grouped together and given identical values on some features, so that they are a priori indistinguishable on that sensory channel. I have shown previously (Smith 2003b) that the intelligent strategy will produce much higher levels of meaning similarity σ under these circumstances, as it is much more able to exploit the underlying information structure in the environment. We can indeed see in table 2 that meaning similarity is much improved under the intelligent tree growth strategy, as this would predict. Much more interestingly, however, the levels of communicative success in these experiments no longer bear any close relationship with the levels of meaning similarity. We can see that the communicative success levels are very high under both strategies; in particular, even when agents have very dissimilar conceptual structures ($\sigma = 0.35$) under the probabilistic strategy, the use of mutual exclusivity means that the hearer can learn to associate the relevant meanings with the signals and communicate much more successfully than results without mutual exclusivity would suggest.

Agents have different meaning creation processes, which promote very different patterns of conceptual growth. Specifically, the speaker, who creates meaning in response to discrimination game failure, has much more conceptual structure than the hearer, who creates meaning is response to the need to understand unfamiliar words. Moreover, in accordance with Grice (1975)'s conversational principles, the agents use meanings in communication which provide sufficient information to identify the target object, but which are not unnecessarily specific. The meanings which the hearer creates under these circumstances are therefore necessarily communicatively relevant, because they can be used to discriminate at least one unfamiliar object from a group of others and therefore describe that object within a communicative episode.

Although the hearer has far fewer meanings, this leads to a situation where those meanings it does have are more relevant and useful for communication, and so the level of communicative success is much higher than might be expected.

0.6 Summary

I have presented in this paper a model of independent meaning creation and communication, which avoids the signal redundancy paradox and can yet produce successful communication through the inference of meaning from context. The inference of meaning is a crucial factor in the evolution of language, because it can explain both the genesis and the incremental development and complexification of negotiated communication systems. Individuals with non-identical semantic representations are able to communicate successfully, while *variation*, necessary to drive language change, and *flexibility*, necessary to allow mutation in semantic representation without catastrophic communication breakdown, both occur naturally as by-products of the meaning inference process itself.

The level of meaning similarity between agents has previously been shown to be very important in predicting the likely level of communicative success in previous simulations. In the experiments presented here, however, I have introduced an assumption of mutual exclusivity into the hearer's interpretation process, and the creation of meaning in order to disambiguate the reference of unfamiliar words. These modifications lead to the development of fewer, but more relevant meanings in the hearer's conceptual structure, and lower levels of meaning similarity between agents; despite this conceptual divergence, using the mutual exclusivity bias produces relatively high levels of communicative success. Biases such as mutual exclusivity, therefore, might have evolved because they allow communicative systems based on the inference of meaning to be shared between individuals with different conceptual structures.

Key Further Readings

Bloom (2000) provides an ideal starting-point for investigating the many aspects of children's word learning. A more detailed discussion of the mutual exclusivity bias in particular, including its theoretical implications and experimental evidence for its development can be found in Merriman and Bowman (1989). Briscoe (2002) is an important collection of papers for those interested in computational simulations of language evolution, while the particular model of communication and meaning creation described here was inspired by that described in Steels (1999).

Acknowledgements

Andrew Smith was supported by ESRC Research Grant RES-000-22-0266.

References

Barrett, M. D. (1986), Early semantic representations and early word usage. In S. A. Kuczaj & M. D. Barrett (eds.), *The acquisition of word meaning*. New York: Springer, 39–67.

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*. Cambridge: Cambridge University Press, 111–172.

Bloom, P. (2000), *How children learn the meanings of words.* Cambridge, MA: MIT Press.

Brighton, H. (2002), Compositional syntax from cultural transmission. *Artificial Life, 8*(1): 25–54.

Briscoe, E. (ed.). (2002), *Linguistic evolution through language acquisition: Formal and computational models.* Cambridge: Cambridge University Press.

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

Christiansen, M. (1994), *Infinite languages, finite minds: Connectionism, learning and linguistic structure.* PhD thesis, University of Edinburgh.

Clark, E. V. (1987), The principle of contrast: a constraint on language acquisition. In B. MacWhinney (ed.), *Mechanisms of language acquisition.* London: Erlbaum.

Croft, W. (2000), *Explaining language change: an evolutionary approach.* Harlow: Pearson.

Gigerenzer, G., & Todd, P. M. (1999), Fast and frugal heuristics: the adaptive toolbox. In G. Gigerenzer, P. M. Todd, & the ABC Research Group (eds.), *Simple heuristics that make us smart*. Oxford: Oxford University Press, 3–34.

Grice, H. P. (1975), Logic and conversation. In P. Cole & J. L. Morgan (eds.), *Syntax and semantics* (Vol. 3). New York: Academic Press, 41–58.

Harnad, S. (1990), The symbol grounding problem. *Physica, D 42*: 335–346.

Heine, B., & Kuteva, T. (2002), *World lexicon of grammaticalization.* Cambridge: Cambridge University Press.

Hopper, P. J., & Traugott, E. C. (1993), *Grammaticalization.* Cambridge: Cambridge University Press.

Hurford, J. R. (2002), Expression/induction models of language evolution: dimensions and issues. In E. Briscoe (ed.), *Linguistic evolution through language acquisition: Formal and computational models*. Cambridge: Cambridge University Press, 301-344.

Hutchins, E., & Hazlehurst, B. (1995), How to invent a lexicon: the development of shared symbols in interaction. In N. Gilbert & R. Conte (eds.), *Artificial societies: the computer simulation of social life.* London: UCL Press.

Jackendoff, R. (2002), *Foundations of language: Brain, meaning, grammar, evolution.* Oxford: Oxford University Press.

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

Landau, B., Smith, L. B., & Jones, S. S. (1988), The importance of shape in early lexical learning. *Cognitive Development*, *3*: 299–321.

Lieven, E. V. M. (1994), Crosslinguistic and crosscultural aspects of language addressed to children. In C. Gallaway & B. J. Richards (eds.), *Input and interaction in language acquisition*. Cambridge: Cambridge University Press, 56–73.

Macnamara, J. (1972), The cognitive basis of language learning in infants. *Psychological Review*, *79*: 1–13.

Markman, E. M. (1989), *Categorization and naming in children: problems of induction.* Cambridge. MA: MIT Press.

Markman, E. M., & Hutchinson, J. E. (1984), Children's sensitivity to constraints on word meaning: Taxonomic vs. thematic relations. *Cognitive Psychology*, *16*: 1–27.

Markman, E. M., & Wachtel, G. F. (1988), Children's use of mutual exclusivity to constrain the meaning of words. *Cognitive Psychology, 20*: 121–157.

Merriman, W. E., & Bowman, L. L. (1989), The mutual exclusivity bias in children's word learning. *Monographs of the Society for Research in Child Development, 54*(3–4).

Nehaniv, C. L. (2000), The making of meaning in societies: Semiotic and information-theoretic background to the evolution of communication. In *Proceedings of the AISB symposium: Starting from society — the application of social analogies to computational systems, 19–20 April 2000*. AISB, 73–84.

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

Origgi, G., & Sperber, D. (2000), Evolution, communication and the proper function of language. In P. Carruthers & A. Chamberlain (eds.), *Evolution and the human mind: modularity, language and meta-cognition*. Cambridge: Cambridge University Press, 140–169.

Pinker, S., & Bloom, P. (1990), Natural language and natural selection. *Behavioral and Brain Sciences, 13*: 707–784.

Quine, W. v. O. (1960), *Word and object.* Cambridge, MA: MIT Press.

Smith, A. D. M. (2001), Establishing communication systems without explicit meaning transmission. In J. Kelemen & P. Sosík (eds.), *Advances in Artificial Life: Proceedings of the* 6^{th} *European Conference on Artificial Life*. Heidelberg: Springer-Verlag, 381–390.

Smith, A. D. M. (2003a), *Evolving communication through the inference of meaning.* PhD thesis, Philosophy, Psychology and Language Sciences, University of Edinburgh.

Smith, A. D. M. (2003b), Intelligent meaning creation in a clumpy world helps communication. *Artificial Life, 9*(2): 175–190.

Steels, L. (1996), Perceptually grounded meaning creation. In M. Tokoro (ed.), *Proceedings of the International Conference on Multi-agent Systems.* Cambridge, MA: MIT Press.

Steels, L. (1999), *The Talking Heads experiment* (Vol. I. Words and Meanings). Antwerpen: Laboratorium. (Special pre-edition)

Vogt, P., & Coumans, H. (2003), Investigating social interaction strategies for bootstrapping lexicon development. *Journal of Artificial Societies and Social Simulation*, *6*(1). (http://jasss.soc.surrey.ac.uk/6/1/4.html)