A cross-situational strategy for damping homonymy in The Naming Game

Joachim De Beule and Bart De Vylder
{joachim,bartdv}@arti.vub.ac.be

Artificial Intelligence Lab
Free University of Brussels
Pleinlaan 2, Brussel, 1050, Belgium

Abstract

Recently there has been a growing interest in the properties and formal analysis of multi-agent systems bootstrapping a communication system (see e.g. .(Baronchelli et al., 2005) and the contribution of (De Vylder and Tuyls, 2005), this workshop.) Although very interesting and promising results were obtained in these studies, major simplifications were made. For example, although much larger populations are considered than was the case in most earlier work, both the cited works (1) assume the possibility of meaning transfer, i.e. the hearer always exactly knows the speaker's intended meaning independent of whether he understands the speaker's utterance and (2) only consider single-word utterances.

In this draft we first consider what happens when relaxing the meaning-transfer assumption, and propose a cross-situational learning scheme that allows a population of agents to still bootstrap a common lexicon under this condition. We empirically show the validity of the scheme and thereby improve on the results reported in (Smith, 2003) and (Vogt and Coumans, 2003) in which no satisfactory solution was found. We then continue to identify some problems that arise when abandoning the single-word utterance simplification. However, no solid solutions are proposed for the identified problems. It is not our aim to reduce the importance of previous work, instead we are excited by recent results and hope to stimulate further research by pointing towards some new challenges.

Draft prepared for the *Workshop on Semiotic Dynamics of Language Games,* Bagnovignoni (Siena - Italy) 8-11 November 2005

Introduction

The use of computer experiments and computational modeling has been a continuous source of interesting results in the scientific fields of language evolution, acquisition and emergence (see e.g. (Steels, 1995), (Cangelosi and Parisi, 2001).) More recently, new technologies and web-applications that support or exploit the self-organization of communication systems are emerging and provide an additional field to which these results and techniques can be applied. However, most of these models lack a solid theoretical foundation, and researchers are only recently taking up this challenge. For example, although many experimental studies have shown the successful emergence of a communication system in a population of multiple autonomous agents in the absence of any central control, only very recently a set of sufficient convergence criteria was formulated supporting this finding (see contribution of (De Vylder and Tuyls, 2005), this workshop.) And it was shown only by (Baronchelli et al., 2005) that the emergence of a successful communication system through self organization scales to large populations, clearly a necessary property for large-scale web-applications.

However, both of these cited works make very strong simplifications. More specifically, both consider only single-word utterances and assume a *meaning transfer*: when a speaker utters a word, the hearer knows what the intended meaning is. These simplifications greatly reduce both the complexity of the language emergence task and the difficulty of analyzing and understanding the dynamics involved.

In the following we first consider the effects of removing the meaning-transfer simplification. We propose a cross-situational learning mechanism that enables the agents to establish a successful and minimal communications system. Although we could not formulate a formal proof, we support this claim with some results obtained from an actual simulation, and thereby improve on the results reported in (Smith, 2003) and (Vogt and Coumans, 2003) in which no satisfactory solution was found.

We then go on to identify some problems that arise when considering multiple-word utterances. It is not our aim to reduce the importance of previous work, instead we are excited by recent results and hope to stimulate further research by pointing towards some new challenges.

The Naming Game without Meaning Transfer

As is the case in the naming game as described in (Baronchelli et al., 2005), the game we consider is played by a population of N agents trying to bootstrap a common lexicon for some set O of objects. At each time-step t two players are randomly selected from the population. One of them is the speaker, the other is the hearer. They are both presented with the current context $C_t \subset O$, which contains a random subset of at least two objects from O. The speaker randomly selects one of the objects in C_t as the topic to which he wants to draw the hearer's attention to. He therefore calculates a word referring to the object and transmits it to the hearer. This may involve inventing a new word. In the following, unless otherwise stated, the object to which a word refers will be called the meaning of that word. Likewise, we will also refer to a set of meanings M, which uniquely identify the objects in a one-to-one fashion. As usual, the interaction is considered a success when the hearer succeeds in identifying the topic chosen by the speaker.

As in the work described by (Baronchelli et al., 2005) or by (De Vylder and Tuyls, 2005), the agents do not directly use the information about the success of the game. In this case, however, this implies that the hearer does not know which object the speaker chose. Other assumptions remain the same, i.e. all agents have an equal chance of interacting and we assume a zero probability that the same word is invented twice. Still, our relaxed assumptions imply the possibility of homonymy and that the complete environment has to be taken into account, i.e. it cannot be reduced to one single object.

An agent will be characterized by its production and interpretation behavior. The production behavior determines the probability with which the agent produces a word for some object. This also includes the mechanism that determines when to introduce a new word. The interpretation behavior determines how the topic is guessed given a context and a word. However, whether the topic is guessed rightly or not does not influence the agents since they do not receive feedback about the outcome of the game. Giving the agents such feedback would again introduce meaning transfer, at least in the case of a successful game.

Thus, each time step t, the only information transmitted between a speaker and a hearer is the fact that the speaker produced a particular word w for one of the objects in the context C_t. This information is insufficient to determine the intended meaning of the word w (the object it refers to) and cross-situational learning is required. If the hearer was to learn a stable language, he could wait until the word w is observed again at time $t' > t$, concluding that the meaning of w should be in $C_t \cap C_{t'}$, and continuing with this strategy until it is exactly known. However, in a lexical emerge task without meaning transfer in which new words are introduced and in which the meaning of existing words may change, this strategy may fail because of inconsistent observations that reduce $C_t \cap C_{t'}$ to the empty set. Thus, a more intelligent cross-situational learning scheme is required. In the following section such a scheme is proposed.

Intelligent Cross-Situational Learning

In this section we propose a learning scheme that is capable of estimating a word/meaning mapping that changes over time from incomplete information. The information consists of consecutive $\langle w, C_t \rangle$ pairs of a word w and a set of objects C_t of which one is apparently referred to by w. Previously, (Smith, 2003) proposed a Bayesian learning mechanism that estimates the probability of some meaning m occurring given the occurrence of a word w in a similar setting. Basically, it consists of storing all co-occurrences of words and meanings. However, such a mechanism on the one hand neglects the possibility that the learning target may change over time and on the other insufficiently uses available information. For example, if an agent at time t observes a word w_1 with a context $C_t = \{m_1, m_2\}$ and, at some later time step t', observes the same word with context $C_{t'} = \{m_1, m_3\}$, it seems logical to conclude that the meaning of w_1 should be m_1. However, only taking co-occurrences into account results in w_1 referring to m_1 with a probability of 0.5 and either to m_2 or m_3 with a probability of 0.25.

The mechanism we propose works independent for different words. Therefore, we will explain the learning scheme for a given word w^* given the subsequent contexts with which it appears. If w^* occurs first at time t with context C_t, the agent associates a probability distribution $s_t : M \longrightarrow [0,1]$ with it, such that

$$s_t(m) = \begin{cases} \frac{1}{|C|} & \text{if } m \in C \\ 0 & \text{otherwise} \end{cases}$$

This implements the fact that all objects in C_t have an equal probability of being the meaning of the word w^*, while all other meanings have a zero probability.

Next, if w^* is observed again at time t' in a context $C_{t'}$, the probability distribution $s_{t'}$ is defined as follows. Let $\gamma = \sum_{m \in C_t} s(m)$ and $\delta = 1 - \gamma$. Furthermore, let $\gamma' = (1-\alpha)\gamma + \alpha$ and $\delta' = 1 - \gamma'$. Then

$$s_{t'}(m) = \begin{cases} \beta(\gamma) s_t(m) \frac{\gamma'}{\gamma} + (1 - \beta(\gamma)) \frac{\gamma'}{|C_t|} & \text{if } m \in C_t \\ s_t(m) \frac{\delta'}{\delta} & \text{if } m \notin C_t. \end{cases}$$

with $\beta(\gamma) = \sqrt{1-(1-\gamma)^2}$, a definition which is motivated further on.

In words, γ is the total probability of all meanings consistent with the current observation (all objects in C_t). At time t', this probability is increased to $\gamma' \geq \gamma$ according to the parameter α. As such, α represents the strength with which the new information at time t is validated more important then the information gathered before time t. Furthermore, this new probability γ' should be distributed among the consistent meanings such that if the new information is in agreement with the current state (γ close to 1), then the relative probabilities among the consistent meanings should be more or less conserved (i.e. strong associations remain strong and weak ones remain weak). Therefore we require $\beta(1) = 1$. However, if the new information is not in agreement with the current state (γ low), then we want γ' to be more or less distributed evenly among all meanings in C_t.

Therefore we also require $\beta(0) = 0$. Moreover, we want that all scores of objects in C_t increase if $\gamma < 1$, because this guarantees convergence to a unique interpretation if the contexts are random but always contain a certain object. It is easily verified that a necessary condition for this is that

$$\beta(\gamma) > \frac{\gamma}{\gamma'} = \frac{\gamma}{(1-\alpha)\gamma+\alpha}$$

for $\gamma < 1$. From this it follows that $\beta'(1) \leq \alpha$.[1] In order for the update mechanism to work for all values of α, we chose $\beta'(1) = 0$. The specific definition of β given meets all of these requirements.

In any case, the total probability δ of inconsistent associations is weakened to $\delta' \leq \delta$.

We will refer to this updating mechanism which transformed s_t in $s_{t'}$ as a function u such that

$$s_{t'} = u(s_t, C_t).$$

To illustrate this estimation function, assume that $\alpha = 0.3$ and that $C_t = \{m_1, m_2\}$ and $C_{t'} = \{m_1, m_3\}$. Then initially $s_t(m_1) = s_t(m_2) = 0.5$ and $s_t(m_3) = 0$. After observing $C_{t'}$ we have $s_{t'}(m_1) \simeq 0.61$, $s_{t'}(m_2) \simeq 0.35$ and $s_{t'}(m_3) \simeq 0.04$.

Agent Architecture

At time-step t, an agent can be described by a tuple $\langle W_t, \sigma_t, \phi_t \rangle$. W_t is the set of words the agent has encountered until that moment. $\sigma_t : W_t \times M \longrightarrow [0,1]$ is a function which associates meanings with words, such that, $\sigma_t(w,m)$ gives the agent's estimation of the probability that word w means m. It might seem that the agents would have to know all the possible meanings M in advance, but this is not the case: as can be verified in the following, σ_t will always be zero for meanings not yet encountered.

$\phi_t : W_t \longrightarrow [0,1]$ is a function which associates scores with words, which will be used to dampen synonymy. Initially, for each agent holds $W_0 = \emptyset$.

Before explaining the production and interpretation behavior of an agent and the way he updates his internal state, we first define the interpretation function of an agent $\langle W_t, \sigma_t, \phi_t \rangle$ as the function $g : W_t \longrightarrow M$ with

$$g(w) = \text{argmax}_{m \in M} \sigma(w,m).$$

If multiple meanings in M have a maximum value, one is chosen at random. Therefore $g(w)$ is possibly a stochastic value.

Production Suppose that the speaker at time t is given by $\langle W_t, \sigma_t, \phi_t \rangle$, the context is C and the topic he will speak about is $m^*(\in C)$. The production behavior of an agent will not depend on the context. The speaker searches for words $W' \in W_t$ which according to him interpret as m^*: $W' = \{w \in W_t | g(w) = m^*\}$. As $g(w)$ can be stochastic, also can W'. If $W' = \emptyset$ (which is always the case if m is encountered for the first time) the speaker invents a new word $w^*(\notin W_t)$. If $W' \neq \emptyset$ then he chooses the word w^* from W' with the highest score: $w^* = \text{argmax}_{w \in W'} \phi_t(w)$. Again, if multiple words have a highest score, one is selected at random.

Only if the speaker invented a new word he updates his internal state. Obviously, $W_{t+1} = W_t \cup \{w^*\}$. Furthermore, the word-meaning scores of known words stays the same, but additionally we have

$$\sigma_{t+1}(w^*, m) = \begin{cases} 1 & \text{if } m = m^* \\ 0 & \text{otherwise} \end{cases}$$

Finally the new word gets score 1: $\phi_{t+1}(w^*) = 1$ (scores for other words are left unchanged.)

To summarize, the speaker has produced a word w^* for meaning m^* in context C, thereby possibly changing its internal state. As will be described next however, the major state change occurs at the hearer side.

Interpretation Suppose that the hearer at time t is given by $\langle W_t, \sigma_t, \phi_t \rangle$, the context is C and the word received is w^*. If this word is unknown to him ($w^* \notin W_t$) then we obviously have $W_{t+1} = W_t \cup \{w^*\}$ and the word-meaning association scores for w^* are initialized as follows:

$$\sigma_{t+1}(w^*, m) = \begin{cases} \frac{1}{|C|} & \text{if } m \in C \\ 0 & \text{otherwise} \end{cases}$$

If $w^* \in W_t$, the association scores involving w^* are altered according to the updating function u defined before:

$$\sigma_{t+1}(w^*, \cdot) = u(\sigma_t(w^*, \cdot), C).$$

We now describe the updating of the word-scores, using the auxiliary definition $\phi_t(w^*) = 1$ if $w^* \notin W_t$. First, the interpretation m' of w^* is determined as $m' = g(w^*)$ (with g using σ_{t+1}). Next, the set of synonyms S for w^* is determined as those words in $W_{t+1} \setminus \{w^*\}$ which also have interpretation m' (according to g). Finally, the score of w^* is increased: $\phi_{t+1}(w^*) = (1-\theta)\phi_t(w^*) + \theta$, and the scores of the synonyms are 'lateraly inhibited': $\phi_{t+1}(w) = (1-\theta)\phi_t(w)$ for $w \in S$. The other scores remain the same.

Experiments

Measures

In order to get insight in the way the population of agents comes to agreement about the emerging language, we define some measures on the population's state. This population state can conceptually be summarized in a semiotic graph. Such a semiotic graph is a bipartite, directed graph in which nodes represent meanings and words and in which edges only go from meaning nodes to word nodes or vice versa. An edge going from a meaning to a word

[1] β' is the derivative

node represents a possible production and an edge going from a word to a meaning node represents a possible interpretation (without context.) Each edge is weighted with the weights representing the probability to observe the associated production (resp. interpretation) when a randomly chosen speaker (resp. hearer) produces (resp. interpret) the meaning (resp. word) represented by the starting node of the edge. The sum of the weights of the outgoing edges of a node is at most one, but may be lower. This because there is the possibility that an agent has not (yet) encountered a certain object or word, in which case no production or interpretation is done (only when building the semiotic graph, not during a game). In figure 1 an example of a simple semiotic graph is shown.

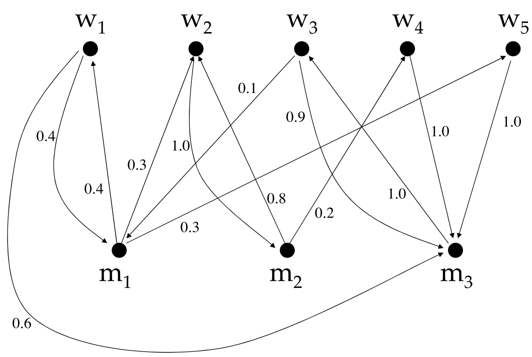

Figure 1: Example semiotic graph representing the state of a population at some point in time. The nodes marked m_i represent a meaning, those marked w_i represent words.

As a first measure we define the **return** of a graph as the probability that, starting from a random meaning node, one returns to that node after taking two steps (thus first going to a word node and then back to a meaning node), with the probability of taking an edge equal to the edge's weight. For example, in figure 1, the chance for returning to m_1 starting from it is 0.16, for m_2 it is 0.8 and for m_3 0.9. Thus the return of this graph is the average 0.62. If a hearer would not take into account the context to guess the meaning of a given word, then the return equals the communicative success.

We will also define two other measures for the amount of synonymy and homonymy present in the graph. For this the notion of the **effective out-degree** of a node is needed, which is related to the number of edges starting in the node, but takes into account the weights of these edges. We assume that the weights of the outgoing edges are normalized such that their sum equals 1. If there are k edges each having an equal probability $1/k$, then the effective out-degree equals k. If, however, the k edges have differing probabilities, then the effective out-degree should be lower. Moreover, if one of the edges has a probability close to one, the effective out-degree should also be close to one (but still slightly higher). Therefore we define the effective out-degree of a node as the number of edges with equal probability needed for a hypothetical node to have the same associated Shannon information as the original node. More precisely, if a node has k outgoing edges with weights $x_i, 1 \leq i \leq k$, its Shannon information is given by

$$\sum_{i=1}^{k} -x_i \log(x_i).$$

A node with k' outgoing edges with equal probability $1/k'$ thus has a Shannon information of $\log(k')$. By definition this information should be equal to the information associated with the original node, from which it follows that

$$k' = \exp(\sum_{i=1}^{k} -x_i \log(x_i)).$$

The effective out-degree k' is not necessarily integer. For example, the effective out-degree of m_1 in figure 1 is 2.97 and of w_3 it is 1.38.

We now define the **synonymy** present in a semiotic graph as the average effective out-degree of the meaning nodes. The synonymy present in the graph in figure 1 is thus 1.87.

The **homonymy** present in a graph is defined as a weighted average of the effective out-degrees of the word nodes, where each node is weighed with its probability of being the result of a production. This probability is proportional to the sum of the node's incoming edges. The homonymy of the example graph is 1.26.

A final measure is the number of words present in the graph. This measure gives an indication of the parsimony of the language. Ideally, the number of words should equal the number of meanings.

Results

We will now present some results of a simulation involving $N = 20$ agents evolving a lexicon to communicate about 100 objects ($|O| = |M| = 100$). Each interaction the speaker and the hearer are presented with a context containing 5 objects ($\forall t \geq 0 : |C_t| = 5$.) The evolution over time of the return, the number of words used by the agents in the population, the homonymy and the synonymy are presented in figure 2. As can be seen, the agents eventually reach a coherent successful language without synonyms or homonyms.

The return, which is related to the communicative success (and necessarily equivalent with it when is 1), starts at a small (chance level) value. As new words are introduced and as their meanings starts to settle, the return gradually increases until finally it becomes maximum at around $t = 105000$.

The maximum number of words present in the population is reached approximately at $t_{\max} \simeq 10000$ and is equal to 1194, which is of the order of $N|O|/2 = 1000$ as would be expected. The cross-situational learning mechanism as proposed in the previous sections allows the agents to eliminate homonyms, which partly explains why the number of words decreases steadily after $t_{\max}$.

In contrast to the findings in (Vogt and Coumans, 2003), our agents do reach a complete coherent language, where coherence is defined as the chance of two random speakers producing the same word for the same meaning. The main difference between Vogt's agents and the ones defined in this paper is the use of a synonymy-damping mechanism which explains further why the number of words used by the agents eventually drops to the number of objects $|O|$.

The Multi-Word Guessing Game

In this section we briefly touch on some aspects of the syntax-free multi-word guessing game. In such a game, agents are allowed to utter multiple words at once and the meaning of an utterance is defined as the union of the meanings of its constituent words. Obviously, the transmission of multiple words is only a possibility when the meaning to be expressed consists of multiple and independent parts.

In the previous sections we silently assumed that the meaning of a word could be identified with the object in the world (or context) referred to by the word. However, in reality things are more complicated, as is also illustrated by (Quine, 1960) in his famous 'gavagai' example: what exactly would a speaker of an unknown language refer to when uttering 'gavagai' while pointing to a fleeing rabbit? In other words, what could be the meaning of the word gavagai?

More formally, assume that the observation of an object or event in the world results in a collection of categories C, for example containing the object's color, size, position etc. In this case, even if the world or the context presented to a speaker and a hearer playing a naming game would only contain a single object, still the hearer would not be able to determine the meaning of a word describing it if any combination of the observational categories is a valid candidate.

If a speaker is only allowed to communicate one category at a time and under the assumption that the speaker and hearer both have access to the same set of categories, then the game would be equivalent to the naming game without meaning transfer as described in the previous sections.

However, if a speaker is allowed to select multiple categories from C_t to assemble the meaning he will communicate, then the possibility for multiple word utterances arises. To see why multiple categories might be needed, assume for example that there are categories representing the color and size of objects. Now if the current context contains three objects of which the first is big and red, the second big and blue and the third is small and red, then if the speaker wants to draw a hearer's attention to the big red object no single category suffices and both the categories big and red need to be communicated.

One strategy that could be adopted by the agents is to always use single-word utterances. This would result in a holistic language in which all possible meanings (i.e. combinations of categories) are expressed by a different word. This is in contrast to a compositional language in which different fragments of meaning would be expressed by different words. However, such a holistic strategy might have a number of disadvantages like increased lexicon size, decreased learnability, less robust against the presence of noise in transmission etc (see e.g. (Nowak and Krakauer, 1999; Kirby, 2000; Brighton, 2002; Smith and Kirby, 2003)). On the other hand, it might be beneficial to express certain combinations that occur very often in a holistic fashion for the same or other reasons. In any case, the possibility to at least partially evolve a compositional language can be considered.

Thus, the information now transmitted between agents are consecutive $\langle U_t, C_t \rangle$ pairs of multi-word utterances $U_t = \{w_1, ..., w_u\}$ and sets of categories $C_t = \{c_1, ..., c_c\}$, with $u \leq c$. In the next sections we will briefly describe some additional problems that arise in this setting compared to the case when only single word utterances are considered.

The Meaning Assignment Problem

One of the difficulties that needed to be solved after relaxing the meaning transfer assumption in the naming game was how to determine the meaning of an utterance. This problem persists in the case of the multi-word guessing game. In addition, a second problem now arises: even if the meaning of an utterance is correctly determined, the meaning of its parts (the individual words constituting the utterance) is still unknown. For example, if the meaning of a particular utterance $U = \{w_1, ..., w_n\}$ is determined to be $M_U = \{c_1, ..., c_n\}$, this still leaves $n!$ possible word/meaning associations or meaning assignments.

The Credit Assignment Problem

Assume that some interaction involving a set $A = \{\langle w_i, m_i \rangle, i = 1..n > 1\}$ of used (in the case of a speaker) or inferred (in the case of a hearer) word/meaning associations was found to be unsuccessful, for example because the utterance did not have the intended effect on the hearer. The agent's now face the problem of determining what elements of A were responsible for the failure. If they would neglect this problem and for example simply punish all associations used, also the successful ones would be punished. This could prohibit the agents to evolve a successful communication system, as even successful elements of language are demoted and thus might not have a chance to become part of the emerging language.

Deciding when to use compositionality

Assume that a speaker wants to express some meaning $M = \{c_1, \ldots, c_n\}$. Assume further that he did not encounter any of the categories in M before. He thus should invent at least one new word, namely one covering the entire meaning M. However, subsequent interactions with the environment could reveal that it would be beneficial to have several words, individually covering only subsets of M but together covering it completely.

Another strategy could be to always introduce different words for all separate categories. However, this would rule out the possibility to have specialized (holistic) words for frequently occurring structured meanings. In addition, this presupposes that all agents are able to independently divide their observational spaces in some shared and hence universal set of atomic categories. Third, this strategy would probably not exhibit the fastest possible convergence rate since many new elements of language (words) are introduced and thereby also many uncertainties that need to be resolved. Finally, the meaning of newly acquired words by children seems to be holistic initially and only later become more specialized (i.e. more re-usable), which makes this strategy less interesting from a cognitive modeling point of view.

In any case, in a multi-word guessing game it needs to be spelled out how to fragmentize some previously unencountered meaning. This problem even persists when the entire meaning to be expressed is covered since it could always be that the wrong choice was made or that another agent made another choice.

How to reach a minimal language

In the naming game without meaning transfer a sort of lateral inhibition was introduced in order to achieve a competition between synonyms (the purpose of the function ϕ in section 'Agent Architecture'.) This is needed to reach coherence in the population, i.e. to enforce that all agents eventually speak the same language instead of a set of languages.

However, in a multi-word guessing game an additional type of competition is needed: namely between holistic and compositional ways to express the same structured meaning. If this is neglected then the resulting language could be highly redundant in a comparable way as a language containing many synonyms is.

A possible strategy might be to demote all applicable but unused verbalizations: much in the same way that synonyms compete for the same meaning, so could any two verbalizations expressing the same meaning be considered competing for that meaning.

Conclusion

We have identified three problems that arise when relaxing the meaning transfer assumption in a naming game. First, since during a single game a hearer can no longer determine the meaning of a word, a cross situational learning mechanism is required. We have proposed such a mechanism that is capable of estimating a word/meaning mapping that changes over time. Second, the conditions that determine when a speaker needs to invent a new word need to be extended beyond the obvious case of uncovered meaning: also when he does not know a word which he himself would *interpret* as the target meaning should he invent a new one. Third, for a coherent language to emerge some synonymy-damping mechanism is needed implementing a kind of lateral inhibition between competing synonyms.

We have presented the results of a simulation of the proposed mechanisms and which empirically illustrates the validity of the scheme. We have thereby improved on the results reported in (Smith, 2003) and (Vogt and Coumans, 2003) in which no satisfactory solution was found.

We have also identified four problems that additionally arise when multiple word utterances are allowed: (1) The meaning assignment problem, i.e. the problem of deciding which words cover what parts of an utterance's meaning; (2) The credit assignment problem, i.e. how to separate successful from unsuccessful elements of language in a globally unsuccessful communication attempt; (3) The need to spell out a mechanism that determines how to fragmentize meaning; and (4) The need for a more general damping mechanism that not only dampens synonyms but also possibly otherwise competing verbalizations. Although no solid solutions to these problems were proposed, we hope that we succeeded in providing a set of interesting challenges.

References

Baronchelli, A., Felicci, M., Cagioti, E., Loretto, V., and Steels, L. (2005). Sharp transitions towards shared vocabularies in multi-agent systems. arXiv:physics/0509075 v1.

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

Cangelosi, A. and Parisi, D., editors (2001). *Simulating the Evolution of Language*. Springer Verlag, London.

De Vylder, B. and Tuyls, K. (2005). Towards a common lexicon in the naming game: The dynamics of synonymy reduction. In *Workshop on Semiotic Dynamics of Language Games*, Bagnovignoni, Siena, Italy.

Kirby, S. (2000). Syntax without natural selection: How compositionality emerges from vocabulary in a population of learners. In Knight, C., Hurford, J., and Studdert-Kennedy, M., editors, *The Evolutionary Emergence of Language: Social function and the origins of linguistic form*. Cambridge University Press.

Nowak, M. A. and Krakauer, D. C. (1999). The evolution of language. *Proc. Nat. Acad. Sci. USA*, 96:8028–8033.

Quine, W. (1960). *Word and Object*. The MIT Press, Cambridge, MA.

Smith, A. D. M. (2003). *Evolving Communication through the Inference of Meaning*. PhD thesis, Theoretical and Applied Linguistics, School of Philosophy, Psychology and Language Sciences, University of Edinburgh.

Smith, Kenny and, B. H. and Kirby, S. (2003). Complex systems in language evolution: the cultural emergence of compositional structure. *Artificial Life*, 9(4):371–386.

Steels, L. (1995). A self-organizing spatial vocabulary. *Artificial Life Journal*, 2(3):319–332.

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

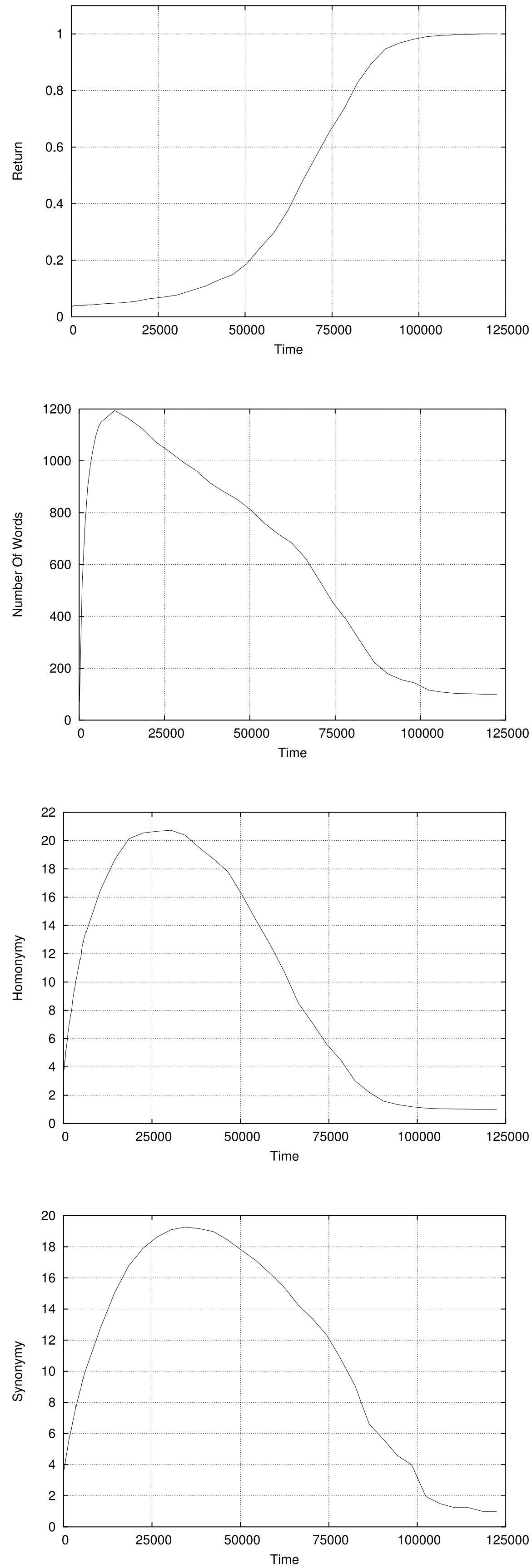

Figure 2: Evolution over time of the return, the number words used by the agents in the population, the homonymy and the synonymy as defined in the text. The graphs were obtained for $N = 20$ agents, $|O| = |M| = 100$ objects and context sizes $|C_t| = 5, t \geq 0$. The forgetting parameter α was set to 0.2 and the synonymy inhibition parameter θ was 0.3.