Acquiring words across generations: introspectively or interactively?

Zoran Macura
Department of Computer Science
King's College, London
The Strand, London WC2R 2LS
zoran.macura@kcl.ac.uk

Jonathan Ginzburg
Department of Computer Science
King's College, London
The Strand, London WC2R 2LS
jonathan.ginzburg@kcl.ac.uk

Abstract

How does a shared lexicon arise in population of agents with differing lexicons, and how can this shared lexicon be maintained over multiple generations? In order to get some insight into these questions we present an ALife model in which the lexicon dynamics of populations that possess and lack metacommunicative interaction (MCI) capabilities are compared. We ran a series of experiments on multi-generational populations whose initial state involved agents possessing distinct lexicons. These experiments reveal some clear differences in the lexicon dynamics of populations that acquire words solely by introspection contrasted with populations that learn using MCI or using a mixed strategy of introspection and MCI. The lexicon diverges at a faster rate for an introspective population, eventually collapsing to one single form which is associated with all meanings. This contrasts sharply with MCI capable populations in which a lexicon is maintained, where every meaning is associated with a unique word. We also investigated the effect of increasing the meaning space and showed that it speeds up the lexicon divergence for all populations irrespective of their acquisition method.

1 Introduction

A key feature of natural language is metacommunicative interaction (MCI)—utterance acts in which conversationalists acknowledge understanding or request clarification. The need to verify that mutual understanding among interlocutors has been achieved with respect to any given utterance—and engage in discussion of a clarification request if this is not the case—is one of the central organising principles of conversation (Clark, 1996). However, hitherto there has been little work on the emergence and significance of MCI meaning.

What significance does MCI have for linguistic interaction within a community? Pretheoretically, they serve as a device for ensuring a certain state of equilibrium or lack of divergence gets maintained within a linguistic community. The plausibility of this speculation can be assessed by converting it into more concrete questions such as the following:

(1) a. Given a community A where clarification requests do not get expressed, and community B where they do, how do the two communities evolve with respect to vocabulary drift.

b. How does this vocabulary drift change once a gradual turnover of community members is introduced?

In previous work we have shown how language converges for different types of populations in a mono-generational model (Ginzburg and Macura, in press). We also compared the performance of mono-generational and multi-generational populations and showed how the introduction of infants and mortality in the model affects the lexicon dynamics (Macura and Ginzburg, in press). In this paper we take a closer look at multi-generational populations, in particular the effect of varying meaning space—number of different plants in the environment—on the results.

In the next section we describe the computational model, including how gradual turnover of agents is implemented. In Section 3 we present the experiments and assess the validity of the proposed model. Finally, in Section 4, we draw some conclusions.

2 The Model

In our previous work we have shown how language converges for different types of populations within a single generation (Ginzburg and Macura, in press). In this type of model there is no generational turnover of agents and the transmission of language is horizontal, whereas the communication is between adult agents of the same generation (e.g. Steels (1998)). In multi-generational models such as the iterated learning model (e.g. Kirby et al. (2004); Smith (2005)) language is vertically transmitted from one generation to the next, where the adult agents are allowed to speak to the child agents only. So in these models there is no horizontal communication (i.e. between adults of the same generation).

We present a model which implements both horizontal (adult-adult) and vertical (adult-child) language transmission (see Vogt (2005) for a similar approach). The model contains an ALife environment in which the lexicon dynamics of populations that possess and lack MCI capabilities are compared. The environment is modelled loosely after the Sugarscape environment (Epstein and Axtell, 1996), in that it is a spatial grid containing different plants. Plants can be perceived and disambiguated by the agents. Agents walk randomly in the environment and when proximate to one another engage in a brief conversational interaction concerning plants visible to the agents.[1]

In the next section we look at the communication protocol in more detail, followed by a closer look at the implementation of generational turnover.

2.1 Communication

Agents can talk about the plants in the environment by making syntactically simple utterances—essentially one consisting of a single word. Every agent has an internal lexicon which is represented by an association matrix (see Smith (2005) for a

similar approach). The lexicon stores the association scores for every meaning–representation pair (i.e. plant–word) based on individual past experiences. Agents don't have an invention capability therefore are only able to talk about the plants that they have a representation for.

Communication is a two sided process involving an intrinsic asymmetry between speaker and addressee: when talking about a plant in his field of vision, the speaking agent necessarily has a lexical representation of the plant (a word with the highest association score for the plant chosen as the topic), which he sends to the hearing agent. There is no necessity, however, that the addressee agent is able to interpret this utterance. If unable to do so (meaning that the hearing agent doesn't have the word in her lexicon, or that the plant it associates with the word is not in her context) the way that the agent tries to ground it depends on the agent's type.

Three types of communicative agents exist in the model; agents capable of making a clarification request (CR agents), agents incapable of doing so (introspective agents), and hybrid agents that use both CRs and introspection.

An introspective agent learns the meanings of words through disambiguation across multiple contexts. Upon hearing a word the agent looks around her and for every plant in her context (field of vision) she increases its association score with the word heard. This strategy is akin to the cross-situational statistical learning strategy used by inferential agents in Smith (2005), and to selfish learners in Vogt and Coumans (2003).

A CR agent on the other hand can resort to a clarification request upon hearing a word. If hearing the word for the first time (no associations with the word in her lexicon) or if there are no plants in her context, a clarification request is raised. Otherwise the agent checks the plants in her context and if there is a mismatch between her internal state and the context (agent thinks that the word heard refers to a plant not in her context) she again resorts to raising a clarification request. The speaking agent answers this clarification request by pointing to the plant intended, after which the hearing agent increases the association score of the word heard with the pointed plant. However, if the perceived plant is in her context then the hearing agent only reinforces its association score with the word heard.

---

[1]An agent's field of vision consists of a grid of fixed size originating from his location. Hence proximate agents have overlapping but not identical fields of vision.

A hybrid agent has a capability of either using the CR strategy or the introspective strategy. The agent only resorts to a clarification request if she cannot ground the word heard (there are no plants in her context or there is a mismatch between her internal state and the context). When hearing an unknown word and having some plants in the context the agent follows the introspective strategy.

After updating her lexicon[2] the hearing agent chooses the plant with the highest association score for the word heard. If this perceived plant matches with the speakers intended plant then the conversational interaction is deemed as a success. Neither agent is given any feedback on the outcome of their conversational interaction (see Smith (2005) for a similar approach).

2.2 Generational Turnover

A typical approach when modelling a multi-generational population is the introduction of mortality and child agents. The iterated learning model (Kirby et al., 2004) is an example of a multi-generational model where the language transmission is vertical (i.e. from one generation to the next). In such models the adult agents are always the speakers and child agents are always the hearers. The agents play a number of language games, which defines the length of a generation. At the end of a generation, the adults are removed from the model, the children become the new adults, and new children are introduced. This way of implementing generational turnover in the iterated learning model and other multi-generational models (e.g. Vogt and Coumans (2003)) is very rigid.

We propose a multi-generational model which is more realistic and resembles closer a human community (e.g. a tribe). In order to extend the mono-generational model described in (Ginzburg and Macura, in press) into a multi-generational model, there is a need to introduce a gradual agent turnover. This is done by introducing mortality. Every agent has a maximum age which is set randomly when the agent is born, and it lies in the range of ±20% from agent to agent. Upon reaching his maximum age the agent dies. Thus it is very unlikely that the whole adult population dies out at the same time as the adult agents are of different ages and have different maximum ages.

In order to keep the population size stable, we also introduce natality. So for every agent that dies a new infant agent is born to a random adult agent in the model. The infant agent inherits the parent's type (introspective, CR or hybrid). Infants have an empty lexicon, with no knowledge of the meaning space or the word space. Each infant follows the parent around and is only able to listen to the parent's dialogues with other agents. In fact an infant only hears the dialogues in which her parent is the speaker. So the assumption here is that an infant learns only the words uttered by her parent. An infant cannot be a speaker and learns exclusively by introspection. Every infant agent has an adulthood age which is set randomly and is about a sixth of the agent's lifespan. When reaching the adulthood age the infant stops following her parent and becomes an adult, meaning that it is able to walk around independently, engage in dialogues with other adult agents and become a parent. An infant can die only if her parent reaches the maximum age and dies.

This multi-agent model implements both vertical and horizontal language transmission as adult agents can communicate with each other as well as parent agents can communicate with their children. There is no clear distinction of when a generation starts and ends, like in the other multi-generational models, because there is continual agent turnover which makes calculating the results more intricate (see Section 3).

3 Experimental Results

This section describes different setups and experiment results for the model described in Section 2. In order to test the questions raised in (1) we ran several experiments in which agents posses distinct lexicons, and clarification requesting (CR) and introspective capabilities.

Before creating a population of agents, the environment is created containing 40 different plants (which represent 40 different meanings). There are three instances of every plant and they are randomly distributed in the environment.

The population in the simulations described here is made up of 40 agents that are also randomly distributed in the environment at the start. 20% of the initial population is made up of infants (i.e. 8 infant agents). Agents form two different communities each of whose members initially share a common lexicon. The initial com-

---

[2]Only the hearing agents update their lexicons after a conversational interaction.

munity lexicons are distinct from each other (in that no meaning has the same representation associated with it). Agents can be either of the same or different type within the community. Apart from the differences in the initial lexicons and types between the agents, all other properties are the same.

Once the simulation starts the agents begin walking randomly in the environment. At every time step agents' age increase and each agent moves to a random position in the environment. After moving an agent looks for other agents (that fall into his field of vision). If an agent sees another agent then two of them enter a dialogue where the 'see-er' is the speaker and the 'seen' is the addressee. After a dialogue the agents continue walking in a random direction. When an agent reaches his maximum age he dies and a new infant is born.

The performance of the model is based upon these behaviours which are collected at regular intervals in a simulation run:

- *Lexical Accuracy*: the population average of correctly acquired words. A word is said to be correctly acquired if it is associated with the same meaning as in either of the two initial lexicons.

- *Meaning Coverage*: the average number of meanings expressible by the overall population. There is no requirement that the meanings have correct associations.

- *Word Coverage*: the average number of words expressible by the population (correctness not taken into account).

- *Communicative Success*: the percentage of successfully completed conversations. A successful conversation is when the intended meaning by the speaker matches the perceived meaning by the hearer.

- *Method of Acquisition*: the percentage of conversational interactions that follow the introspective strategy or the CR strategy.

- *Distinct Lexicons*: the total number of distinct lexicons in the population. A lexicon is distinct only if there is no other lexicon in the population with which it shares all plant-word associations, so even if two or more lexicons have 19 out of 20 same plant-word associations they are regarded as distinct.

- *Lexical Convergence*: the percentage of agents sharing a lexicon. Agents share a lexicon if and only if all the plant-word associations are the same in their respective lexicons. Lexical convergence of 1 implies that all the agents use the same words for every plant in their lexicons.

The initial conditions and model parameters affect the above behaviours in complex ways. To determine what consequences arise when a single parameter is manipulated there is a need to control all other parameters and keep them constant whilst only manipulating the parameter being investigated.

Each parameter has a default value throughout the experiments, unless it is being investigated. The default and investigative values are shown in Table 1. In this paper we investigate the effect of increasing the meaning space on the lexicon dynamics of different populations.

Parameter	Default	Investigative
population size	40	-
adulthood age	5000 $\pm$1000	-
max age	30000 $\pm$5000	-
meaning space	40	20, 40, 60

Table 1: Default and investigative parameter values used during the experiments.

We ran four types of experiments with different population make-ups, namely introspective populations, CR populations, hybrid populations and mixed populations (made up of both introspective and CR agents in a 1:1 ratio). For all different experiments, 10 trial runs were carried out for statistical analysis.

In the first set of experiments the default parameter values as shown in Table 1 were used (Sectction 3.1). Then experiments with varying meaning space (Section 3.2) were carried out in order to get some insight into how it affects the outlined behaviours.

3.1 Multi-generational Experiments

The population in these experiments is kept constant to around 40 agents at any moment in time and the ratio of adults to infants is roughly 3:1. The agent life span is limited to around 30,000 ticks ($\pm$20%). Results were taken at every 20,000 ticks. The simulation is stopped when it reaches 2

million ticks, which means after around 70 generations.

The lexical accuracy initially drops very sharply for every population (Figure 1). At the beginning of the simulation there are a total of 80 words in the population (40 words from each community). As the words compete with one another there is a point when one word becomes dominant for a given plant and the majority of agents start using it. Thus the other competing words for the same meaning are used less frequently. The fact that the infant agents only learn the words uttered by their parents makes it very unlikely that the infrequently uttered words will pass to the next generation. After about three generations (100,000 ticks) the lexicon stabilises for the CR and hybrid populations, whilst for the mixed and introspective populations it keeps diverging.
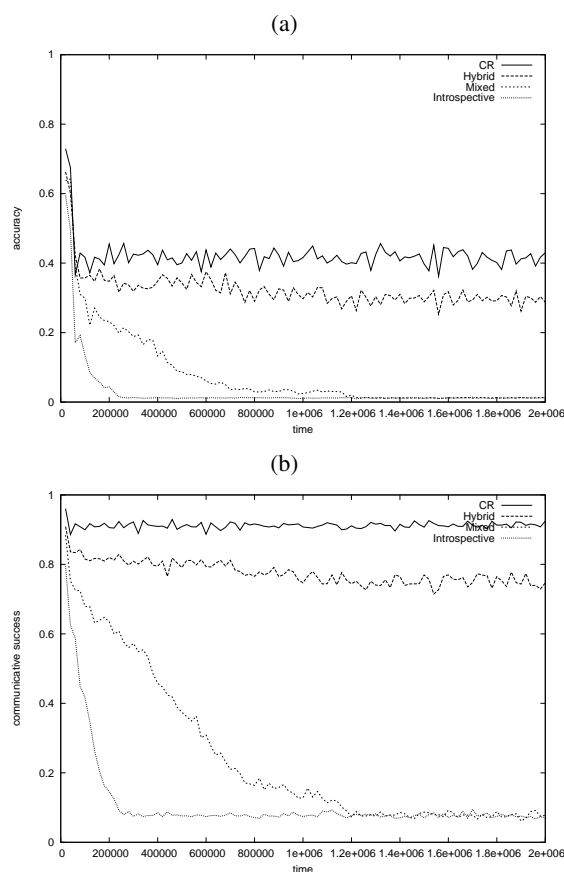

Figure 1: (a) Lexical accuracy, (b) Communicative success.

The reason for this stabilisation in CR and hybrid population can be explained by looking at Figure 2. Figure 2(a) shows that the meaning coverage for different populations is stable (all of them are able to express nearly every meaning).

The word coverage however drops rapidly along with the lexical accuracy, as seen in Figure2(b). This is an indication that only the dominant words are surviving. Once the word coverage drops to around 50% the lexicon stabilises. Around 40 different plants are expressible by the population at this stage, so every plant is associated with one word. These words can be successfully passed onto the next generation as they are used with greater frequency.
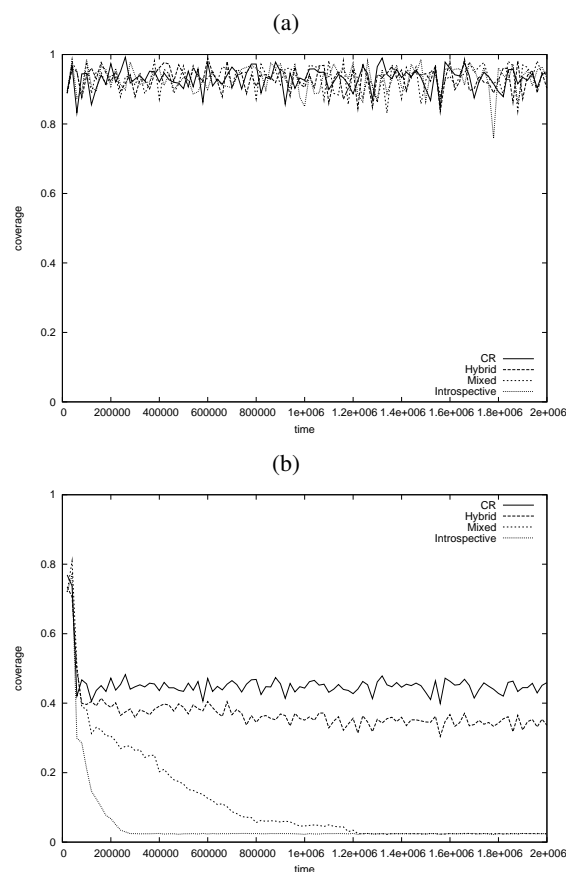

Figure 2: (a) Meaning and (b) Word coverage.

This is not the case for the mixed and introspective populations. The lexicon keeps diverging very rapidly and eventually reaches nearly 0% convergence (very few words have the association with the same plants as in the initial lexicon). Looking again at Figure 2 explains why this happens. The word coverage also drops very sharply, where in the end only one word is known by the whole population. The meaning coverage is comparable with other populations (where are all able to express nearly all the plants) so it is easy to see that every plant in the population is associated with this single word. The divergence is considerably slower in the mixed population than in the introspective.

The communicative success is in turn affected by the lexical accuracy as can be seen in Figure 1. The reason is that the higher the lexical accuracy is, the more similar the lexicons are between the agents in the population. Thus the more plant-word associations the agents share the more successful communications they are likely to have. Note that even though the lexicon is diverging at a fast rate initially, the agents in CR and hybrid populations are still able to communicate successfully about different plants.

The percentage of conversational interactions where introspective or CR strategy has been employed is shown by Figure 3. It can be seen that the populations in which CRs can be expressed (CR, hybrid and mixed) perform much better than the ones in which CRs can't be expressed (introspective). An interesting observation is that the clarification strategy in the mixed populations raises for the first 80,0000 ticks and then levels off. This is because as the lexicon is steadily diverging the agents are less successful in communicating, leading to more CRs being raised.
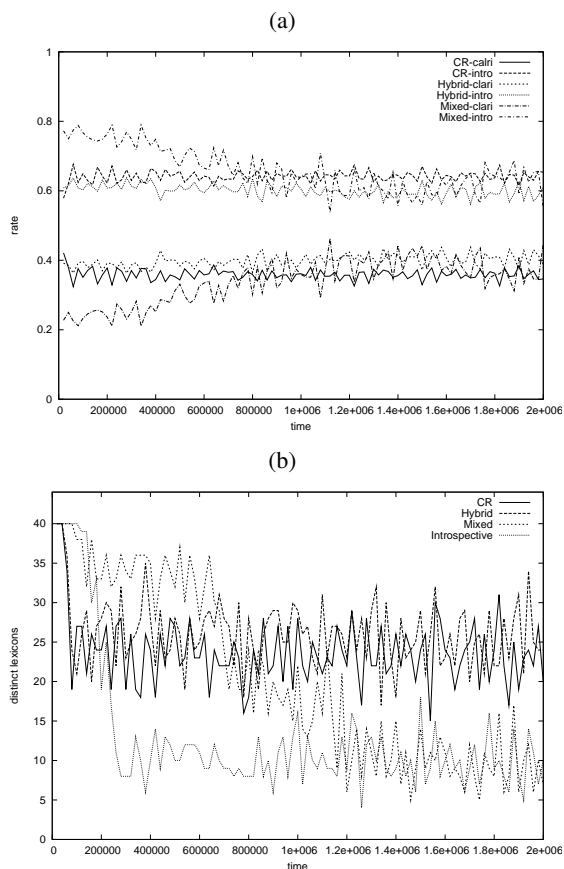

Figure 3: (a) Method of acquisition, (b) Distinct lexicons.

None of the populations converge to a single common shared lexicon (Figure 3(b)). One reason for this is that infant agents often have incomplete lexicons which differ from other agents, and this brings up the number of distinct lexicons. Another reason derives from the way common lexicons are calculated. Two or more agents are said to share a common lexicon if and only if all the plant-word associations are the same in their respective lexicons. But as there are 40 meanings it is very unlikely that all the agents will have the same associations. Thus, even though they might share the majority of the associations their lexicons are considered as distinct. We can induce from Figure 1 that the convergence for CR and hybrid populations is high where between 80% to 95% of the plant-word associations are shared. The lower number of distinct lexicons in the introspective and mixed populations might suggest that they have converged to a common lexicon. Strictly speaking, this is true, but as we have shown one word is used for representing every plant so the majority of agents converge to the same lexicon containing only this single word.

3.2 Meaning Space Variation

In this set of experiments we manipulate the meaning space—the number of different plants in the simulation. Increasing the meaning space involves increasing the differentiation among types of plants. The actual number of tokens remains constant (i.e. 120 plants). Thus when the meaning space is 20 there are six instances of each plant in the environment, whilst when the meaning space is 60 there are only two instances.

The effect of increasing the meaning space is similar for the different types of populations, thus we only present the results of CR populations. Figure 4 shows that increasing the meaning space from 20 to 60 causes a fall of around 20% in both the lexical accuracy and communicative success.

Meaning coverage is affected to a lesser extent but there is still a slight drop as the meaning space increases (Figure 5(a)). Word coverage, however, drops more significantly (Figure 5(b)). One reason for this is that as the meaning space increases the actual number of plants stays constant (e.g. for meaning space = 60 there are only two instances of each plant type in the environment). Therefore the agents are less likely to talk about all different plants as they encounter each one infrequently.
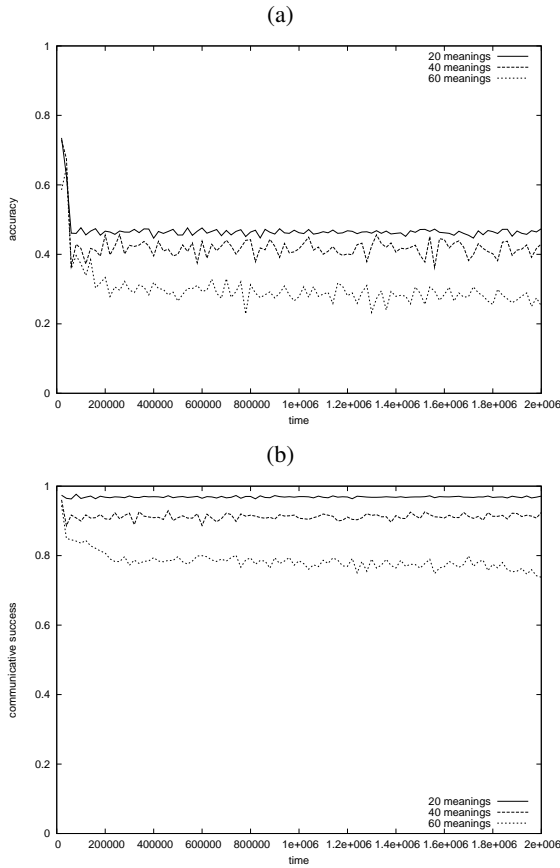
Figure 4: (a) Lexical accuracy, (b) Communicative success.

Figure 5: (a) Meaning and (b) Word coverage.

The percentage of clarification requests increases as more plant types are introduced (Figure 6(a)). The reason for this is, presumably, that there is greater uncertainty as to the referent of a word heard. This uncertainty rises as more plants are introduced, causing the agents to resort to clarification requests more often. Figure 6(b) shows that the number of distinct lexicons also rises as the meaning space increases: as there are more possible meanings it is less likely that agents will have the same association for all the meanings in their lexicons.

4   Conclusions and Future Work

In this paper we have discussed how metacommunicative interaction (MCI) serves as a key component in the maintenance of a linguistic interaction system. We ran a series of experiments on multi-generational populations in which lexicon dynamics of the populations that posses and lack MCI capabilities were compared.

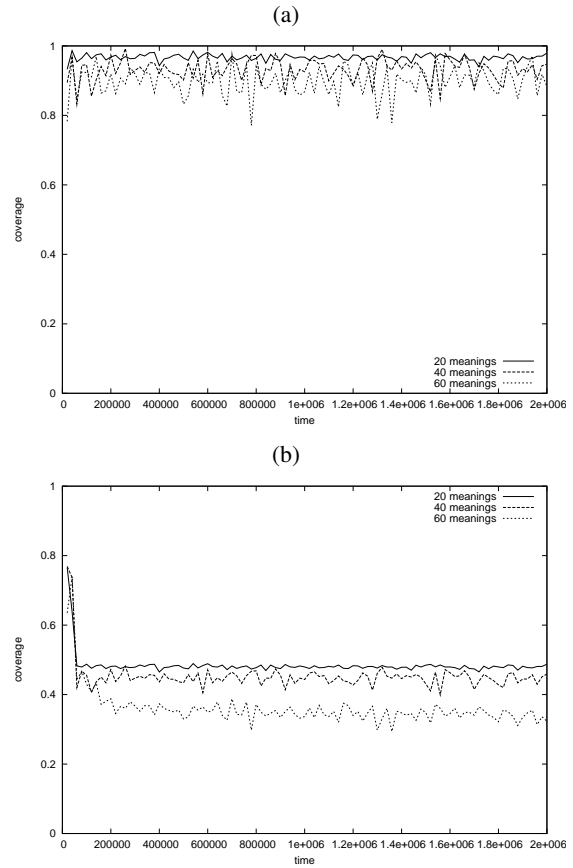We showed that limiting life span of agents in the multi-generational model raised some clear differences in the lexicon dynamics between the MCI capable and incapable populations. The main effect demonstrated is that in the introspective (and to a lesser extent mixed) populations the lexicon diverges continually, ending up with a situation where every agent in the population uses the same word to represent every plant in the environment. On the other hand MCI capable populations are able to maintain the lexicon, and the adult agents converge to a common lexicon.

We also investigated the effect of increasing the meaning space and showed that it speeds up the lexicon divergence for all populations irrespective of their acquisition method.

While this confirms our initial theorising, much work remains to butress it as a fundamental dividing line between MCI-ful and MCI-less populations. In our current experiments we are seeing that increasing the maximum age of agents improves the lexicon stability and convergence. Further work needs to be done in order to get more insight into this issue.

A crucial issue, which given space considerations we can only discuss here telegraphically, is
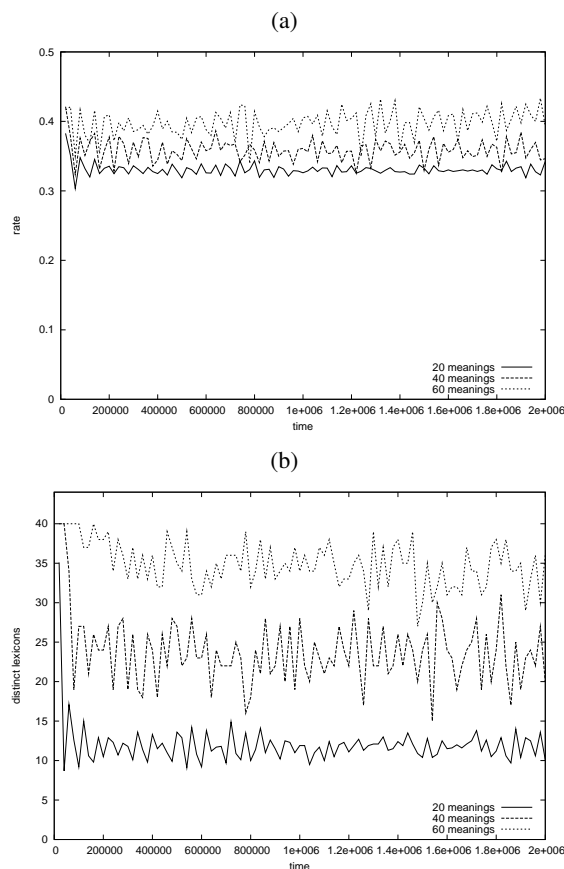
Figure 6: (a) Method of acquisition, (b) Distinct lexicons.

the relevance of the current simulation to *real human language use*.[3] There are a variety of simplifications in the current set up, possibly the grossest one is that agents employ a language lacking any sort of syntactical complexity. This would in turn lead to a massive increase in the size of the (potential) meaning space. While it is certainly an interesting and important extension to the current work, it is at least worth pointing out why such a move need not alter the current results beyond recognition. Recent corpus research on the distribution of clarification requests (see e.g. Rodriguez and Schlangen (2004); Purver (2006)) makes it clear that the lion's share of CRs in human conversation concern clarification of reference, of deixis, and of mishearing. Moreover, there is no evidence for CRs that concern syntactic ambiguity (e.g. attachment or scope). Hence, even in real human language the main communication difficulties seem to center on referential or lexical uncertainty.

---

[3] As emphasized by a Brandial reviewer.

References

Herbert H. Clark. 1996. *Using Language*. Cambridge University Press, Cambridge.

Joshua M. Epstein and Robert Axtell. 1996. *Growing Artificial Societies: Social science from the bottom up*. MIT Press.

Jonathan Ginzburg and Zoran Macura. in press. Lexical acquisition with and without metacommunication. In Caroline Lyon, Chrystopher L. Nehaniv, and Angelo Cangelosi, editors, *Emergence of Communication and Language*. Springer Verlag, London.

Simon Kirby, Kenny Smith, and Henry Brighton. 2004. From ug to universals: linguistic adaptation through iterated learning. *Studies in Language*, 28(3):587–607.

Zoran Macura and Jonathan Ginzburg. in press. Lexicon convergence in a population with and without metacommunication. In E. Tuci P. Vogt, Y. Sugita and C. Nehaniv, editors, *Proceedings of the Third International Symposium on the Emergence and Evolution of Linguistic Communication (EELC06)*, Lecture Notes in Computer Science, Berlin. Springer-Verlag.

Matthew Purver. 2006. Clarie: Handling clarification requests in a dialogue system. *Research on Language and Computation*, to appear.

Kepa J. Rodriguez and David Schlangen. 2004. Form, intonation and function of clarification requests in german task-orientaded spoken dialogues. In *Proceedings of the Eighth Workshop on Formal Semantics and Dialogue*.

Andrew D. M. Smith. 2005. The inferential transmission of language. *Adaptive Behavior*, 13(4):311–324.

Luc Steels. 1998. The origins of ontologies and communication conventions in multi-agent systems. *Autonomous Agents and Multi-Agent Systems*, 1(2):169–194, October.

Paul Vogt and Hans Coumans. 2003. Investigating social interaction strategies for bootstrapping lexicon development. *Journal of Artificial Societies and Social Simulation*, 6(1), 1.

Paul Vogt. 2005. On the acquisition and evolution of compositional languages: Sparse input and the productive creativity of children. *Adaptive Behavior*, 13(4):325–346.

121