EXPLORATIONS IN SYNTHETIC PRAGMATICS

Christian Balkenius
Simon Winter

Lund University Cognitive Science
Kungshuset, Lundagård
S–222 22 Lund
Sweden

christian.balkenius@fil.lu.se
simon.winter@fil.lu.se

Abstract: We explore a number of pragmatic principles of communication in a series of computer simulations. These principles characterize both the environment and the behavior of the interacting agents. We investigate how a common language can emerge and when it will be useful to communicate, rather than try the task without communication. When we include the cost of communicating, it becomes favorable to communicate only when expectations are not met.

1 INTRODUCTION

How can a common language emerge without a central authority? Who decides on word meaning? When is it more efficient to perform a task alone, and when is it more efficient to ask others? We have studied these questions in computer simulations of a minimal environment where two agents must communicate about a simple task.

Very recently, a number of researchers have investigated these questions. (Hutchins and Hazlehurst, 1995, Mataric, 1993, Moukas and Hayes, 1996, Noble and Cliff, 1996, Steels, 1996a; 1996b; forthc., Yanco and Stein, 1993, Yanco, 1994). Our approach relates to the Adaptable Synthetic Robot Language (ASRL) paradigm developed by Yanco and Stein (1993) and Yanco (1994).

Yanco (1994) identifies two distinctions among ASRLs. First, whether the language is pre-engineered or developed by the agent itself. Second, whether the agents are capable of adapting the language to their own needs or not.

Moukas (1996) also mentions the distinction between direct and stigmetric communication. Direct communication consists in sending information intentionally to the recipients, while stigmetric communication consists in deducing the behavior of the other agents from environmental cues.

Since our main aim is to explore pragmatic principles of communication, it is important to keep the basic setting as transparent as possible. Hence, the simulations we present in this paper can be characterized as fixed, pre-engineered and direct.

Our setting consists of two agents engaged in a simple game. The turns alternate between the two agents, and their task is to choose one of two alternatives. At the end of each turn the agent tries to communicate an expression of its choice to the other agent, and thus exhibits a cooperative behavior.

One of the alternatives gives a reward, but not the other. To be successful in the game, the agent should try to choose the rewarding alternative all the time. The problem is to know which alternative is better. The agent can base its choice either on previous experiences of the task or on information given by the other agents. However, as the agents from the outset have no common language, the other agent will not know the meaning of the communicated label, and will have to try the task, to figure it out.

A central finding is that the meanings associated with the labels will stabilize when the appropriate strategy is used by the two agents. We have also explored factors that determine when communication will be useful. The first of these is the cost of communication and the second is the rate of change in the environment. If the cost of communication is very small compared to performing the actions, communication will be an interesting alternative, as soon as one is not completely sure whether the world has changed since the last trial.

2 THE EMERGENCE OF A LEXICON

What is required for a common language to evolve? In this section we explore the simple scenario mentioned above where two initially ignorant agents come to agree on the meaning of two labels.

Our starting point for this presentation will be the following situation. An agent finds itself in

front of two closed doors. A prize is placed behind one of the doors and if the agent chooses to open the correct one, it will win the prize. In this case, the agent can choose either door, and the chance of winning is 50 percent.

Now consider a more advanced version of the game. Let us assume that the game is played repeatedly by two players X and Y who take turns at opening the doors. Every time the agent chooses correctly, it will receive a new prize. We will also assume that the correct choice stays the same over a number of trials, that is, if one door was correct on the last trial, it is likely to be correct on the current trial too. Finally, we allow the two agents to exchange a message between each trial. This message is posted on the wall in between the two doors, and must state either 'A' or 'B'.

In a situation like this, it would be useful for the two agents if they could cooperate and tell the other one which alternative is correct. The problem, of course, is that they are not allowed to meet before the game and decide on which alternative to call 'A' and which to call 'B'. To gain anything from the communication, the significance of 'A' and 'B' must be established throughout the game in some way. The central goal of this paper is to investigate strategies of the two agents which will result in consensus regarding the meaning of the two messages.

We have already mentioned the first requirement, the stability of the environment. The correct choice must not change too much. It is obvious that if the probability that the correct choice is altered between two trials is too high, the messages passed between the two agents will not be of any use. We call this *the principle of stability*.

Let us look at this game more abstractly and simply represent it as the two choices in figure 1. Each agent has two alternatives, L and R, which they must choose repeatedly.

L   R

Figure 1. The task consists of iterative choices of either L or R, where one of them is the correct choice.

The interaction between the two agents and the environment is shown in figure 2. The two agents, X and Y, can communicate with each other with the two messages 'A' and 'B'. They both also have the same two possible actions to choose from: L and R. When these actions are performed in the environment, they may result in a reward, r_X and r_Y. We set the reward to 1 if the correct action was chosen and 0 otherwise. The goal of each agent will be to collect as many rewards as possible.

However, we do not assume that the agents use the size of these rewards in their learning. The rewards are only used to evaluate the performance of the agents. In the simple scenario we envision, the agents themselves have no access to these rewards, except that it tells them that the selected alternative is correct. Since the communicating agents are embodied in their environment, this approach avoids the symbol grounding problem (Harnad, 1990).
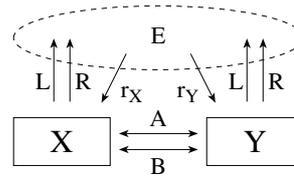
Figure 2. Two agents X and Y communicates about a common task. They can chose between the two words A and B, and at each trial they can perform one of two actions L or R. When the agent chooses correctly, it will receive a reward r_X or r_Y respectively.

To emphasize the role of communication in this task, we will assume that the agents have no memory of the correct choices on the previous trials. The only information they can use is the message sent from the other agent.

Figure 2 also illustrates the second obvious requirement that is needed for language to emerge. The interaction of one agent with its environment has something to say about the actions of the other. This will be called *the principle of a common environment*. In this context, this means that the agents act *as if* there were a common environment. It is the *assumption* of a common environment that makes this principle work, not that it exists in an objective sense.

We now turn to the agents themselves and consider two important questions. What strategies can the agent use to construct the meaning of the two messages, and what structures does the agent need for those strategies? We will start with the second question.

We will assume that each agent structures its experience with the environment and the other agent as a table. The inclination to choose action a when message m is received is represented by the table entry I_{ma}. The agents derive the probability of choosing action a when receiving m, from the formula,

$$p(a, m) = I_{ma} / \Sigma_{i \in M} I_{ia},$$

where M is the set of all messages. To be successful, this strategy requires that the other agent tries to communicate the correct alternative. This will be our first pragmatic principle: *an agent acts as if the other agent tries to cooperate.*

Similarly, to choose the message to communicate the correct choice, each message M is selected according to the probability,

$$p(m, a) = I_{ma} / \Sigma_{j \in A} I_{mj},$$

where A is the set of all actions. This strategy assumes that the agent wants to transmit the correct message to the other agent. This is our second pragmatic principle: *an agent cooperates by trying to transmit the correct message.*

The different inclinations for the case with two action and two messages are shown in table 1.

	Choice	
	L	R
Message A	I_{AL}	I_{AR}
Message B	I_{BL}	I_{BR}

Table 1. Linguistic inclinations.

The simplest way to model update of the table structure is to consider the values in the table as *subjective probabilities*. According to the 'principle of ignorance' (Luce and Raiffa, 1957), all values should be set initially to 0.5, to signify that all inclinations are equal, that is, that there is no reason to select one alternative over the other.

2.1 A Symmetrical Update Rule

We now need a strategy to update the values in the table. The strategy we suggest is that the agent should try to act according to the message it receives. By doing so, it will learn about the consequences of its choice. If the choice is correct, it can assume that the received message should be associated with the performed action and updates its table accordingly. This is our third pragmatic principle: *an agent should try to do what the other agent says.*

By interacting with the environment in this way, the agent will have a chance of learning about the intended meaning of the message. By using this type of strategy, it will be the interaction with the environment and the other agent that serves to structure the lexicon. The agents do not receive any direct positive or negative feedback about word meaning from the other agent.

Let us see how these considerations can be formulated as an update rule for the inclination table. The main idea is that if any tendency to agree on the meaning of a message emerges should be reinforced by further interaction between the agents. This can be described as an update of the inclinations with the values in table 2.

		Choice	
		Correct	Incorrect
Message	Received	$+\delta$	$-\delta$
	Other	$-\delta$	$+\delta$

Table 2. The changes to the values in the table when the agent chooses the correct alternative.

The table describes the changes to the various inclinations when the correct alternative is chosen. No changes are made when the agent chooses incorrectly. We keep the inclinations in the range 0 – 1, where 1 will represent a fully stabilized word meaning. If the value moves outside this range, the value is set to the closest value within the allowed interval. Since the there are only two alternatives and two messages, we can simultaneously update the lexicon for both words.

The value δ describes how fast the agent should change its lexicon. In all our simulations, δ is set to 0.02, but this value is not at all critical. If a smaller value is used, the agent will need a longer time to determine which message indicates which choice. A larger value will make learning faster, but may also cause oscillations in the interaction between the agents if the environment is noisy. As a consequence,

they will never learn a common lexicon. If δ is too small, the values of the matrix will stay close to 0.5 all the time and the probability that the lexicon will stabilize will become very small.

This update rule is obviously not optimal for this two-agent task, but will be much more reasonable in a multi-agent context with more choices and words.

Note that with this update rule, both the received message and the one which was not received are updated in the table. This means that the whole table could in principle be coded by a single parameter. However, we will see below that all the values are needed in the more general case.

To investigate this update rule, we have run an number of computer simulations. In all these simulations, alternative L was the correct one. Figure 3 shows how the inclination to use the message A to mean L develops over time for the two agents. As can be seen, the value for both agents start out at 0.5 and approaches 1.0 in about 200 trials. At this time, both agents have acquired the same lexicon and can successfully communicate about the task. Of course, the meanings of the messages are arbitrary and not determined initially. As a consequence, there are two ways in which a lexicon can stabilize. In some simulations the two agents decide to use the message A for L and B for R, in others they choose the other way around.
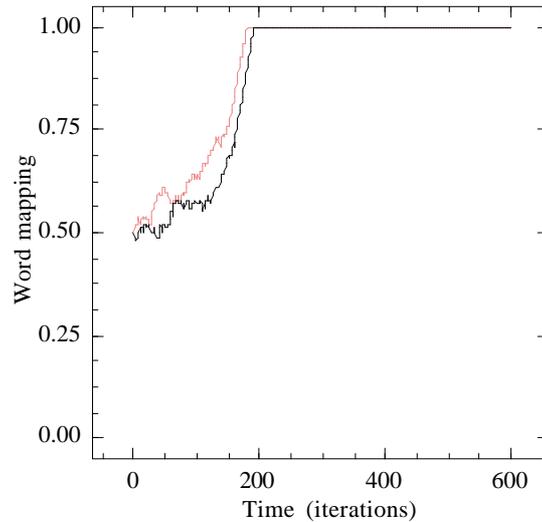
Figure 3 Simulation A1. The development of the inclination to use A for L for the two agents. The lexicon stabilizes after close to 200 trials.

3
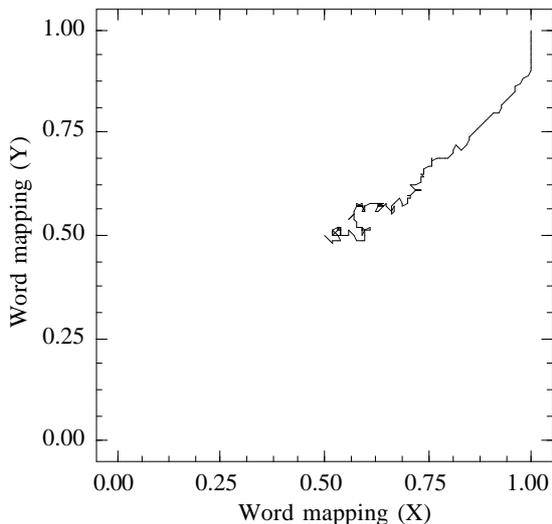
Figure 4. Simulation A2. The development of a single word meaning in the two agents. The word mapping changes in both agents simultaneously indicating that this is a cooperative task.

Figure 4 shows the development of the lexicons for the two agents in relation to each other. The graph shows the value of one agent plotted against the value of the other for the same word-mapping. Both agents learns approximately at the same rate but agent X reaches the stable lexicon slightly faster. The interpretation of this graph is that the establishment of the lexicon is truly a cooperative task. Both agents change their inclinations together. In section 3 below, we will show an example where this is not the case.

To conclude, we have shown that using the three pragmatic principles above, and a simple update rule, a stable lexicon will emerge with which the two agents can communicate about the game.

2.2 An Asymmetrical Update Rule

While the update rule above certainly works, it is unrealistic in one important aspect. It assumes that there are only two alternatives in the environment and that there are only two possible messages. Although this is the case in the simple game we described above, it cannot be true in a more general setting. We must thus assume that there are a large number of possible messages and actions and an update rule cannot rely on this number.

A related problem is that a successful trial where, for example, A is used to mean L is taken as evidence for the association of B with R. This resembles Hempel's paradox where an observation of a non-black non-raven is taken as evidence for the fact that all ravens are black. We certainly do not want an update rule that works in this way.

Fortunately, this problem is easy to overcome. We simply remove the lower right update from the update rule in table 2. The resulting update rule avoids this problem, but not without some sacrifice. Since the update is not done symmetrically, the sums over the rows and columns will no longer be 1. It is, thus, no longer possible to interpret the inclinations as probabilities directly. However, it is easy to derive the desired probabilities when needed, as

described in the beginning of section 2. Since we divide each value with the sum of its row or column, the values can still be used to select appropriate actions or messages.

Another point of concern in the new update rule is the asymmetry between increase and decrease in the table. It is no longer obvious that the value δ should be used both to increase and decrease the values in the table. In a more general setting, there are reasons why these values should be different, but these will not bother us here. We will discuss some of these alternatives in the next section.

Figure 5 shows a simulation using the new update rule. The graph shows how the inclination to use A for L (black) and B for R (gray) develops over time. Remember alternative L was correct all the time. The general conclusion to draw from the figure is that only one of the values stabilize. When the association of A with L reaches its maximum, this word will be used all the time and the value for B and R will not change any more. The value at which the B–R association stays is entirely random.

The interesting property of the asymmetrical update rule is that messages that do not need to be used do not converge to any specific value, that is, the agent does not learn about situations which do not occur. It is thus possible to use an update rule which does not change values for events that do not occur. This is a form of lazy learning where the agents only agree on the meaning of messages they have any use for, which is much more realistic than the previous rule.
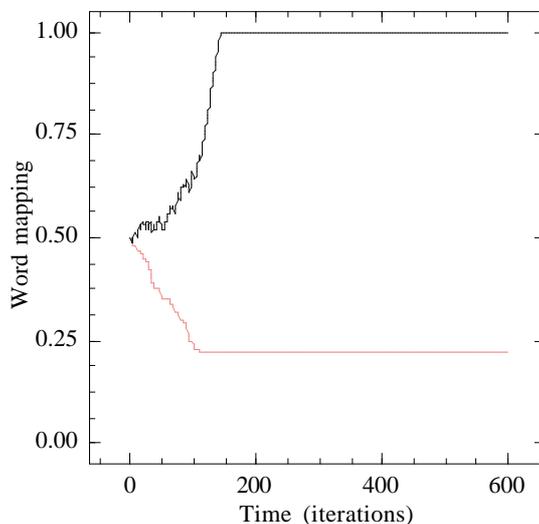
Figure 5. Simulation C (1). The development of the inclinations to use the words A and B for the correct choice. Only the meaning of one of the words stabilizes at 1.00.

2.3 Alternative Rules

The two update rules described above are by no means the only possible. There exist many alternative methods when actions and messages are selected. In this section we will describe some of these alternatives.

An obvious alternative to the additive change to the inclinations in the table is to use a Bayesian

approach instead. In this case, the probability $p(m, a)$ is set to the conditional probability $p(m \mid a)$, which in turn could be derived from counted co-occurrence of m and a. That is,

$$p(m, a) = p(m \mid a) =$$

$$p(m \cap a)/p(a) = N_{ma}/N_a,$$

where N_{ma} are the number of times m and a have been used together, and N_a is the number of times a has been used at all. The probability for choosing a certain action based on the received message could be calculated in a similar way. In the Bayesian approach however, it is necessary that the agents take into account *all* their previous interactions with the environment and each other.

The Bayesian update rule is similar to having a large δ and immediately update the inclinations to either 0 or 1. In our simulations above, this would of course produce an stabilization time of one (1) iteration instead of our typical 200. In this case, the first choice made will have dramatic consequences on the subsequent trials. The construction of the lexicon will no longer be a cooperative process.

In the update rule we have used, the choice is stochastic based on the inclination table. The interpretation of this is that the agent will express its uncertainty by sometimes choosing the 'other' alternative even if there is a marked bias for one of the alternatives. In this simple setting, with absolute knowledge of all the states in the 'world,' this is not motivated, but in a more complex environment exploration of alternatives is necessary. If an agent immediately decides that one combination of a message and an action is correct, it will not learn about other possibilities. This is the well known exploration–exploitation problem (Kaelbling, et al., 1996).

It is, however, possible to bias the choice of message or action to the one with the highest probability. In the extreme case, the agent could choose the alternative with the highest probability all the time, that is, it could use a greedy strategy (Sutton, 1996). This would also be the Bayesian solution. A more moderate strategy could be to use a greedy selection most of the time and to try out other possibilities with some small probability (Sutton, 1996) or to derive some more advanced probability density function from the inclination table. A common method in reinforcement learning is to choose alternatives according to the Bolzmann distribution generated by the individual inclinations (Balkenius, 1995). However, a more advanced simulation will be necessary to explore these possibilities.

3 POWER AND PERSUASION

In the above examples, the establishment of a lexicon was a cooperative task, but does this always have to be the case? Is it possible for one agent to have greater power than the other over word meaning? In this section we show that this can indeed be the case. If one of the agents comes to the game with a preset lexicon, it will be able to convince the other one that it is the correct one. As we will see,

the process is more akin to stubbornness than to real power.
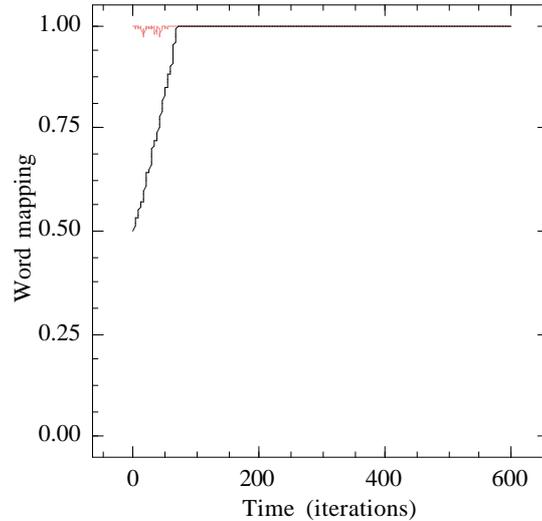
Figure 6. Simulation H (1). When agent X (gray) comes to the game with a predefined lexicon, the other agent will conform in less than 100 trials.
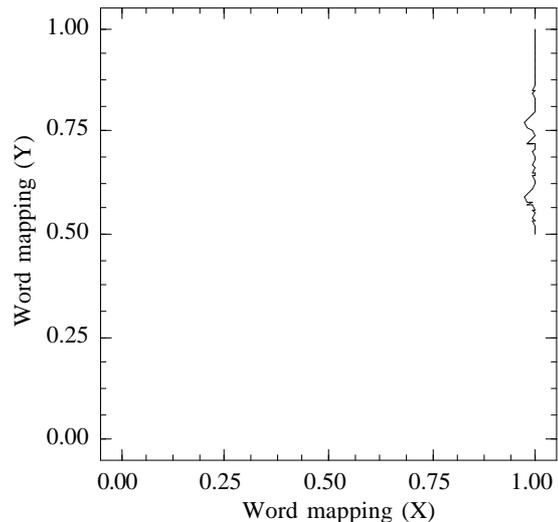
Figure 7. Simulation H (2). The relative change in the lexicons of the two agents. Only agent Y changes its values to any large extent.

We ran a number of simulations where the table for agent X was set to a stabilized lexicon. The lexicon for the other agent was initialized to 0.5 for all values. The resulting simulation is shown in figure 6. The value for agent X (in gray) stays mainly the same all the time while the value for the other agent moves at a high pace toward 1. It is interesting to note that learning for agent X was much faster in this case compared to when they had to cooperate. Since agent X acts as if its lexicon was correct, the game looses one degree of freedom which helps the acquisition of a common lexicon. Figure 7 shows the same simulation in an alternative way as the relative change of each agent. Since agent X is reluctant to change its lexicon, agent Y has no choice than to use the same mapping between words and actions.

These simulations show that the power to decide on word meaning can be modeled simply as an initially larger separation between the different words. Agent X has better ability to discriminate which will be transferred to agent Y as a result of their interaction. This does not mean that agent Y has nothing to say about word meaning. If it manages to construct its own discrimination, it can in principle convince the other agent that this is the correct one, but the probability for this will be very low.

The presented simulation used the extreme case in which one agent had a completely converged lexicon while the other had none at all. In general, the agents can have lexicons anywhere between these two extremes, and their relative influence on the emerging lexicon will be proportional to this. It is also possible for different agents to have lexicons that are more or less converged in different areas. It can know the meaning of some words better than others.

4 A CHANGING WORLD

We saw above that the lexicon used by the agents would only converge for the words that where necessary to solve the task. Since the same alternative was correct all the time, the agents choose to use only a single word. An objection to these simulations is that if the same alternative is correct all the time, it would be easier to remember this instead of trying to communicate with the other agent between each trial.
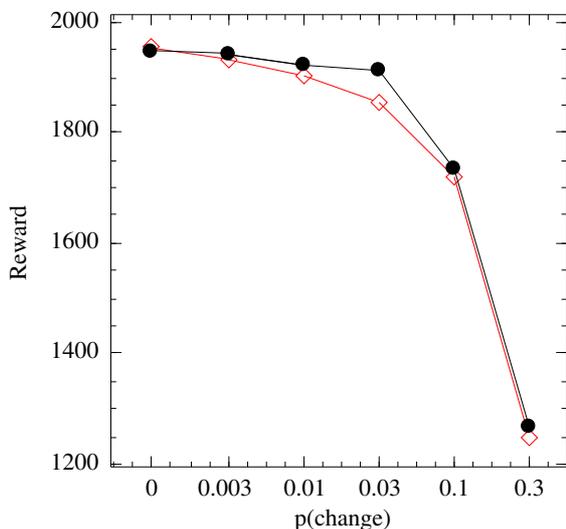
Figure 8. The utility of communication. The filled black dots indicate communication, and the gray unfilled reliance on own experience. The average reward decreases when the environment changes more frequently, but the situation can be overcome to some extent if the agents communicate with each other. N.B. the logarithmic scale. Values are calculated as the mean over 10 runs.

To make communication more useful, we introduce an element of chance in the game. Instead of keeping the correct alternative fixed, we change which action is correct with some probability. In this case, the agents can gain something from communicating with each other. When the correct alternative changes, the first agent to notice this change can inform the other about the change. Because of this communication, the other agent will receive more rewards than if no communication took place. Since the agents cooperate, both will gain something from this communication in the long run.

The basic reason for communicating is that two agents can make more observations than a single one. If they communicate about their findings, both agents will gain more experience than an agent that does not communicate. In the simple game used here, the usefulness of communication is rather limited since the game is so easy, but there is nevertheless a little to gain from communicating with each other.

We ran a number of simulations that tried to address these questions. As could be expected, it turned out that agents that communicates with each other will initially be worse off than agents that do not communicate. The reason for this is that it takes some time for the lexicon to stabilize. During this time, the communicating agents will do a lot mistakes and lose many rewards. The agents that do not communicate will only lose their reward when the environment changes and will gain more rewards during this period.

In the long run, however, the communicating agents will earn more rewards, since on the average they will only miss a reward on every second change of the environment. In the example game used throughout this paper, the effect is very small however, and it did not seem possible to set up a simulation where it would be possible to show this effect in a graph of limited size. Again, we expect this effect to be much larger if more than two alternatives were present.

This is also illustrated in figure 8, where the average reward after 2000 trials are plotted against the probability that the correct choice will change. The black dots indicate the situation where the two agents communicate with each other, while the gray unfilled dots show the situation where each agent only uses its own previous experience to choose. The graph shows that the largest gain of communication is obtained when the world changes. However, the gain diminishes when the changes are so frequent that the agents have not the time to report them to each other before the next change occurs.

When the environment changes, it becomes necessary to send two different messages rather than a single one. Figure 9 shows how the word meaning changes over time. The dotted line shows when the correct alternative changes. When the black bar is drawn in the bottom of the graph, alternative L is correct. When it is draw in the top, alternative R is the correct one. The graph shows the development of the values of I_{LA} (black) and I_{RB} (gray). When L is the correct alternative, the main change is in the value for I_{LA}. When R is correct, the main change is in I_{RB}. This illustrates the general principle that agents communicate about and learn words for the current state of their environment (see figure 5).

5 THE COST OF COMMUNICATION

In the previous section we saw that in a changing world, it can pay to communicate with each other. We did assume, however, that the communication itself was free. What happens if we put a cost on communication? In this case it would be favorable only to communicate when one has something to say. An obvious strategy would be to only communicate if one believes that the other agent does not know the correct alternative. This is summarized in our last pragmatic principle: *an agent should communicate only when its own or the other agent's expectations fail.*

Unfortunately, this requires that each agent keeps a model of the other agent, which seems like a large effort only to avoid some redundant communication. However, there is a simpler way to avoid unnecessary communication. Let us assume that each agent remembers the last message sent or received. It can then compare the message it would otherwise have sent with this previous message and refrain from talking if the two messages are identical.

We thought initially that agents using this strategy would acquire their lexicons at a slower pace than agents that communicate all the time. Our simulations did not confirm these expectations, however. It turned out that the time for the lexicon to stabilize is identical in the two cases. If the agent act according to the principle above, no information is lost even though the rate of communication is much lower. Consequently, there is no change in the speed of convergence of the lexicons.

The simulation shown in figure 10 shows the accumulated reward for agents that communicate all the time (gray) and agents that communicate only when expectations are not met (black). After the characteristic first period when the lexicon is acquired, the two curves take off at different slopes. Since a cost of 0.2 was withdrawn from the accumulated reward every time the agent communicated, agents that only communicate when necessary will bring in more rewards than agents that communicate all the time.
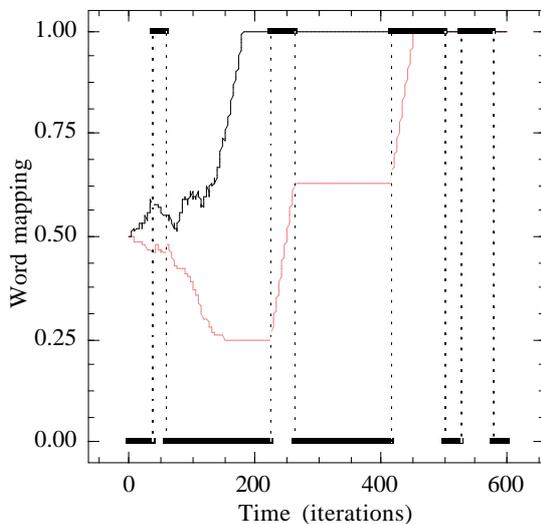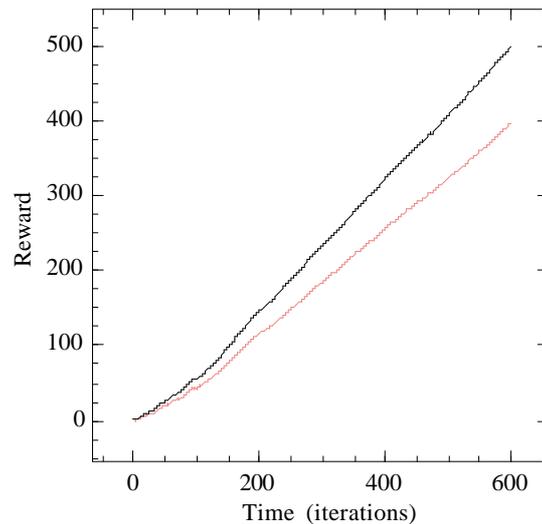
Figure 10. Agents that only communicate when the environment changes bring in more rewards than agents that communicate all the time. The directions of the graphs change approximately at the time when the lexicon has been established.

6 DISCUSSION

The simulations that we have set up are deliberately kept simple. We believe that it is fruitful to discuss some of the fundamental bases of language simulation before getting to a level where such discussions are impossible because of the rapidly increasing complexity. The task so far more resembles the conventionalizing of left- or right-hand side driving, and the changing world corresponds to the government's decision to change this convention, as happened last time in Sweden 1967. When the drivers come to a new road they communicate what they think is the correct lane to the others.

This brings up the issue of *stakes* in the game. If we conceive of the game as one of traffic conventions, it becomes clear that the speed and accuracy of the conventionalizing process is important, as all other cases will lead to inevitable collisions, normally associated with great loss.

In language there are no such strong environmental constraints. If the linguistic emphasis is on descriptive language, as in Hutchins and Hazlehurst (1995) or Steels (1996), the stakes are even smaller, and the connection between linguistic conventions and action is weaker. (In a larger context, the gains of linguistic ability have to be reconsidered, as the largest gain is perhaps the function of language to structure our cognition.)

Our pragmatic approach places itself somewhere in the middle. Our agents are rewarded when they develop a functioning language, but they cannot take advantage of their increasing rewards to change their behavior strategies, only enjoy the reward they are getting.

The traffic analogy also breaks down as soon as we consider extensions of the game. Three directions of growth are obvious: more words, more actions, and more agents. As soon as one of these dimensions is

Figure 9. Simulation J (8). See the text for further explanation.

7

changed, the dynamics of the game will change radically.

To introduce more actions and more words means that the knowledge obtained from an incorrect choice is of much less value than before, and the update system has to be reconsidered. It will also be possible to introduce an asymmetry between the number of words and the number of actions, and to force the agents to assign the words that are needed rather than the words that are available. Following Steels (1996) it would also be possible to let the agents themselves construct and choose words depending on the distinctions that they need to make.

With more possible actions and a more complicated word–action structure, the question of what is meant by a certain label will arise, and the underdetermination of natural language will come into play. The general problem is known as Gavagai after Quine, but since then several *constraints* have been formulated for what meanings can be assigned to. The most well-known of these constraints are the contrast principle and the whole-object assumption. (Baldwin, 1994, Clark, 1987, Markman, 1991)

The introduction of *more agents*, on the other hand, will give rise to "social" problems, where the agents depart from total cooperation. Some interesting issues are:

The introduction of credibility, i.e. the judgement of the predictions of the other agents. This assumes the modelling of the other agents with respect to different factors as well as keeping track of how well a certain agent does in reporting the correct action to the others.

An element of competition. If linguistically transmitted knowledge becomes valuable, agents can be induced to use it as a means in trading. Combined with a credibility model, agents can choose only to communicate with the ones that have shown credible in the past. This will lead to the formation of coalitions.

The formation of coalitions. In a multi-agent environment it is possible to model a situation where isolated language islands emerge, where only some of the agents establish a common lexicon.

Distributed reward. Yanco and Stein (1993), in their simulations of leader–follower communicative behavior, introduced task-based reinforcement so that neither the leader or the follower is rewarded until the task is performed correctly. This idea combined with more free coalitions can be used for the performance of more complex tasks, where the task is impossible to perform for an isolated agent, but where collaborating agents together can perform the task and share the reward.

The purpose of all these extensions is to investigate the possibility to simulate the pragmatic principles found in natural language and studied in e.g. Winter (1994; 1996), Winter and Gärdenfors (1995).

7 SUMMARY AND CONCLUSION

In the preceding, we have given an account for some basic simulations of primitive communicative behavior. In contrast to many other models (for example Hutchins and Hazlehurst, 1995) these are based on a simple table representation rather than on artificial neural nets. This has several advantages. It reduces the complexity and the run-time of the simulations. A typical run takes about 1 second on a Power Macintosh™.

Our simulations are based on direct communication, where the agents' communication is deliberate and distinct from the rest of their behavior. This is in contrast to stigmetric communication, where the agents deduce the information communicated from changes in the environment (Moukas and Hayes, 1996).

The simulations were based on a number of principles which characterize both the environment and the behavior of the agents. The principle of a common environment makes sure that the agents have something to communicate about, while the principle of stability assures that the environment is deterministic enough for communication to be useful.

The pragmatic principles that are modelled in this environment resembles the cooperative principle of Grice (1975). The agent acts as if the other agent tries to cooperate, tries to do what the other agent says, and cooperates by trying to transmit the correct message. When the agents use update rules for their lexica which exploit these principles, the emergence of a common language is based on the cooperation of the agents.

We have strived for a minimal implementation of these principles to allow a clear analysis of the strategies used. In more complicated systems, many interesting properties are obscured by the complexity arising from the interacting principles.

To explore the cost of communication, we introduced a changing environment. In this case, it was favorable for the agents to communicate only when their expectations were not met. It was also possible to model different power over language as an initial difference in the lexica of the two agents. An agent with an initially better discrimination between messages will have greater power over the resulting common lexicon.

REFERENCES

Baldwin, D., (1994), "Update on Inductive Mechanisms for Language Acquisition", Dept. of Linguistics, University of Oregon, Manuscript.

Balkenius, C., (1995), *Natural Intelligence in Artificial Creatures*, Ph.D. thesis, Lund University Cognitive Science, 37.

Clark, E., (1987), "The Principle of Contrast – A Constraint on Language Acquisition". In MacWhinney (ed.) *Mechanisms of Language Acquisition*, Lawrence Erlbaum Ass., Hillsdale, NJ.

Grice, H. P., (1975), "Logic and Conversation". In Cole and Morgan (eds.) *Syntax and Semantics*, Academic Press, New York.

Harnad, S., (1990), "The Symbol Grounding Problem", *Physica*, **D 42**, 335–346.

Hutchins, E. & Hazlehurst, B., (1995), "How to invent a lexicon: the development of shared symbols in interaction". In N. Gilbert and R. Conte (eds.) *Artificial Societies: The Computer Simulation of Social Life*, UCL Press, London.

Kaelbling, L. P., Littman, M. L. & Moore, A. W., (1996), "Reinforcement Learning: A Survey", *Journal of AI Research*, **4**.

Luce & Raiffa, (1957), *Games and Decisions*, John Wiley, New York.

Markman, E. M., (1991), "The whole-object, taxonomic, and mutual exclusivity assumptions as initial constraints on word meanings". In S. A. Gelman and J. P. Byrnes (eds.) *Perspectives on Language and Thought – Interrelations in Development*, Cambridge U P, Cambridge.

Mataric, M. J., (1993), "Designing Emergent Behaviors – From Local Interactions to Collective Intelligence". In S. W. Wilson, J.-A. Meyer and H. L. Roitblat (eds.) *From Animals to Animats II*, The MIT Press, Cambridge, MA.

Moukas, A. & Hayes, G., (1996), "Synthetic Robotic Language Acquisition by Observation". In P. Maes, M. J. Mataric, J.-A. Meyer, J. Pollack and S. W. Wilson (eds.) *Fourth International Conference on Simulation of Adaptive Behavior*, Cape Cod, Massachusetts, 568–579, The MIT Press.

Noble, J. & Cliff, D., (1996), "On Simulating the Evolution of Communication". In P. Maes, M. J. Mataric, J.-A. Meyer, J. Pollack and S. W. Wilson (eds.) *Fourth International Conference on Simulation of Adaptive Behavior*, Cape Cod, Massachusetts, 608–617, The MIT Press.

Steels, L., (1996a), "Emergent Adaptive Lexicons". In P. Maes, M. J. Mataric, J.-A. Meyer, J. Pollack and S. W. Wilson (eds.) *Fourth International Conference on Simulation of Adaptive Behavior*, Cape Cod, Massachusetts, 562–567, The MIT Press.

Steels, L., (1996b), "Perceptually Grounded Meaning Creation", Artificial Intelligence Lab, Vrije Universiteit Brussel, Draft.

Steels, L., (forthc.), "Synthesising the Origins of Language and Meaning Using Co-evolution, Self-organisation and Level Formation". In J. Hurford (ed.) *Evolution of Human Language*, Edinburgh U. P., Edinburgh.

Sutton, R. S., (1996), "Generalization in Reinforcement Learning: Successful Examples Using Sparse Coarse Coding". In *Advances in Neural Information Processing Systems 8*, MIT Press, Cambridge, MA.

Winter, S., (1994), "Förväntningar och kognitionsforskning", Lund University Cognitive Studies, No. 33.

Winter, S., (1996), "Anticipation and Violin Strings", Lund University Cognitive Studies, No. 44.

Winter, S. & Gärdenfors, P., (1995), "Linguistic Modality as Expressions of Social Power", *Nordic Journal of Linguistics*, **18**, 137–166.

Yanco, H. & Stein, L. A., (1993), "An Adaptive Communication Protocol for Cooperating Mobile Robots". In S. W. Wilson, J.-A. Meyer and H. L. Roitblat (eds.) *From Animals to Animats II*, The MIT Press, Cambridge, MA.

Yanco, H. A., (1994), *Robot Communication – Issues and Implementations*, M. Sc. thesis, Massachusetts Institute of Technology.