Bootstrapping grounded symbols by minimal autonomous robots

Paul Vogt
IKAT / Infonomics - Universiteit Maastricht
P. O. Box 616 - 6200 MD Maastricht - The Netherlands
p.vogt@cs.unimaas.nl

Abstract

In this paper an experiment is presented in which two mobile robots develop a shared lexicon of which the meanings are grounded in the real world. The robots start without a lexicon nor shared meanings and play language games in which they generate new meanings and negotiate words for these meanings. The experiment tries to find the minimal conditions under which verbal communication may begin to evolve. The robots are autonomous in terms of computing and cognition, but they are otherwise far simpler than most, if not all animals. It is demonstrated that a lexicon nevertheless can be made to emerge even though there are strong limits on the size and stability of this lexicon.

1 Introduction

Most research on the evolution of language concentrates on the emergence of syntax (Knight et al. 2000). But, one of the first and possibly most important prerequisites for the evolution of human languages is that humans evolved the capability to form symbols (Deacon 1997). This can be inferred from the fact that human language is highly symbolic. The question how symbolic communication emerged is often overestimated in its complexity and it could be argued that the transition towards syntactic communication may not be so difficult once symbolic communication is in place as shown by simulations on the emergence of syntax, e.g., (Kirby 2001). Although communication as such does not need to be symbolic - ants for instance communicate non-symbolically -, humans tend to communicate with symbols. The

ability to form, manipulate and interpret symbols is what makes humans cognitive agents. This is conform the physical symbol system hypothesis (Newell and Simon 1976). Humans think, reason, act and communicate in symbols. However, sensation of the world is mainly a non-symbolic event. Yet humans are very well capable in constructing symbols from these non-symbolic events. Moreover, humans are possibly the only known species that learn a symbolic communication system.

There are other species that have symbolic communication, but this is most likely innate, for instance, the alarm calls of vervet monkeys (Seyfarth and Cheney 1986). Nonhuman species that have learned communication systems tend to communicate non-symbolically. Although some scientists believe that human symbolic communication is innate too, many others believe that these symbols are learned (Bloom 2000; Ellman 1993; Tomasello 1999). In this paper the assumption is adopted that the evolution of symbolic communication can be viewed as an adaptive complex dynamical system that evolved culturally in a similar way ant paths are formed (Steels 1997). Words and meaning are thought to co-evolve based on three mechanisms: (cultural) interaction, individual adaptation and self-organisation. Based on these mechanisms agents interact with their environment in order to sense their environment and to communicate. They adapt their memories to develop word-meaning associations about things they sense. Together, the interactions and adaptations ensure a self-organisation of a lexicon that is grounded in reality. One of the biggest unresolved questions concerning the development of grounded symbols is how can agents learn what is meant with the utterances of speakers? This question is relevant for both artificial agents and humans.

In the past decade different aspects of language evolution have been studied computationally, for an overview see, e.g., (Briscoe 2001; Cangelosi and Parisi 2001; Steels 1997). Most of this work is based on the adaptive complex dynamical systems approach. The studied aspects include simulations on lexicon formation (Hurford 1989; Steels 1996a; Oliphant 1999), meaning formation (Steels 1996b), phonetics (De Boer 2000; Redford et al. 2001), concept formation (Cangelosi et al. 2000; De Jong 1999) and the emergence of syntax (Hashimoto and Ikegami 1996; Kirby 2001). In addition some experiments on physical robots have been done to investigate the emergence of lexicons that are grounded in reality (Steels and Vogt 1997; Steels et al. 2002; Steels and Kaplan this issue; Vogt in press).

This paper is based on the robotic experiments reported previously in (Steels and Vogt 1997; Vogt 2000b). In these experiments two minimal autonomous robots try to develop a set of symbols that are grounded in reality. Initially, the robots

have no symbols at all, these are all constructed during the experiments. The term *minimal autonomous robots* is used to indicate that the experiments are done with autonomous robots, which have a very limited physical architecture and operate in a very limited environment. It has been shown in the previous publications how these robots could develop such a set of symbols. This paper investigates what the minimal conditions are for these robots to develop a shared set of grounded symbols, where the main focus is on *physical* conditions. In the end these conditions may shed light on why humans can develop a shared and grounded symbol system very well, while other species cannot. Other conditions relating to, for instance, population dynamics, learning abilities, parameter settings and the impact of non-verbal cues have been investigated elsewhere, see, e.g., (Steels et al. 2002; Vogt 2000b; Vogt 2001).

The conditions that are investigated basically involve conditions regarding the physical bodies of the robots, their sensorimotor skills and their environment. In most research, such conditions are assumed to be given and are not part of the experimental setup. In embodied systems, such as robots, humans and other living animals, these conditions are considered to be extremely important. As it is assumed that language, and cognition in general, is both embodied and situated - conform (Barsalou 1999; Clancey 1997; Lakoff 1987; Pfeifer and Scheier 1999) -, it is important to investigate the impact of such conditions when investigating the origins and evolution of language. One reason is that the nature of symbols acquired by embodied agents depends to a high degree on the architecture of their bodies (Barsalou 1999; Lakoff 1987). It would therefore be naive to expect that robots with minimal bodies will acquire human-like symbols. In addition, because the symbols are acquired from the interactions of agents with their environment, both the sensorimotor skills of the agents and the complexity of the environment influence the ability to ground symbols as well as the nature of these symbols. That these factors are crucial can also be seen in the research with the Sony AIBO - a four legged pet robot (Steels and Kaplan this issue) and with the LEGO robots used in (Billard and Hayes 1997; Billard and Dautenhahn 1999). In both studies experiments are done in which more or less minimal autonomous robots develop a shared set of grounded symbols.

To deal with a dynamical changing environment and the instability of the robots, the experiment is designed following the *behaviour-based approach* towards robotics (Arkin 1998; Brooks 1990; Steels and Brooks 1995). According to this approach behaviours - including symbolic communication - are acquired or designed from the

bottom-up. The approach strongly relates to the situated view of cognition.

This paper is organised as follows: The next section will briefly introduce the definition of symbols that has been adopted for this work. Section 3 will explain what a minimal autonomous robot is and briefly indicate the consequences of working with such robots with respect to bootstrapping grounded symbols. The language game model by which the robots develop a lexicon is described in section 4. Section 5 presents some experimental results, which are discussed in section 6. Finally, section 7 gives conclusions.

2 Semiotic symbols

When robots try to develop a shared lexicon of which the meaning is grounded in the real world, they need to solve what Harnad (1990) has called the symbol grounding problem. The symbol grounding problem deals with the question how seemingly meaningless symbols become meaningful in the real world (Harnad 1990). This problem is fundamental when symbols are viewed in the traditional sense (Clancey 1997; Pfeifer and Scheier 1999). In the traditional sense symbols are sometimes names for categories (Harnad 1993) or sometimes associations between names and meanings as in the Saussurean sign (De Saussure 1974)[1].

In this paper the definition of symbols provided by the theory of semiotics as introduced by Peirce (1931) is adopted. According to Peirce, a symbol can be viewed as a sign when certain conditions are met. A sign in the Peircean sense consist of three elements (Chandler 2001):

Representamen: The form which the sign takes (not necessarily material).

Interpretant: The sense made of a sign.

Object: To which the sign refers.

Rather than using Peirce's terms, the terms adopted in this article are *form* for representamen, *meaning* for interpretant and *referent* for object. According to Peirce, a sign becomes a symbol when its form, in relation to its meaning is arbitrary or purely conventional. As a consequence, the relationship must be learnt (Chandler 2001). The relation can be conventionalised in language. To distinguish this definition of symbols from the traditional definition, they will be called *semiotic symbols.*

[1] In Saussure's terminology, the name is called signifier and the meaning is called the signified.

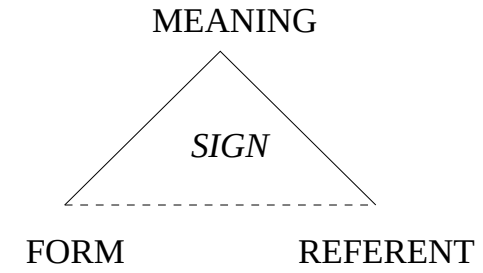


Figure 1: A semiotic triangle shows how a referent, meaning and form are related as a sign.

How the three units of the sign are related is often illustrated with the semiotic triangle as introduced by Ogden and Richards (1923) and shown in figure 1. It is assumed that the meaning of a semiotic symbol depends on how the relation between form and referent is established by the agent. The meaning is represented by a memorised category that becomes activated through the interaction of an agent with the referent and form. Referents are considered as real world entities, which could be physical objects, actions, features like colours, abstract notions or even other semiotic symbols. In this paper only physical objects will be considered as referents; however, most of the discussions will apply to other notions of referents as well. Forms are in principle arbitrary of shape, so they could be anything. In communication systems the forms include gestures, spoken utterances, written forms, etc. In the robotic implementation these are strings of letters taken from the English alphabet.

Philosophically, it could be argued that a semiotic symbol is *per definition* grounded, because the symbol already has a meaningful intrinsic relation with the referent, and hence with the real world, see (Vogt in press) for a discussion. Thus the symbol grounding problem is perhaps not a philosophical problem, but only a technical problem that focuses on the construction of semiotic symbols.

3 Minimal autonomous robots

Minimal autonomous robots are physical robots that operate autonomously in a certain minimal environment and that have a minimum of sensorimotor equipment with respect to the task at hand. The robots are autonomous in the sense that their behaviour is not controlled by a human[2].

[2]In the experiment the human experimenter does intervene sometimes in order to speed up the experiment. However, these interventions do not influence the autonomy of the robots with

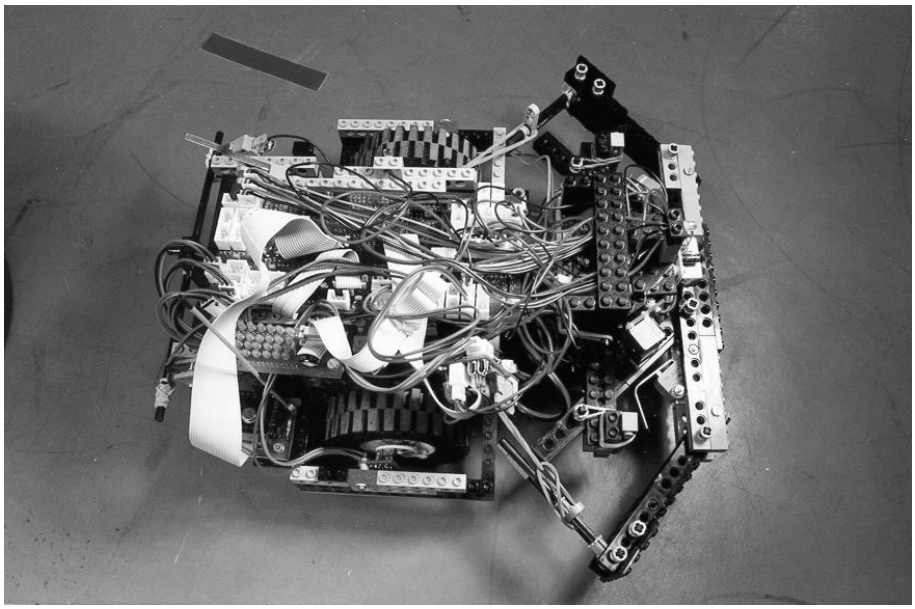


Figure 2: One of the robots that is used in the experiment.

The environment is minimal in terms of its complexity. It is a closed squared arena of $6.25m^2$ with a flat surface. The objects are four light sources placed at different heights and which emit light through a narrow horizontal slit. The ambient lighting conditions change over time, for instance because the intensity of daylight fluctuates. The objective of the experiment is that the robots develop a shared set of semiotic symbols referring to the light sources.

Using minimal robots means that no sophisticated robots such as humanoids or fine-tuned machines with state-of-the-art high-tech sensorimotor equipment are used. The robots' physical bodies come closer to those of insects, but are far more limited.

For the experiment two LEGO vehicles, equipped with sensors, motors, a radio module and an on-board computer are used, see figure 2. The robots have very limited sensory capabilities, consisting of light sensors (LDRs), which only detect the intensity of incoming light. The light sensors are mounted at different heights, corresponding to the light sources in their environment so that these can be distinguished properly. In addition to the light sources, the robots have infrared (IR) sensors and bumpers that are used for obstacle avoidance and for control of their physical behaviour, which is necessary to interact with each other and their environment. All these sensors are unreliable in the sense that they are subject to a lot of noise that comes both from the environment and the electronic hardware.

respect to the objective of the experiment.

Two wheels that are independently driven by two motors, control the robots' movements. To ensure stability, the robots have also a caster wheel. The motors that drive the robots are connected to the wheels by a series of gearings. This whole motor system is unreliable, because the gears are subject to wear, the motors to noise and the wheels to slippage and friction on the flat surface.

The physical behaviours of the robots are synchronised using the radio module. This module is also used to transmit data to a stand alone PC.

This stand alone PC is a Pentium PC that potentially allows more complex computing than an insect brain. This PC is used for most of the cognitive processing, because the on-board computational resources are not sufficient. The physical behaviours required for the robots' sensing is processed on-board and the symbol formation and communication is processed on the PC. One advantage is that this further speeds up the experiments. Off-line processing of sensorimotor data is not uncommon in the field of evolutionary robotics and yield results that are more consistent with reality than simulations (Lund et al. 1997).

As a consequence of using such limited and noise-sensitive sensors and motors, the robots have poor sensorimotor skills. For instance, when a robot senses the same light source on different occasions, the sensation might not be exactly the same. Or when a robot rotates around its axis one full circle, the time they take on different occasions will most likely vary and the circle will hardly ever be exactly 360 degrees. The speed of a robot is often not constant, although the motors are given a constant signal. These are just some examples of the poor sensorimotor skill; many more exist.

Given that the robots operate in a minimal environment and have minimal bodies, it would be wrong to expect that the robots could be able to acquire semiotic symbols that are similar to the symbols used in human languages. The categories such minimal robots can construct are limited in the way the robots can perceive and interact with their environment. For instance, the robots have no means to sense the colour or shapes of objects. Hence they will not be able to form concepts relating to colour of shape.

In communication the interaction between agents requires a sufficient amount of coordinated behaviour, for instance, in order to establish joint attention. The use of minimal robots with unreliable sensorimotor skills cause these robots to have extreme difficulties in coordinating their behaviour. As a consequence, the robots have a limited capability to develop a shared lexicon.

4 The model

The experiments are based on the language game model that was introduced by Steels (1996a). In a language game, two agents - a speaker and a hearer - try to convey the meaning of some referent by exchanging a word-form. When the agents fail, they can adapt their individual lexicons in order to improve their performance on future occasions. By playing a series of language games, a lexicon emerges that is shared on a global level. At the start of each experiment, the agents have no categories in their ontologies and their lexicons are empty. Both private lexicons and ontologies are constructed during the experiments.

Three different language games have been implemented on the minimal autonomous robots: the guessing game, the observational game and the selfish game. The games mainly differ in how both agents agree on the reference of the communication. In the guessing game, the hearer tries to guess what referent the speaker tries to convey. Corrective feedback is used to verify whether they succeed in doing so. The guessing game is also implemented in the Talking Heads experiment (Steels et al. 2002). In that experiment, the robots are embodied as pan-tilt cameras places on tripods and the environment contains geometrical figures pasted on a white board that have to be named.

In the observational game, the speaker informs the hearer what referent is the topic of the game prior to the verbal communication, thus establishing joint attention. Corrective feedback is not used to verify whether both agents communicate about the same referent as they already established joint attention.

In the selfish game, neither corrective feedback, nor joint attention is established. Both agents just infer the meaning of the verbal communication by counting the co-occurrences of word-forms and meanings.

In this paper the observational game will be presented in more detail. The other games will be discussed briefly in section 6. In short, the observational game is organised as follows:

1. The robots get together and stand close to each other with their backs facing each other. One robot arbitrarily takes up the role of speaker, the other becomes the hearer.

2. Each robot rotates around its axis to sense its environment one by one. As the robots are not located at the same position, the resulting spatial views differ for the two robots.

3. The sensing is segmented such that the sensory data relating to the sensing of a light source are extracted. This segmentation results in a set of segments.

4. From the segments feature vectors are extracted. Each segment can then be described by a feature vector that represents the sensory data by certain features. The set of resulting feature vectors constitutes the context of the game. Due to the differences in spatial views, the context of the speaker typically varies from the context of the hearer.

5. The speaker of the game selects an arbitrary feature vector from the context as the topic of the game. It informs the hearer which referent relates to the topic. This way joint attention is established.

6. Both speaker and hearer try to find a category that distinguishes the topic from all other feature vectors in their context. This is done by means of the discrimination game. When no distinctive categories can be found, new categories are constructed.

7. The speaker, when it finds a distinctive category tries to name this category by searching its lexicon for form-meaning associations that are consistent with the distinctive category.

8. The hearer, when received the utterance, tries to interpret the uttered name in relation to the topic. It does so by searching its own lexicon for associations that are both consistent with the uttered name and the distinctive category.

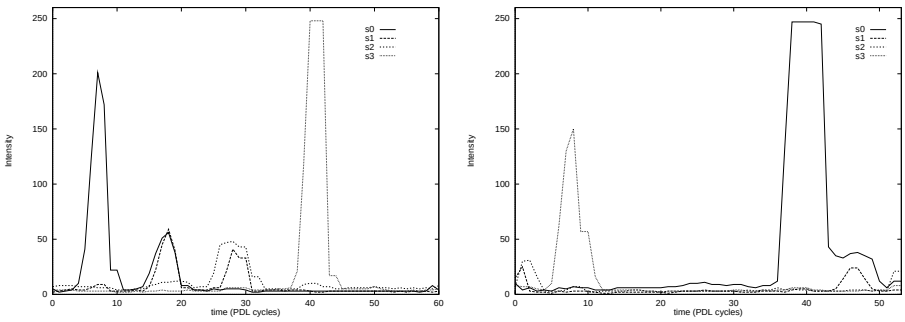
9. If the hearer finds a consistent association, the observational game succeeds. Otherwise it fails. Both robots know the outcome of the game.

10. Depending on the outcome of the game, the robots adapt their lexicons. New forms may be constructed, existing ones may be adopted and the associations between forms and meanings may be strengthened or weakened.

The above outline is a very brief description of the observational game of which some parts will be discussed in more details below. The reader interested in the technical details is referred to, e.g., (Vogt 2000b; Vogt in press).

4.1 Sensing, segmentation and feature extraction

The robots need to get together at close distance in order to detect their environment as similar as possible. In previous experiments, such as reported in (Steels and



(a) robot 0 (b) robot 1

Figure 3: The spatial view of two robots acquired during the sensing event for one observational game. The intensities of the four light sensors are plotted on the y-axis, the time is plotted on the x-axis (the time is given in PDL units, each unit is $\frac{1}{40}s$).

Vogt 1997), the robots did this completely autonomous. However, this behaviour took on the average 2 minutes for each language game. As an experiment may need thousands of language games, this is very time consuming. To speed up the experiments, the experimenter intervenes to bring the robots close to each other by hand. Such interventions do not influence the fundamental properties of the experiment, because the essential part of a language game follows after the robots stand close to each other.

After the robots come together and stand opposite of each other, they start sensing their environment to obtain a spatial view that forms the basis of the context of the observational game. The robots sense their environment by rotating around their axis twice, while they record the intensity of the four light sensors during the middle 360 degrees[3]. Figure 3 shows the result of a sensing event for both robots during one observational game. Figure 3 (a) clearly shows four peaks wherever the first robot detected a light source. In each peak a different sensor reads the highest intensity. The sensor that reads the highest intensity corresponds in height to the light source that caused the sensor's stimulation. The robots are not aware of this information. As figure 3 (b) shows, the robots do not acquire the same spatial view. This is due to the fact that the robots are positioned at different locations. In this case, the second robot did not clearly detect all four light sources because,

[3]Previously the robots only rotated once (Steels and Vogt 1997), but it has been found that the on- and off-set of rotation was a large source of noise.

for instance, the first robot may have obscured the visual field. As a result, the robots will not be capable of constructing a highly similar context.

The sensing results in a spatial view of the robots' surroundings, which is described by a set of raw sensory data that may vary in length for the two robots and for different games. In order to identify the different light sources from this data set, the robots extract connected regions that correspond to a peak. Each connected region is called a segment and is represented, again, by the raw sensory data that has been measured within this region. As the regions may differ in size, the segments differ in size. It is assumed that each segment corresponds to the sensing of a light source.

In order to have a computationally manageable categorisation, each detected light source should have a consistent representation; the segments do not provide this. To come with a such a representation, a feature vector is constructed from the segment. It is possible to design feature vectors such that they already bear some useful information concerning the sensory data in the segments. Humans and other robots such as the Talking Heads, for instance, extract features that contain information concerning the colour of a segment (in terms of, e.g., hue, saturation and brightness), the size, the spatial location etc. The minimal autonomous robots, however, have sensors that can only detect the intensity of light. Each robot therefore extracts features by looking within a segment for the highest measured intensities for each sensor. These values are then divided (or normalised) by the maximum measured intensity within the segment of all sensors. This way, the sensor that has measured the highest intensity within a segment will yield a feature with a value of one; while all other sensors will yield features lower than one. Thus a feature vector is constructed for each segment where each feature in the vector is a value that relates to the maximum intensity of a sensor in that segment.

The way features are extracted is extremely important, because good features can already contain a lot of information about referents. Having feature vectors that already bear some invariant information allows more consistent categorisation and naming. In (Steels and Vogt 1997), for instance, the robots extracted features corresponding directly to the absolute intensities of the light sensors, hence they also contained distance information. Since distances between a robot and a light source vary a lot on different occasions, the feature vectors vary a lot too. This, in turn, leads to inconsistent categorisation and much confusion amongst the robots.

4.2 Topic selection

When the robots acquired a context for the observational game, the speaker chooses an arbitrary feature vector as the *topic* of the game. The speaker then informs the hearer what referent corresponds to the topic, so that both robots know what the topic is. This way the robots establish joint attention.

The idea is that the speaker points at the referent and the hearer tries to find out what the speaker is pointing at. This all seems reasonable, but here a major limitation of the minimal robots becomes apparent. Implementing pointing behaviour physically on the robots was extremely difficult and several attempts failed. In (Steels and Vogt 1997), for instance, it was implemented by having the speaker orient in the direction of topic. While the speaker rotated in that direction, four perpendicular beams of infrared were emitted, such that the hearer could approximate the angle at which the speaker was looking. This worked well under the condition that both robots were initially facing each other perfectly at close distance. However, slight disturbances in the initial orientation of the robots already leads to failures. As the sensorimotor skills of the robots are unreliable, they were most of the time not facing each other perfectly. In the end, all attempts to implement some kind of pointing behaviour physically failed.

Joint attention is therefore *simulated* by having the robots inspect each other's internal states. The speaker presents the hearer the feature vector of the topic. In turn, the hearer selects from its context the feature vector that best matches the given vector as the topic and joint attention is established. Although this works reasonably well, sometimes the hearer did not detect the speaker's topic. In such cases, the hearer may have selected a different topic than the speaker. Naturally, this implementation is far from plausible - agents are not supposed to have access to each other's internal states. It has been implemented this way in order to study the bootstrapping of semiotic symbols further under the assumption that joint attention could be implemented physically.

4.3 Meaning formation

The meaning formation is based on the *discrimination game* model (Steels 1996a). Below follows a brief description of the discrimination game, which is slightly altered from the exact implementation for reasons of clarity. For exact details, consult, e.g., (Vogt 2000b; Vogt in press).

In order to find a *distinctive category*, all the feature vectors are categorised.

Categories are represented as prototypes (or points) in a feature space, which is the space that is spanned by all possible feature vectors that a robot can acquire. Categories are defined as regions in this feature space in which all points are nearest to this prototype. A feature vector is categorised with that category for which the feature vector lies within the region of the category. Or in other words: a feature vector is categorised with the category for which the prototype is nearest to the feature vector.

A category is distinctive when it is a category for the topic, but not for any other feature vector in the context. When a distinctive category is found, the discrimination game is a success and the prototype of this category is shifted towards the feature vector of the topic. By shifting the prototype the category becomes a more representative sample of the feature vectors it categorises.

The discrimination game fails when no distinctive category is found. In this case a new category is added to the ontology. The feature vector of the topic is taken as the prototype of this new category. Recall that initially, the ontology is empty, so this way the discrimination game controls the construction of categories.

To allow both general and specialised categories, each robot has a number of feature spaces at its disposal in which the density of categories differ. The categories in the more dense feature spaces are more specific, while the categories in less dense spaces are more general. The robots play a discrimination game for each feature space of which they have six available in the experiment. Hence, the meaning formation could result in a set of total of six distinctive categories. This property allows the robots, in principle, to categorise their sensing by their superordinate and subordinate categories. For instance, if the robots were to categorise a dog, they would be able to categorise them both by its superordinate category *dogs* and by its subordinate category, e.g., *bulldog*.

The way categories are represented is not extremely important for the discrimination game to work (De Jong and Vogt 1998; Vogt 2000b). The Talking Heads, for instance, use binary trees (Steels et al. 2002) and De Jong uses binary subspaces (De Jong and Vogt 1998; De Jong 2000). For the minimal autonomous robots a prototype representation is used as they form a more plausible representation than binary distinctions (Lakoff 1987; Rosch et al. 1976). Prototypical categories can be viewed as convex regions inside the feature space and thus resemble *conceptual spaces* (Gärdenfors 2000).

It is important to notice that the discrimination games are played by each robot individually, just as the sensing, segmentation and feature extraction. This way,

each robot will develop its own ontology of categories that, although perhaps similar, differs from the ontology of another robot. The discrimination game will thus result in different distinctive categories for each individual robot.

4.4 Producing an utterance

Once the speaker acquired one or more distinctive categories, it will try to find a matching form-meaning association that it has stored in its lexicon. The lexicon of the robots is a set of associations between a form and a meaning of which the strength is given by a score. The meaning corresponds to a certain category the robot has stored in its ontology and that has been associated with a form.

For the production, the speaker orders the distinctive categories according to some preference criteria. These criteria tend to prefer general categories that have been used successfully in previous games. One by one, the distinctive categories are selected and the speaker searches in its lexicon for associations of which the meaning corresponds to that distinctive category. This may yield more than one association, because meanings can be associated with several forms and vice versa. If the speaker finds one or more matching associations, it selects that association for which the score is highest. If no association is found, the speaker explores the next distinctive category until it finds one or more associations or until all distinctive categories have been explored.

If, in the end, no matching association has been found, the speaker may invent a new form and adds the new form-meaning association to the lexicon. This is done with a certain probability that has been set to 0.1 in the current experiment. If this probability is too high, too much synonymy and polysemy will emerge in the lexicon. If it is too low, the lexicon does not get off the ground fast enough (Vogt 2000b).

When an association is found or a new one is invented, the speaker utters its form to the hearer. Although in animal or human communication this utterances are transmitted physically and the robots may use radio communication, in the current experiment this is processed on a PC where no physical transmission is necessary.

4.5 Interpreting an utterance

When the hearer 'receives' the utterance, it will try to interpret this utterance in relation to the distinctive categories of the topic. For this, the hearer searches its

lexicon for associations of which the form matches the utterance *and* of which the meaning matches one of the distinctive categories.

If there are one or more associations, the hearer will select that association for which the association score is highest, regulated by the mentioned preference criteria. In that case the observational game is considered successful. If there are no associations found, the observational game is a failure and the lexicon has to be expanded. The outcome of the game is signalled back to the speaker, so both robots know the outcome.

4.6 Adapting the lexicon

Depending on the outcome of the game, the robots have various ways of adapting the lexicon. They can expand the lexicon and they can adapt the association scores. The invention of new forms has already been discussed, below follow the remaining two adaptations:

Failure In case the hearer does not know the uttered form with respect to one of the distinctive categories of the topic, the observational game is a failure and the lexicon has to be expanded. In this case, the hearer adopts the form and associates it with the distinctive categories of the topic. These associations are added to the lexicon. In turn, the speaker lowers the association score of the used association.

Success In case the observational game is a success, both robots increase the association score of the used association and they laterally inhibit competing associations. An association is competing when either its form is the same as the uttered form, but not its meaning or vice versa. This latter adaptation implements a preference for the agents that meanings have a one-to-one relation with a form.

These adaptations, together with the cultural interactions, ensure a self-organisation of the global lexicon. For a more detailed discussion on the self-organisation of language games, see, e.g., (De Boer 2000; De Jong 2000).

5 Experimental results

Several experiments have been done with (variants of) the above model, for a detailed report on the results of these experiments see (Vogt 2000b). This section presents the results of one such experiment.

For the experiments a data set of approximately 1,000 situations was recorded beforehand. Each situation contains the sensory data of both robots' sensing as if they played one observational game. These situations are processed further off-board on a PC. The recorded situations are fed to the segmentation, feature extraction, discrimination- and naming games, and they are re-used for playing 10,000 observational games. The experiment has been repeated for 10 runs with different random seeds. (Thus assuring different initial conditions for each run.) As for each observational game one robot is arbitrarily assigned the role of speaker, who randomly selects a topic, the chance that the exact conditions of a game re-occurs is small. It has been calculated that there are approximately 7,000 possible conditions.

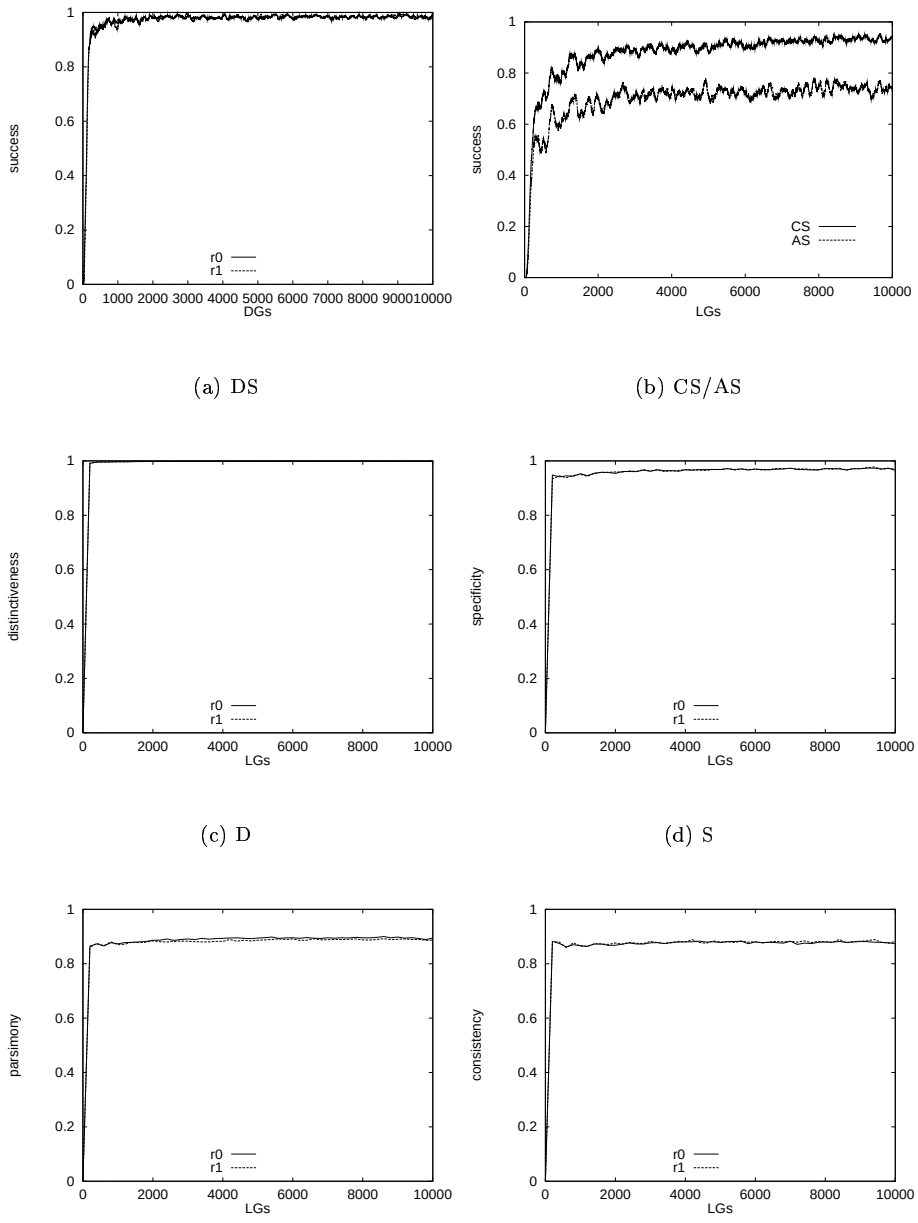
From the acquired data set some statistics have been extracted. The a priori success is calculated from the average context size and is scaled for the potential understandability. Statistics revealed that the average context size was 3.35 segments. The *potential understandability* indicates an upper limit in the number of successful games to be expected. As shown in figure 3, the robots do not always sense the same spatial view and therefore cannot select the same topic. It has been calculated from the recorded data set that in 80.5% of the games the robots do select the same topic. Hence the a priori chance that the robots would randomly choose the same segment as the topic is calculated to be 23.4%. This reveals another major limitation of the minimal robots. Due to their minimal sensory and coordination abilities, the robots fail to construct a coherent context properly. Given these statistics, the remainder of this section presents the results of the experiment. First by looking at quantitative results and then by looking at the quality of an emerged lexicon.

5.1 Quantitative results

The success of the experiment averaged over the 10 trials is presented with a number of measures. The results are shown in figure 4. The plots in the left column, figures (a), (c) and (e), show the results relating to the discrimination games. The plots in the right column, figures (b), (d) and (f), show the results in relation to the communication.

5.1.1 Discrimination games

The plot in figure 4 (a) shows the *discriminative success* (Steels 1996b), which is the average number of successful discrimination games over the past 100 observational games. Recall that a discrimination game is successful when the individual robot



(a) DS

(b) CS/AS

(c) D

(d) S

(e) P

(f) C

Figure 4: The measured success of the observational games. The figure shows (a) the discriminative success, (b) the communicative success (upper line) and the actual success (lower line), (c) the distinctiveness, (d) the specificity, (e) the parsimony and (f) the consistency.

was able to find distinctive categories in relation to the feature vector of the topic. The figure shows that the discriminative success increases towards a value near 98% within the first 1,000 observational games. This means the robots are very good in constructing an ontology with which they can form distinctive categories of their sensing.

The remaining measures relating to the discrimination game inspect the qualitative evolution of the ontologies. These measures are developed by De Jong (2000) and are based on the information entropy introduced by Shannon (1948). Below follows a brief description of these measures; the equations to calculate these measures can be found in (De Jong 2000; Vogt 2000b; Vogt in press).

Figure 4 (c) shows the *distinctiveness*. The distinctiveness is a measure for inspecting the quality of the robots' categorisation and indicates more or less how uniquely the use of a meaning predicts its categorisation of a referent. So, if the distinctiveness is high, the meanings a robot uses relate mostly to the same referent. When it is low, the meanings are used for various referents. As figure 4 (d) shows, from the moment the robots start to use distinctive categories in the observational games they almost uniquely relate to a particular referent.

The *parsimony* indicates how consistent a particular referent is categorised with a meaning. If the parsimony is high, a referent is consistently categorised by the same meaning. Figure 5 (e) shows that the robots do not categorise each referent consistently with the same meaning. In fact, the robots use many meanings in the different observational games. It has been observed that in one experiment the robots each used about 500 meanings to categorise no more than 4 referents! This can be explained by the fact that the robots have many different views of the referents, depending on, for instance, the robots' locations, changing lighting conditions and unreliability of the sensors. Although the parsimony is rather low, the robots construct a small lexicon rather well as will be shown below.

5.1.2 Communication

The upper line of figure 4 (b) shows the *communicative success* (Steels 1996a). This is calculated by averaging the number of successful observational games over the past 100 games under the assumption that joint attention is established perfectly. As can be seen, the success rate already exceeds 60% after a few hundred games. After that it slowly increases to almost 95% after 10,000 games. In the last 5,000 games the communicative success is more or less stabilised. Recall that when the speaker presents the feature vector of the topic to the hearer, the latter selects the

feature vector that is most similar to the presented vector. But the hearer does not always sense the topic, it only does in 80.5%. Thus the communicative success is measured from the viewpoint of the robots, which is not objective. The lower line of figure 4 (b) shows the *actual success*. This is measured by averaging over the past 100 games the number of successful observational games where both robots refer to the same light source. The actual success has a similar evolution as the communicative success, but is about 20% lower. So, the actual success increases to a value near the potential understandability and the robots learn to communicate rather well.

The two remaining measures are, again based on the entropy measures introduced by (De Jong 2000). These measures inspect the quality of the emerging communication system. One of them - the *specificity* - indicates how uniquely the use of a form refers to a particular referent. It is shown in figure 4 (d) that when the robots use a particular name, they use it almost unambiguously to name one referent. The specificity increases towards a value of 0.95 around which it stabilises, meaning that the global lexicon reveals little polysemy.

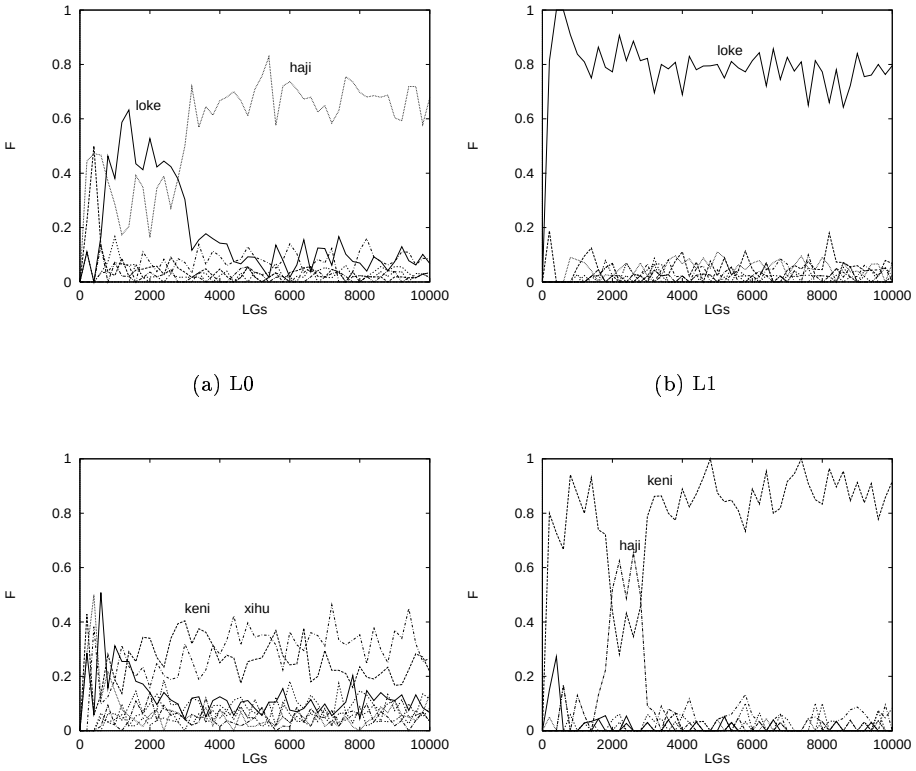
The second measure concerning the quality of the lexicon is called the *consistency* and is used to monitor how consistently the robots name a particular referent. The consistency stabilises around 0.85 and is lower than the specificity. In this sense it is similar to the parsimony, which is lower than the distinctiveness. This indicates that the robots do not always use the same name in relation to a particular referent. But, although the robots may use up to 500 different meanings, the number of names that are used in a typical experiment is around 15 of which the robots tend to use about 7 frequently and the other 8 seldomly.

These measures all indicate that the robots are capable of developing a grounded and shared lexicon. The robots may use different meanings or forms to categorise or name some referent, but when a meaning or form is used, they almost uniquely refer to the same referent. This becomes clearer by looking at the formation of the lexicon of one typical experiment in more detail.

5.2 Tracking the emergence of semiotic symbols

One way to inspect the emergence of semiotic symbols is to look at various *competition diagrams*. A competition diagram can be used to monitor the evolution of meanings and forms of an individual run.

Figure 5 shows the competition of forms that are used to name the various referents (light sources). On the x-axis of the figure the number of observational



(a) L0 (b) L1

(c) L2 (d) L3

Figure 5: The referent-form competition diagrams for light sources (a) L0, (b) L1, (c) L2 and (d) L3 show the co-occurrence frequencies of the competing forms relative to the occurrence frequency of the referents (y-axis). The frequencies are calculated over every 200 observational games (x-axis). Note that elements that compete with very low frequencies are left out for clarity. All graphs, except (b), show the results of both robots r0 and r1.

games are shown. The y-axis shows the occurrence frequencies of forms that are successfully used to name a referent relative to the occurrence of that referent. The frequencies are measured over every 200 observational games.

The competition diagrams show that there are several forms competing to name one referent. For referent L0 there emerges a winning form 'haji' after approximately 2,500 games, see figure 5 (a). At first 'haji' competes strongly with 'loke', which already from the start wins to name L1 as shown in figure 5 (b). As a result, the competing associations of 'loke' with meanings that refer to L0 are laterally inhibited more often than the associations with L1. This strengthens the competition of 'haji' to name L0 and both 'haji' and 'loke' become clear winners for light sources L0 and

L1 respectively. Such competitions are very typical.

The competition for light source L2 is much harder, see figure 5 (c). No clear winning forms can be found; there are two forms 'xihu' and 'keni' that are frequently used throughout the experiment. Note, however, that 'keni' is also a winning form to name referent L3 as figure 5 (d) shows. Besides being a synonym for referent L2, 'keni' has a polysemous relation with referents L2 and L3. Being both part of a synonymous and polysemous relation is not necessarily related. Some symbols have been observed upon inspection of the lexicon that are used either synonymous or polysemous. (Note, by the way, that in all four graphs there is a lot of competition going on at the bottom.)

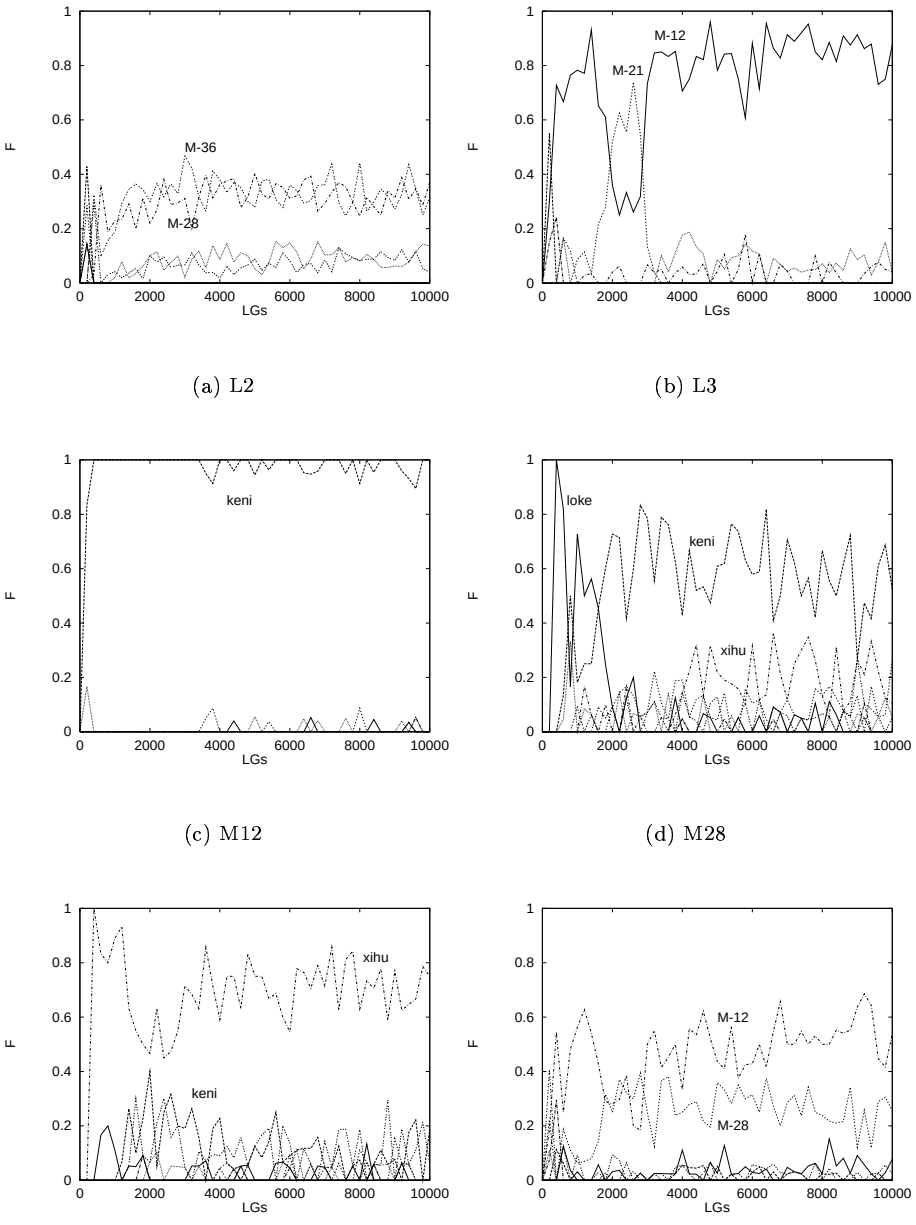
Figures 6 (a) and (b) show the referent-meaning competition diagrams for referents L2 and L3 respectively. Like in the referent-form diagrams of figure 5, the competition for light source L2 is strongest. The competition for L3 is less strong. The meaning-form competition for meaning M12 in figure 6 (c) shows that 'keni' is almost uniquely used to name M12. This seems to explain the winning competition of 'keni' for L3. However, 'keni' is also in strong competition for meanings M28 and, to a lesser extent, M36, see figures 6 (d), (e) and (f).

It is possible to inspect the use of semiotic symbols differently by looking at a semiotic landscape as shown in figure 7. In this figure some forms and meanings occur that are not labeled in the competition diagrams. The lexicon clearly shows that the number of meanings outnumbers the number of forms in relation to the referents. So there is a tendency that the relationship between referent and form becomes close to one-to-one (or more specific one-to-few), despite a strong one-to-many relationship between referent and meaning, and between form and meaning.

6 Discussion

The experimental results show that minimal autonomous robots are capable of developing a shared set of semiotic symbols under the assumption that they are able to establish joint attention. Although much better than chance, the (actual) communicative success[4] is lower than 80%. The robots tend to name referents consistently over time. But there is still quite some polysemy and synonymy, which destabilises the lexicon at times. An important question is what factors are responsible for the limited abilities of the minimal autonomous robots to develop a shared set of

[4]In the remainder of this paper, the term 'communicative success' is used rather than actual success, because this term speaks better for itself.



(a) L2

(b) L3

(c) M12

(d) M28

(e) M36

(f) 'keni'

Figure 6: Referent-meaning competition diagrams for referents (a) L2 and (b) L3, the meaning-form competition diagrams for meanings (c) M12, (d) M28 and (e) M36 and (f) the form-meaning competition for form 'keni'.

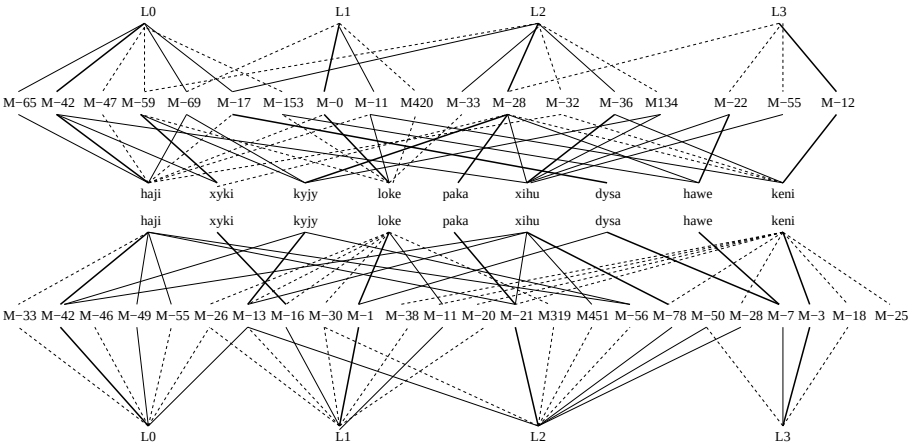


Figure 7: A semiotic landscape of the same run as the competition diagrams. The landscape shows the relations between referents, meanings and forms of the two robots. The line styles indicate the co-occurrence frequencies over the 10,000 games of the form-meanings relative to the occurrence frequencies of forms and of the referent-meanings relative to the referents. Bold lines indicate the winning associations. Thin continuous lines show associations with a relative frequency higher than 5%. Dashed lines show associations with a relative frequency between 0.5 and 5%. Associations that are less frequently used are left out for clarity of the graph.

semiotic symbols? And also to what extent is the experiment scalable with respect to the size of the lexicon?

A number of physical factors were found to limit the cognitive abilities of the robots. The factors that appear to be most crucial are:

1. The robots' sensors are subject to a lot of noise.
2. The robots have limited sensory abilities.
3. The environment is very limited.
4. The robots have poor sensorimotor skills.
5. The robots fail to construct a coherent context.
6. The robots fail to establish joint attention physically.

The first factor has the consequence that the sensation of a light source can differ a lot on various occasions, even when viewed from more or less the same location. This leads to many different categorisations of the light sources and, in turn increases the complexity of communication. It is therefore one of the sources

of synonymy and puts pressure on the communicative success. Nevertheless, the self-organising mechanism of the observational game reduces the number of forms that are used significantly (figure 7). The experiment with the AIBO too revealed many different categorisations (or stored views) of referents (Steels and Kaplan this issue). Steels and Kaplan mention the changing lighting conditions as the source of inconsistent views rather than noise in the sensor. To overcome the problem they use a deliberate classification method for learning the words of the many views, instead of the implicit mechanism used here.

Factors 2 and 3 are probably most important with respect to the size of the lexicon. Four light sources are, of course, very limited for a realistic ecosystem. The same holds for robots with only four light sensors to detect their environment[5]. However, the objective of this experiment was not to investigate the emergence of semiotic symbols in a realistic setting, but rather to investigate the limitation of this emergence using minimal autonomous robots.

One important question is to what extent the lexicon can be expanded? Adding more objects to the environment without changing the robots would not work. Perhaps a small increase in lexicon size can be established, but adding new light sources to the environment reduces their distinctiveness. Experiments in which the environmental distinctiveness was reduced, revealed that this has a negative effect on the performance of the robots (Vogt 2000b). Adding more feature extraction methods and, possibly, more or richer sensors on the robots will have a larger impact. With more features robots are capable of constructing a richer repertoire of distinctions such as spatial relations, distance, colour, etc. This would also allow the environment to be expanded. An increasing repertoire of features, however, may require a different categorisation method than the discrimination game. Steels and Kaplan (this issue) report that their categorisation method does not work with a large number of features.

Factor 4 - having poor sensorimotor skills - contributes for a great deal to the causes of factors 5 and 6. Poor sensorimotor skills reduces the capability of the robots to coordinate their actions, which is a requirement to construct a shared context and to establish joint attention. Improved sensorimotor skills would allow the robots to set up more ideal conditions to view their environment similarly. Also the precondition of facing each other perfectly before pointing in order to establish

[5] Although, the robots also have infrared sensors and bumpers, these are not directly used to form semiotic symbols and thus do not influence the size of the lexicon. These are therefore left out in the current discussion.

joint attention (see section 4.2) would be reached more often.

The robots' inability of reaching a shared context or joint attention - factors 5 and 6 - has a more fundamental cause. When humans notice that they fail to construct a shared context or establish joint attention, they immediately search for the missing information. Such adaptations require some form of *theory of mind* (Bloom 2000) or the ability to reason about the intentions of a speaker (Tomasello 1999). Failures of this kind are always revealed by the hearer when it is unable to identify the speaker's topic. When such failures occur, the hearer has to rescan its environment to search for the missing information. For this it would most probably move from its place as the speaker might block its view. Besides having the proper reasoning mechanisms, the robots will need to have better sensorimotor skills to solve these problems.

That factors 4, 5 and 6 are typical for minimal autonomous robots can be inferred from related work. The AIBO, which is in a sense a minimal autonomous robot too, also has unreliable sensorimotor skills. It has no ability to establish joint attention to most of the objects, except for the red ball (Steels and Kaplan this issue). The human has to push the object in front of the AIBO or has to watch carefully when the AIBO is looking at the object. Except for the red ball, the AIBO has no mechanism to establish attention on objects. Build in mechanisms cause the AIBO to track red objects, but even then it can loose track of the ball. Billard and colleagues have also faced the same problems. In their experiment, a student robot learns the lexicon from a teacher robot while following this teacher robot and communicating with it (Billard and Hayes 1997; Billard and Dautenhahn 1999), see (Vogt 2000a) for a similar experiment. In their experiments the level of success is similar to the ones presented here, and they argue that the main cause of failures came from the robots inability to construct a shared context (Billard and Dautenhahn 1999). When the interactions were done with a human experimenter as teacher, the success rate increased because the human adapts its context by carefully monitoring the robot's behaviour. When more reliable and controlled robots, such as the Talking Heads are used, the problems of factors 4, 5 and 6 are reduced and the performance improves (Steels et al. 2002).

The question remains whether establishing joint attention during the observational game is a necessary strategy? In Western cultures children are taught a lot of novel novel words by, for instance, pointing at an object and saying its name. However, in non-Western cultures, there is evidence that infants learn their first words while no joint attention is established (Lieven 1994).

An alternative would be that children receive corrective feedback on their use of word-meanings. This is what happens in the guessing game that has also been implemented on the minimal autonomous robots (Vogt 2000b; Vogt 2001; Vogt in press). These experiments reveal that the guessing game is at least as good as the observational game. In the observational game a high level of communicative success is established faster, but the lexicons emerging from the guessing game reveal less synonymy and polysemy. The the observational is faster is consistent with the observation reported in (Tomasselo and Todd 1983) that children tend to learn faster when joint attention is established than when not. The way corrective feedback has been implemented, however, is similar to the way joint attention has been implemented, namely by inspecting each other's internal states. For the guessing game factor 6 would be changed in: "The robots fail to evaluate corrective feedback physically." For a comparison between the observational game and the guessing game, see (Vogt 2000b; Vogt 2001).

Like there is negative evidence for joint attention as a necessity for learning the meaning of words, there appears to be negative evidence for corrective feedback as well (Bloom 2000). Another alternative would be to see if a lexicon can emerge without using either joint attention or corrective feedback. Such a game has been implemented as the *selfish game* on the minimal autonomous robots (Vogt 2000b). In this selfish game, the hearer guesses the meaning of the speaker's utterance as in the guessing game. But instead of using corrective feedback to adapt the lexicon, the robots keep track of the co-occurrence between meaning and form. These co-occurrence frequencies then determine the selection of associations.

However, experiments with the selfish game on the minimal robots yielded very poor results (Vogt 2000b). The performance of the robots was not better than chance, because the selfish game can only work when there is enough variation in the context. If the context does not vary enough, the co-occurrence frequency of word-forms with meanings is distributed more or less equally and the robots have no consistent criterion for selecting the proper associations. Yet unpublished simulation experiments reveal that when the environment of the agents is richer and there is sufficient variation in the context, the selfish game does work. See also (Smith 2001) for simulations with a similar model. Still the communicative success is both lower and stabilises slower than in the observational and guessing games. Again, this result is consistent with the observation of Tomasselo and Todd (1983).

The experiments might, at least to some extent, shed light on why other species that humans have no learned symbolic communication system. The first three fac-

tors might explain why lower animals such as ants have limited communication systems. The more an animal can distinguish, the more information it can acquire from its interaction with the environment and the more information can be communicated. It certainly explains why species with different sensory capabilities have different communication systems. This does not explain, for instance, why blind sighted people can learn language equally well as sighted humans. But blind people do have other sensory capabilities and they probably have an ability to form abstract concepts concerning, e.g., colours. The limited sensory abilities also does not explain why nonhuman primates such as bonobos have a limited ability to develop a symbolic communication system.

The experiment shows that poor sensorimotor skills are sources of erroneous communication (factor 4). Especially since they result in poor physical coordination during their interactions to reach a shared context (factor 5) and joint attention (factor 6). Although lower species, such as ants do reveal coordinated behaviours, these are often limited and less complex and less fluent than those exposed by humans. Lower animals lack a rich repertoire of coordinated behaviour skills, which humans do have. Like, for instance, follow an eye-gaze, grasp things, point with their fingers etc.

The sensorimotor skills of nonhuman primates are better than those of lower species. This is perhaps a reason why bonobos can learn a limited number of symbols through intensive interactions and reinforcements given by humans. The reason why they fail to develop symbolic communication systems at a large scale has possibly to do with the lack of a capability to understand that other conspecifics also have intentions (Tomasello 1999). Having this capability would allow species to share attention on something, which is necessary to bootstrap grounded symbols.

7 Conclusion

This paper investigated crucial physical conditions that minimal autonomous robots need to bootstrap grounded symbols. The robots are minimal with respect to their physical architecture and their environment. The robots' bodies are far more simple then living animals, especially those animals that are known to communicate in symbols. The cognitive architecture, on the other hand, allows the robots to develop symbols.

The experiments revealed that the robots can develop a shared lexicon when they interact with a physical environment, have the proper learning mechanisms

and have sufficiently adapted bodies. Acquiring knowledge by interacting with the environment and learning through self-organisation are core principles of situated cognition and requires embodiment (Clancey 1997; Pfeifer and Scheier 1999).

The extend in which lexicons can be developed depends heavily on the physical abilities of agents. Both minimal sensing abilities and a minimal environment form a major limitation on the size of the lexicon. It is extremely important that robots are equipped with reliable sensors from which they can extract features that contain invariant information. Also having poor sensorimotor skills appear to be an important drawback in the ability to develop large and shared lexicons - especially as the robots have difficulties to construct a shared context or fail to establish joint attention, which are both necessary to bootstrap the lexicon.

The poor ability of the robots to agree on the reference of speakers' utterances is problematic. To improve this the robots would probably require something as a theory of mind to form beliefs about speakers' intentions. Investigations how such a theory of mind can be designed or - better - evolved and/or learned should be the next challenge in the research on bootstrapping grounded symbols in *autonomous robots*; preferably not in minimal autonomous robots as these are too limited in order to scale up a lexicon.

Acknowledgements

The research of this paper has been carried out at the Artificial Intelligence Laboratory of the Vrije Universiteit Brussel. The author wishes to thank Luc Steels for providing an excellent research environment and for his valuable comments that helped improve this paper. Many thanks to Miranda Brouwer, Tony Belpaeme and Michel van Dartel for giving critical comments on earlier versions of this paper.

References

Arkin, R. C. (1998). *Behavior-based robotics*. Cambridge Ma.: The MIT Press.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences 22*, 577–609.

Billard, A. and K. Dautenhahn (1999). Experiments in social robotics: grounding and use of communication in autonomous agents. *Adaptive Behavior 7(3-4)*.

Billard, A. and G. Hayes (1997). Robot's first steps, robot's first words ... In Sorace and Heycock (Eds.), *Proceedings of the GALA '97 Conference on*

Language Acquisition - Edingburgh, Human Communication Research Centre. University of Edinburgh.

Bloom, P. (2000). *How Children Learn the Meanings of Words*. Cambridge, MA. and London, UK.: The MIT Press.

Briscoe, E. J. (Ed.) (2001). *Linguistic evolution through language acquisition: formal and computational models*. Cambridge: Cambridge University Press.

Brooks, R. A. (1990). Elephants don't play chess. *Robotics and Autonomous Systems 6*, 3–15.

Cangelosi, A., A. Greco, and S. Harnad (2000). From robotic toil to symbolic theft: Grounding transfer from entry-level to higher-level categories. *Connection Science in press*.

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

Chandler, D. (2001). *Semiotics: The Basics*. Routledge.

Clancey, W. J. (1997). *Situated Cognition*. Cambridge University Press.

De Boer, B. (2000). Self-organization in vowel systems. *Journal of Phonetics 28*, 441–465.

De Jong, E. D. (1999). Autonomous concept formation. In *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence IJCAI'99*.

De Jong, E. D. (2000). *The Development of Communication*. Ph. D. thesis, Vrije Universiteit Brussel.

De Jong, E. D. and P. Vogt (1998). How should a robot discriminate between objects? In R. Pfeifer, B. Blumberg, J.-A. Meyer, and S. Wilson (Eds.), *From animals to animats 5, Proceedings of the fifth internation conference on simulation of adaptive behavior*, Cambridge, Ma. MIT Press.

De Saussure, F. (1974). *Course in general linguistics*. New York: Fontana.

Deacon, T. (1997). *The Symbolic Species*. New York, NY.: W. Norton and Co.

Ellman, J. (1993). Learning and development in neural networks: The importance of starting small. *Cognition 48*, 71–99.

Gärdenfors, P. (2000). *Conceptual Spaces*. Bradford Books, MIT Press.

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

Harnad, S. (1993). Symbol grounding is an empirical problem: Neural nets are just a candidate component. In *Proceedings of the Fifteenth Annual Meeting of the Cognitive Science Society*, NJ. Erlbaum.

Hashimoto, T. and T. Ikegami (1996). Emergence of net-grammar in communicating agents. *BioSystems 38*, 1–14.

Hurford, J. R. (1989). Biological evolution of the saussurean sign as a component of the language acquisition device. *Lingua 77,2*, 187–222.

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

Knight, C., M. Studdert-Kennedy, and J. R. Hurford (Eds.) (2000). *The Evolutionary Emergence of Language: Social function and the origins of linguistic form*. Cambridge University Press.

Lakoff, G. (1987). *Women, Fire and Dangerous Things*. The University of Chicago Press.

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

Lund, H. H., J. Hallam, and W. Lee (1997). Evolving robot morphology. In *Proceedings of the IEEE Fourth International Conference on Evolutionary Computation*. IEEE Press.

Newell, A. and H. A. Simon (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the ACM 19*, 113–126.

Ogden, C. K. and I. A. Richards (1923). *The Meaning of Meaning: A Study of the Influence of Language upon Thought and of the Science of Symbolism*. London: Routledge & Kegan Paul Ltd.

Oliphant, M. (1999). The learning barrier: Moving from innate to learned systems of communication. *Adaptive Behavior 7 (3-4)*, 371–384.

Peirce, C. S. (1931). *Collected Papers*, Volume I-VIII. Cambridge Ma.: Harvard University Press. (The volumes were published from 1931 to 1958).

Pfeifer, R. and C. Scheier (1999). *Understanding Intelligence*. MIT Press.

Redford, M. A., C. C. Chen, and R. Miikkulainen (2001). Constrained emergence of universals and variation in syllable systems. *Lang Speech 44(1)*, 27–56.

Rosch, E., C. B. Mervis, W. D. Gray, D. M. Johnson, and P. Boyes-Braem (1976). Basic objects in natural categories. *Cognitive Psychology 8*, 382–439.

Seyfarth, R. and D. Cheney (1986). Vocal development in vervet monkeys. *Animal Behavior 34*, 1640–1658.

Shannon, C. (1948). A mathematical theory of communication. *The Bell System Technical Journal 27*, 379–423, 623–656.

Smith, A. D. M. (2001). Establishing communication systems without explicit meaning transmission. In J. Kelemen and P. Sosík (Eds.), *Proceeding of the 6th European Conference on Artificial Life, ECAL 2001*, LNAI 2159, Berlin Heidelberg, pp. 381–390. Springer-Verlag.

Steels, L. (1996a). Emergent adaptive lexicons. In P. Maes (Ed.), *From Animals to Animats 4: Proceedings of the Fourth International Conference On Simulating Adaptive Behavior*, Cambridge Ma. The MIT Press.

Steels, L. (1996b). Perceptually grounded meaning creation. In M. Tokoro (Ed.), *Proceedings of the International Conference on Multi-Agent Systems*, Menlo Park Ca. AAAI Press.

Steels, L. (1997). The synthetic modeling of language origins. *Evolution of Communication 1(1)*, 1–34.

Steels, L. and R. Brooks (Eds.) (1995). *The 'artificial life' route to 'artificial intelligence'. Building situated embodied agents*, New Haven. Lawrence Erlbaum Ass.

Steels, L. and F. Kaplan (this issue). Aibo's first words. the social learning of language and meaning. *Evolution of Communication this issue.*

Steels, L., F. Kaplan, A. McIntyre, and J. Van Looveren (2002). Crucial factors in the origins of word-meaning. In A. Wray (Ed.), *The Transition to Language*, Oxford, UK. Oxford University Press.

Steels, L. and P. Vogt (1997). Grounding adaptive language games in robotic agents. In C. Husbands and I. Harvey (Eds.), *Proceedings of the Fourth European Conference on Artificial Life*, Cambridge Ma. and London. MIT Press.

Tomasello, M. (1999). *The cultural origins of human cognition.* Harvard University Press.

Tomasselo, M. and J. Todd (1983). Joint attention and lexical acquisition style. *First Language 4*, 197–212.

Vogt, P. (2000a). Grounding language about actions: Mobile robots playing follow me games. In Meyer, Bertholz, Floreano, Roitblat, and Wilson (Eds.), *SAB2000 Proceedings Supplement Book*, Honolulu. International Society for Adaptive Behavior.

Vogt, P. (2000b). *Lexicon Grounding on Mobile Robots*. Ph. D. thesis, Vrije Universiteit Brussel.

Vogt, P. (2001). The impact of non-verbal communication on lexicon formation. In *Proceedings of the Belgian/Netherlands Artificial Intelligence Conference, BNAIC'01.*

Vogt, P. (in press). The physical symbol grounding problem. *Cognitive Systems Research Journal*. Special Issue on Embodied Cognition.