Symbol grounding in communicative mobile robots

Paul Vogt
IKAT / Infonomics - Universiteit Maastricht
P. O. Box 616 -5200 MD Maastricht
The Netherlands
p.vogt@cs.unimaas.nl

Abstract

This paper reports on an experiment in which two mobile robots solve the symbol grounding problem in a particular experimental setup. The robots do so by engaging in a series of so-called language games. In a language game the robots try to name an object that is in their environment first by categorizing their sensing, then by naming this categorization. When the robots fail to categorize or name, they can adapt their memory in order to succeed in the future. At the start of each experiment, the robots have no categories or names at all. These are constructed during the experiment. The paper concludes that in the investigated experiment the robots solve the symbol grounding problem as a result of the co-evolution of meaning and lexicon.

Introduction

One of the hardest problems in AI and robotics is the so-called *symbol grounding problem* (Harnad 1990). The symbol grounding problem deals with the question how seemingly meaningless symbols become meaningful in relation to the real world. For robots this problem can be translated to the problem how sub-symbolic representations of events and perceptions can be transformed into a symbolic representation such that these symbols can be used meaningfully.

Recently several attempts have been made to solve the symbol grounding problem in learning robots. Examples of such implementations are (Billard & Hayes 1998; Rosenstein & Cohen 1998; Steels & Vogt 1997).

In (Billard & Hayes 1998) a student robot tries to learn a lexicon from a teacher robot by learning through imitation. The teacher robot has a preprogrammed lexicon about some sensorimotor couplings. The student robot follows the teacher by means of phototaxis and while the teacher communicates its sensorimotor actions, the student tries to couple these communicated signals with its own sensorimotor activation. The couplings are represented in a neural network architecture called DRAMA, which implements a Willshaw network.

In the work of (Rosenstein & Cohen 1998) a robot has been developed that categorizes its sensorimotor activity by categorizing the time series of this activity with the so-called *method of delays*. The sensorimotor activity is described by delay vectors, which reconstruct the original time-series in its phase space. By categorizing these delay vectors with the delay vectors of stored prototypes, the robots construct clusters of concepts like *chase*, *escape*, *contact* etc..

Steels and Vogt have conducted experiments where two autonomous mobile robots develop a grounded and shared lexicon from scratch (Steels & Vogt 1997). The robots play a series of so-called *language games* in which they try to communicate the name of a light source that is in their environment. When a language game fails, the robots may adapt their memory in order to succeed in the future.

This paper reports on an experiment that implements the latter work, which has been reported in (Vogt 2000). Difference with the earlier work is that the current experiment 1) incorporates a prototype based categorization, similar to (De Jong & Vogt 1998), 2) reveals significant improvement of performance and 3) is more controlled, so the results can be analyzed more reliably. The next section presents the symbol grounding problem in some more detail and gives a workable definition of a symbol. The implemented model is explained in the section called 'language games'. The subsequent section presents the experimental results. And the final section provides a discussion.

The symbol grounding problem

The symbol grounding problem occupies many robotics scientists as each robot that uses symbols has to deal with this problem in one way or another. Stevan Harnad identified three subphases needed to solve this problem: (1) iconization, (2) discrimination and (3) identification (Harnad 1990).

Iconization is the process in which the sensing takes place. Sensing the world leads to the formation of what Harnad calls an *iconic representation*. This can be compared with the representation of the sensing of a real world object on the retina.

When an agent has formed iconic representations of the objects sensed in the world, the agent has to find if and how one representation is distinctive from another. This is called the problem of *discrimination*. Discrimination still yields sub-symbolic information and can be highly ambiguous (Harnad 1990).

Finding categorizations that *invariantly* recognizes the object is called *identification*. Identification yields the sym-

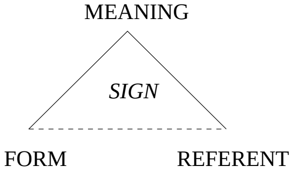


Figure 1: The semiotic triangle illustrates the relations between referent, form and meaning that constitute a sign.

bolic structure that one is interested in. Although the first two subproblems may be very complex, their implementation in this paper is relatively simple. The main focus of this paper deals with the process of identification.

In Harnad's description a symbol is defined in terms of physical symbol systems (Newell 1980). More specifically, it is defined as a *category name* (Harnad 1993). The problem with such a definition is that there is no direct link to reality; this is why the symbol grounding problem is raised in the first place.

In this paper an alternative definition of a symbol is used. This definition is consistent with a physical symbol system and is taken from the theory of semiotics. Such a symbol will be called a *semiotic symbol* to indicate its distinction from the conventional use of symbols. In this theory according to C.S. Peirce (Peirce 1931) a symbol can be equaled with a *sign*. A sign is defined as a relation between a *referent*, *form* and *meaning* as illustrated in the semiotic triangle, see figure 1. Following (Chandler 1994), the three elements are defined as follows[1]:

Referent A referent is the thing "to which the sign refers".

Form A form is "the form which the sign takes (not necessarily material)".

Meaning The meaning is "the sense made of the sign".

According to Peirce, a sign becomes a (semiotic) *symbol* when its form, in relation to its meaning "is arbitrary or purely conventional, so that the relationship must be learned" (Chandler 1994). The relations can be conventionalized in language. A nice side effect of this definition is that the semiotic symbol is *per definition* grounded, because the symbol has an intrinsic relation with the referent. The only task that remains to be solved is to construct the semiotic relation, which remains a hard problem.

Related to the symbol grounding problem is the *anchoring problem* (Coradeschi & Saffiotti 2000). According to Coradeschi and Saffiotti, "anchoring is the process of creating and maintaining the correspondence between symbols and percepts that refer to the same physical object"[2] (Coradeschi & Saffiotti 2000). As with the symbol grounding problem, an anchor in the current paper is created upon the construction of the semiotic symbol. As long as the semiotic symbol proves to be a good one (i.e. when it is re-applicable), the anchor is maintained. This is because, by definition, the semiotic symbol relates to the referent whether or not it is perceived, hence the dotted line in the semiotic triangle (Chandler 1994).

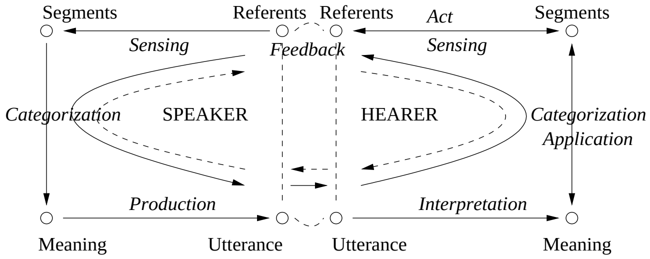


Figure 2: The semiotic square illustrates the guessing game scenario. The two squares show the processes of the two participating robots. This figure is adapted from [Steels and Kaplan, 1999].

Special notice should be given to the notion of meaning. As is assumed here, a meaning is a distinctive categorization of the perception of a referent that is *used* in the language. The first constraint is cf. (Peirce 1931), the latter is cf. (Wittgenstein 1958). Furthermore, it is assumed that the meaning co-evolves with the lexicon (Steels 1997). This means that the development of the lexicon gives rise to the development of meaning and vice versa.

Language games

The goal of the experiment presented here is that two mobile robots develop a shared and grounded lexicon from scratch about the objects that the robots can detect in their environment. This means that the robots construct a vocabulary of form-meaning associations from a tabula rasa with which the robots can successfully communicate the names of the objects.

In the experiments, the robots play a series of *guessing games* (Steels & Kaplan 1999). A guessing game is a variant of the language games in which the hearer of the game tries to guess what referent the speaker tried to verbalize. The basic scenario of a guessing game is illustrated in what Steels calls the semiotic square (Steels & Kaplan 1999), see figure 2. A guessing game is played by two robots. One robot takes the role of the speaker, while the other takes the role of the hearer. Both robots start sensing their surroundings, after which the sensory data is preprocessed. The robots categorize the preprocessed data, which results in a meaning (if used in the communication). After categorization, the speaker produces an utterance and the hearer tries to interpret this utterance. When the meaning of this utterance applies to one or more segments of the sensed referents, the hearer can act to the appropriate referent. The guessing game is successful if both robots communicated about the same referent. The robots evaluate feedback about the outcome of the guessing game, i.e. whether the robots com-

[1] Note that the terminology is somewhat different than Peirce's. Peirce uses *object*, *representamen* and *interpretant* for referent, form and meaning resp.. The adopted terminology is consistent with the terminology used by (Steels & Kaplan 1999)

[2] Note that Coradeschi and Saffiotti use the term symbol in the classical sense. Their use does not refer to the semiotic symbol.

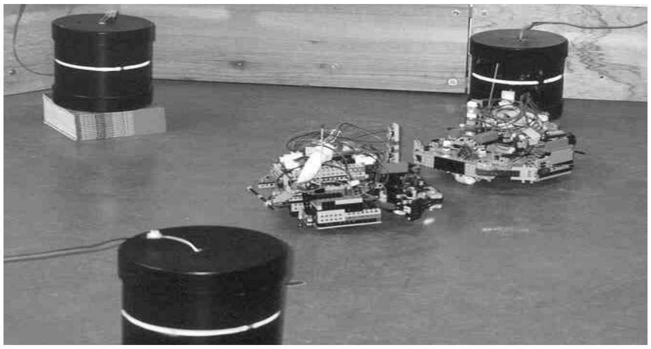


Figure 3: The LEGO robots in their environment.

municated about the same referent. This feedback is passed back to both robots so that they can adapt their ontology of categories (used as meanings) and lexicon.

Below follows a description of the experiments. For more details, consult (Vogt 2000).

The experimental setup

The experiment makes use of two LEGO robots. The robots are equipped with a.o. four light sensors, two motors, a radio module and a sensorimotor board, see figure 3. The light sensors are used to detect the objects in the robots' environment. The two motors control the robots movements. And the radio module is used to coordinate the two robots' behavior and to send sensor data to a PC where most of the processing takes place.

The robots are situated in a small environment ($2.5 \times 2.5m^2$) in which four light sources are placed at different heights. The light sources act as the objects that the robots try to name. The different light sensors of the robots are mounted at the same height as the different light sources. Each sensor outputs its readings on a *sensory channel*. A sensory channel is said to *correspond* with a particular light source if the sensor has the same height as this light source.

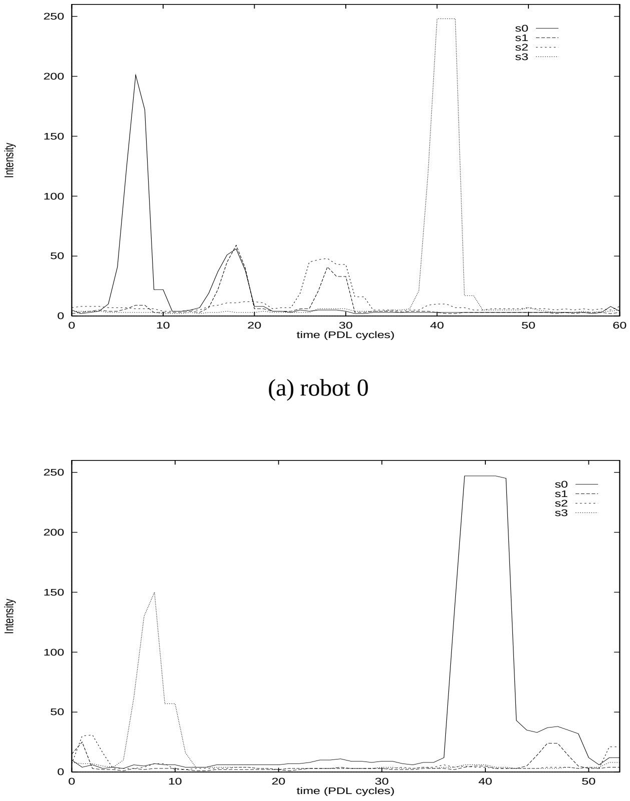
The goal of the experiments is that the robots develop a lexicon with which they can successfully name the different light sources.

Sensing, segmentation and feature extraction

As a first step towards solving the symbol grounding problem, the robots have to construct what Harnad calls an iconic representation. In this work, it is assumed that an iconic representation is the preprocessed sensory data that relates to the sensing of a light source. The resulting representation is called a *feature vector* in line with the terminology from pattern recognition, see e.g. (Fu 1976).

A feature vector is acquired in three stages: sensing, segmentation and feature extraction. The remainder of this section explains these three stages in more detail.

Sensing During the sensing phase, the robots detect what is in their surroundings one by one. They do so by rotating 720^o around their axis. While they do this, they record the sensor data of the middle 360^o part of the rotation. This way the robots obtain a spatial view of their environment for each four light sensors, see figure 4.



(a) robot 0

(b) robot 1

Figure 4: The sensing of the two robots during a language game. The plot shows the spatial view of the robots' environment. It is acquired during 360^o of their rotation. The figures make clear that the two robots have a different sensing, since they rotate at different locations. The y-axis shows the intensity of the sensors, while the x-axis determines the time (or angle) of the sensing in PDL units. A PDL unit takes about $\frac{1}{40}$second, hence the total time of these sensing events took circa $1.5s$.

The robots rotate 720^o instead of 360^o in order to cancel out nasty side effects induced by the robots' acceleration and deceleration.

Figure 4 shows the sensing of two robots during a language game. Figure (a) shows that *robot 0* clearly detected the four light sources; there appears a peak for every light source. The two robots do not sense the same view as can be seen in figure (b). This is due to the fact the robots are not located at the same place.

Segmentation Segmentation is used by the robots to extract the sensory data that is induced by the detection of the light sources. The sensing of the light sources relates in the raw sensory data with the peaks of increased intensity. As can be seen in figure 4, in between two peaks the sensory channels are noisy. So, when the noise is reduced from the sensory data, all connected regions that have non-zero sensory values for at least one sensory channel relate to the sensing of a light source.

The segmentation results in a set of segments $\{S_k\}$, where $S_k = \{\mathbf{s}_{k,0}, \ldots, \mathbf{s}_{k,n-1}\}$, where n is the number of sensory channels (4 in this case) and $\mathbf{s}_{k,i} = (\tau_{k,i,0}, \ldots, \tau_{k,i,m})$. The sensory channel data $\tau_{k,i,j}$ represents the sensory data *after* noise reduction and for all $j = 0, \ldots, m$ there exist a $\tau_{k,i,j} > 0$ (m is the length of the segment). Each segment is assumed to relate to the detection of one light source.

The set of segments constitute what is called the *context* of the guessing game, i.e. $Cxt = \{S_1, \ldots, S_N\}$, where N is the number of segments that are sensed. Each robot participating in the guessing game has its own context which may differ from another.

Feature extraction In order to allow efficient categorization based on invariant properties of the detected segments, features are extracted. Each segment may be of different length. But, in order to categorize the segments efficiently, it is useful to have a consistent representation of the segment. In order to allow a proper categorization of the segments it is useful to extract some invariant features from these segments. Thus the feature extraction is used to describe a segment with a consistent representation that contains some invariant property of the detected signal.

One can see in figure 4 that for each peak there is a different sensory channel that has the highest intensity. After segmentation, each peak is described by a segment that contains the sensory data after noise reduction. The sensory channel with the highest intensity *corresponds* with the light source the segment relates to. This correspondence can be used as an invariant property of the segments.

For all sensory channels, the feature extraction is a function $\varphi(\mathbf{s}_{k,i})$ that normalizes the maximum intensities of sensory channel i to the overall maximum intensity from segment S_k, cf. eq. (1). This way the function extracts the invariant property that the feature of the sensory channel with the overall highest intensity (i.e. the corresponding one) has a value of 1, whereas all other features have a value ≤ 1.

$$\varphi(\mathbf{s}_{k,i}) = \frac{\max_{\mathbf{s}_{k,i}}(\tau_{k,i,j})}{\max_{S_k}(\max_{\mathbf{s}_{k,i}}(\tau_{k,i,j}))} \qquad (1)$$

The result of applying a feature extraction to the data of sensory channel i will be called a feature f_i, so $f_i = \varphi(\mathbf{s}_{k,i})$.

Segment S_k can now be related to a feature vector $\mathbf{f}_k = (f_0, \ldots, f_{n-1})$, where n is the total number of sensory channels. Like a segment, a feature vector is assumed to relate to the sensing of a referent. The space that spans all possible feature vectors $\mathbf{f}$ is called the n dimensional feature space $\mathcal{F} = [0,1]^n$, or *feature space* for short.

The reasons for this feature extraction are manifold. First, it is useful to have a consistent representation of the sensed referents in order to categorize. Second, the normalization to the maximum intensity within the segment (the 'invariance detection') is useful to deal with different distances between the robot and the light source. Furthermore, it helps to analyze the experiments from an observer's point of view and to evaluate feedback. Besides its use during feedback (see below), the robots are not 'aware' of this invariance.

All above processes are carried out by each robot individually. These processes resemble the iconization process of the symbol grounding.

Categorization

In order to form a semiotic symbol from the sub-symbolic feature vectors, the robots have to relate the vectors to categories that are stored in their memories and that may be used as the meaning of the referent the feature vector relates to. The categorization of some referent should be distinctive from the categories relating to the other referents, so it can be used as the meaning of this referent.

Categorization is modeled with the so-called *discrimination games* (Steels 1996b). Each robot individually plays a discrimination game for the (potential) *topic(s)*. A topic is a segment from the constructed context (described by its feature vector). The topic of the speaker is arbitrarily selected from the context and is the subject of communication. As the hearer tries to guess what the speaker intends to communicate, it considers all segments in its context as a *potential topic*. The result of the discrimination game should be a categorization that distinguishes the topic from all other segments in the context. When the robot fails to find such a categorization, it can adapt its ontology in which the categories are stored.

Let a category $c = \langle \mathbf{c}, \nu, \rho, \kappa \rangle$ be defined as a region in the feature space $\mathcal{F}$. It is represented by a prototype $\mathbf{c} = (x_0, \ldots, x_{n-1})$, where n is the dimension of $\mathcal{F}$, and ν, ρ and κ are scores. The category is the region in $\mathcal{F}$ in which the points have the nearest distance to $\mathbf{c}$.

A feature vector $\mathbf{f}$ is categorized using the *1-nearest neighbor algorithm*, see e.g. (Fu 1976). Each robot categorizes all segments this way.

In order to allow generalization and specialization of the categories, different versions of the feature space $\mathcal{F}_\lambda$ are available to a robot. In each space a different resolution is obtained by allowing each dimension of $\mathcal{F}_\lambda$ to be exploited up to 3^λ times. How this is done will be explained soon.

The different feature spaces allow the robots to categorize a segment in different ways. The categorization of segment S_k results in a set of categories $C_k = \{c_0, \ldots, c_m\}$, where $m \leq \lambda$.

Suppose that the robots want to find distinctive categories for (potential) topic S_t, then a distinctive category set can be defined as follows:

$$DC = \{c_i \in C_t \mid \forall (S_k \in Cxt \backslash \{S_t\}) : \neg c_i \in C_k\}$$

Or in words: the distinctive category set consists of all categories of the topic that are not a category of any other segment in the context.

If $DC = \emptyset$, the discrimination game is a failure and some new categories are constructed. Suppose that the robot tried to categorize feature vector $\mathbf{f} = (f_0, \ldots, f_{n-1})$, then new categories are created as follows:

1. Select an arbitrary feature $f_i > 0$.
2. Select a feature space $\mathcal{F}_\lambda$ that has not been exploited 3^λ times in dimension i for λ as low as possible.
3. Create new prototypes $\mathbf{c_j} = (x_0, \ldots, x_{n-1})$, where $x_i = f_i$ and the other x_r are made of already existing prototypes in $\mathcal{F}_\lambda$.
4. Add the new prototypical categories $c_j = \langle \mathbf{c}_j, \nu_j, \rho_j, \kappa_j \rangle$ to the feature space $\mathcal{F}_\lambda$, with $\nu = \rho = 0.01$ and $\kappa = 1 - \frac{\lambda}{\lambda_{\text{max}}}$.

The score ν indicates the effectiveness of a category in the discrimination games, ρ indicates the effectiveness in categorization and κ indicates how general the category is (i.e. in which feature space $\mathcal{F}_\lambda$ the category houses). κ is a constant, based on the feature space $\mathcal{F}_\lambda$ and the feature space that has the highest resolution possible (i.e. $\mathcal{F}_{\lambda_{\text{max}}}$). The other scores are updated similar to reinforcement learning. It is beyond the scope of this paper to give exact details of the update functions.

The three scores together constitute the meaning score $\mu = \frac{1}{3}(\nu + \rho + \kappa)$, which is used in the naming phase of the experiment. The influence of this score is small, but it helps to select a form-meaning association in case of an impasse.

The reason to exploit only one feature of the topic and to combine it with existing prototypes, rather than adopting the complete feature vector is to speed up the construction of categories.

If the distinctive category set $DC \neq \emptyset$, the discrimination game is a success. The DC is forwarded to the naming game that models the naming phase of the guessing game. If a category c is used successfully in the guessing game, the prototype $\mathbf{c}$ of this category is moved towards the feature vector $\mathbf{f}$ it categorizes:

$$\mathbf{c} := \mathbf{c} + \epsilon \cdot (\mathbf{f} - \mathbf{c}) \tag{2}$$

where ϵ is the step size with which the prototype moves towards $\mathbf{f}$. This way the prototype becomes a more representative sample of the feature vectors it categorizes.

The discrimination game models the discrimination phase in the symbol grounding problem.

Naming

After both robots have obtained distinctive categories of the (potential) topic(s) from the discrimination game as explained above, the *naming game* (Steels 1996a) starts. In the naming game, the robots try to communicate the topic.

The speaker tries to produce an utterance as the name of one of the distinctive categories of the topic. The hearer tries to interpret this utterance in relation to distinctive categories of its potential topics. This way the hearer tries to guess the speaker's topic. If the hearer finds a possible interpretation, the guessing game is successful if both robots communicated about the same referent. This is evaluated by the feedback process as will be explained below. According to the outcome of the game, the lexicon is adapted.

The lexicon L is defined as a set of form-meaning associations: $L = \{\text{FM}_i\}$, where $\text{FM}_i = \langle F_i, M_i, \sigma_i \rangle$ is a lexical entry. Here F_i is a form that is made of a combination of consonant-vowel strings, M_i is a meaning represented by some category, and σ is the association score that indicates the effectiveness of the lexical entry in the language use.

Production The speaker of the guessing game tries to name the topic. To do this it selects a distinctive category from DC for which the meaning score μ is highest. Then it searches its lexicon for form-meaning association of which the meaning matches this distinctive category.

If it fails to do so, the speaker will first consider the next distinctive category from DC. If all distinctive categories have been explored and still no entry has been found, the speaker may create a new form as will be explained in the adaptation section.

If there are one or more lexical entries that fulfill the above condition, the speaker selects that entry that has the highest association score σ. The form that is thus produced is uttered to the hearer.

Understanding On receipt of the utterance, the hearer searches its lexicon for entries for which the form matches the utterance, and the meaning matches one of the distinctive categories of the potential topics.

If it fails to find one, the lexicon has to be expanded, as explained later.

If the hearer finds more than one, it will select the entry that has the highest score $\Sigma = \sigma + \alpha \cdot \mu$, where $\alpha = 0.1$ is a constant weight. The potential topic that is categorized by the distinctive category that matches the meaning of the lexical entry is selected by the hearer as *the* topic of the guessing game. I.e. this segment is what the hearer guessed to be the subject of communication.

Feedback In the feedback, the outcome of the guessing game is evaluated. This outcome is known to both robots.

As mentioned, the guessing game is successful when both robots communicated about the same referent. This feedback is established by comparing the feature vectors of the two robots relating to the topics. Previous attempts to implement feedback physically have failed, therefore it is assumed that the robots can do this. Naturally, this problem needs to be solved in the future.

If the hearer selected a topic after the understanding phase, but if this topic is not consistent with speaker's topic, there is a *mismatch in referent*.

If the speaker has no lexical entry that matches a distinctive category, or if the hearer could not interpret the speaker's utterance because it does not have a proper lexical entry in the current context, then the guessing game is a failure.

Adaptation Depending on the outcome of the game, the lexicon of the two robots is adapted. There are four possible outcomes/adaptations:

1. The speaker has no lexical entry: In this case the speaker creates a new form and associates this with the distinctive category it tried to name. This is done with a probability

$P_s = 0.1$.

2. The hearer has no lexical entry: The hearer adopts the form uttered by the speaker and associates this with the distinctive categories of a different randomly selected segment from its context.
3. There was a mismatch in referent: Both robots adapt the association score σ of the used lexical entry: $\sigma := \eta \cdot \sigma$, where $\eta = 0.9$ is a constant learning rate. In addition, the hearer adopts the utterance and associates it with the distinctive categories of a different randomly selected segment.
4. The game was a success: Both robots reinforce the association score of the used entry: $\sigma := \eta \cdot \sigma + (1 - \eta)$. In addition, they lower competing entries (i.e. entries for which either the form or the meaning is the same as in the used entry): $\sigma := \eta \cdot \sigma$. The latter update is called lateral inhibition.

The guessing game described above implements the three mechanisms hypothesized by Luc Steels (Steels 1996a) that can model lexicon development. *(Cultural) interactions* are modeled by the sensing and communication. *Individual adaptation* is modeled at the level of the discrimination and naming game. The selection of elements and the individual adaptations power the *self-organization* of a global lexicon.

So, the selection, generation and adaptation of the lexicon cause, together with the multiple interactions a self-organizing effect in which the lexicon is structured such that a consistent communication system emerges.

The coupling of the naming game with the discrimination games and the sensing part makes that the emerging lexicon is grounded in the real world. The robots successfully solve the symbol grounding problem in some situation when the guessing game is successful. This is so, because identification (Harnad 1990) is established when the semiotic triangle (figure 1) is constructed completely. This is done at the naming level and it is successful when the guessing game is successful.

Experimental results

An experiment has been done in which the sensory data of the sensing phase during 1,000 guessing games has been recorded, see also (Vogt 2000). From this data set it has been calculated that the a priori chance of success when the robots randomly choose a topic is 23.5%. Because the robots do not always detect all the light sources that are present (figure 4), their context is not always coherent. This leads to the fact that there is a maximum success rate that can be reached, called the *potential understandability*. The potential understandability has been calculated to be on the average 79.5%.

The 1,000 recorded situations have been processed on a PC in 10 runs of 10,000 guessing games. Figure 5 shows the communicative- and discriminative success of this experiment. The communicative success measures the number of successful guessing games, averaged over the past 100 games. The discriminative success measures the number of successful discrimination games, also averaged over the past 100 guessing games.

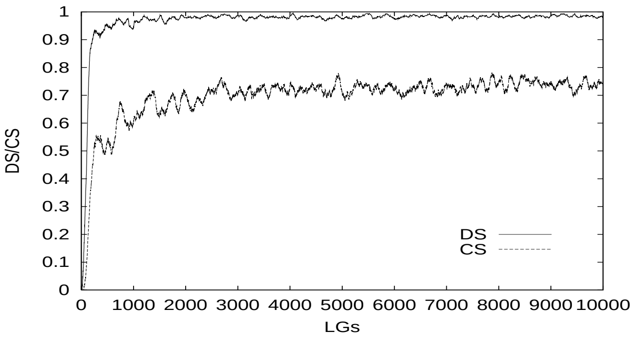


Figure 5: The communicative success (CS) and discriminative success (DS) of the experiments.

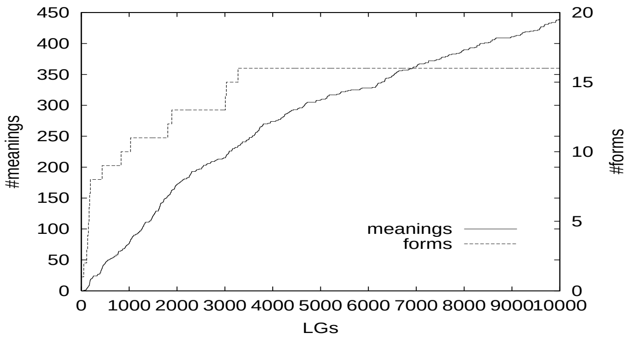


Figure 6: The evolution of the number of meanings and forms that have been used successfully by the robots in the experiment.

As can be seen, the discriminative success reaches a value near 1 very fast. Hence, the robots are well capable of finding distinctive categories for the sensed light sources.

The communicative success is somewhat lower. It increases towards a value slightly below 0.8 near the end. Since this is close to the potential understandability, the robots are well capable to construct a shared lexicon within its limits.

Figure 6 shows the number of different meanings and forms that have been used successfully in the guessing games. As can be seen, the number of meanings are much higher than the number of forms used. There are approximately $28\times$ more meanings used than forms. So, although the robots construct many meanings in relation to the four referents, the robots only use 16 forms to name them.

When investigating the results in more detail it has been observed that the most frequently used forms count up to 7, see figure 7. Moreover, when these forms are used, they almost uniquely refer to one referent. The number of meanings most frequently used are also substantially lower as in figure 6. It is clear that the referents are categorized differently in the different games, but that they are named rather consistently. Although there is some synonymy (one referent is named with on the average two forms most frequently), the amount of polysemy is very low (a form usu-

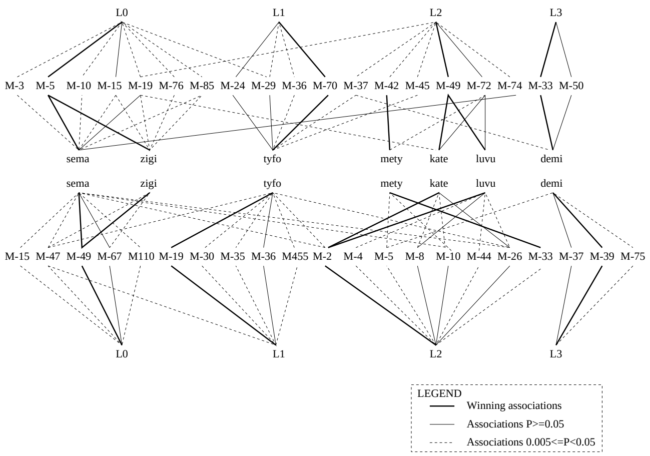


Figure 7: The lexicon of a typical run after 10,000 guessing games shown as a semiotic landscape. This figure shows the relation between the referents (L), meanings (M) and forms of both robots. The connections indicate the occurrence frequencies of their use relative to the occurrence frequency of the referent (between referent and meaning) and relative to the occurrence frequency of the form (between form and meaning).

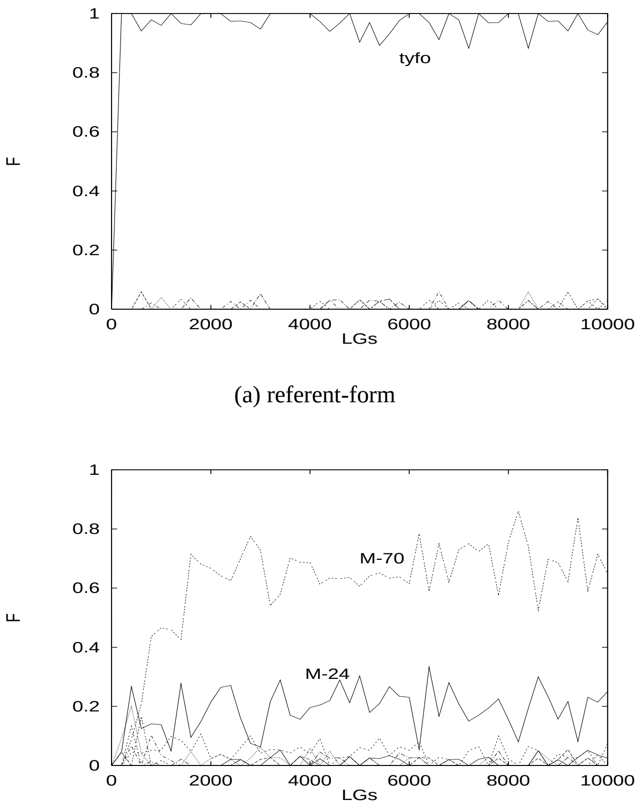
ally names only one referent).

How the occurrence frequencies of the used forms evolve is shown in the competition diagram of figure 8 a. As this figure makes clear, the most frequently used form clearly wins the competition to name light source L1. At the bottom of the diagram, other forms reveal a weak competition. Similar competitions have been observed for the other referents. Such competitions show how the forms are well anchored to the referents they name. It shows that the forms maintain a relation with their referents and could therefore be used in the absence of the referents, which is one of the key issues of anchoring symbols (Coradeschi & Saffiotti 2000). This property is achieved by allowing the forms to be coupled with multiple meanings that categorize a referent on different occasions. As a result, the competition between meanings to categorize a referent is stronger (figure 8 b).

Discussion

This paper presents research that investigates how two mobile robots can solve the symbol grounding problem in a particular experimental setup by employing language games. An alternative definition of a symbol has been adopted from Peirce's theory of semiotics, which has been coined a *semiotic symbol*. A semiotic symbol is defined by the relation between a form, a meaning and a referent.

The experimental results show the robots can do so without any preprogrammed semiotic symbols; these are constructed along the way. For this the robots categorize preprocessed sensory data, which they then incorporate to name them. The semiotic symbols are always constructed to name some referent. When either the categorization or the naming fails, the robots adapt their memories so that they may improve performance on future occasions.



(a) referent-form

(b) referent-meaning

Figure 8: (a) The referent-form competition diagram from the same run as in figure 7. This diagram shows the competition between forms to name referent light source L1. (b) The referent-meaning diagram shows the competition between meanings to interpret light source L1. In both diagrams the y-axis shows the occurrence frequencies of successfully used forms or meanings over the past 200 games relative to the occurrence of the referent. The x-axis shows the number of games played.

The results suggest that the co-evolution of meaning and form is very crucial in the robots' ability to solve the symbol grounding problem. As the sensing of the robots may differ a lot in different situations, there emerge many different distinctive categories (which are employed as meanings). Hence there are one-to-many relations between referent and meaning. On the other hand, the interactive adaptation during the guessing games, allows the emergence of one-to-many relations between form and meaning. Thus the invariance of the identification (needed to solve the symbol grounding problem, cf. (Harnad 1990)) creeps in at the form-meaning level. This leads to the conclusion that communication is very beneficial in solving the symbol grounding problem. The same conclusion applies to the anchoring problem as well.

When comparing the results of this experiment with the ones reported in (Billard & Hayes 1998), where a student robots learns a vocabulary from a teacher robot, one sees one striking similarity and one difference. Most important similarity is that the student robot learns and uses the vocabulary correctly in about 71% of the communication. The errors in learning were mainly due to differences in sensorimotor activation between the two robots, i.e. the inability to construct a coherent context. This is the same reason why the robots in this experiment could not communicate correctly to a very high degree, cf. the potential understandability. The main difference with Billard and Hayes' work is the learning speed. In their experiments, the student robot learns the vocabulary in approximately 70 teaching examples. In these experiments it takes about 1,000 guessing games until a high level of success is reached. This difference is to be sought in the fact that the teacher robot in Billard and Hayes' experiments is preprogrammed with the lexicon. In the guessing games, both robots start from scratch. Clearly, the latter task is much more difficult, because initially much synonymy and polysemy enter the lexicon in both robots. It then takes a while to disambiguate the system.

It should be noticed that the current experiment solves the symbol grounding problem in a relatively simple experiment. It remains to be shown whether the model still works in a more complicated environment and with different sensor modalities. One such experiment is the Talking Heads experiment (Steels & Kaplan 1999) where two immobile robots equipped with a camera play guessing games to develop a lexicon about geometrical figures. This experiment shows that the model works with a different sensor modality, although the robots' environment is still rather simple. The model is also used to investigate human-robot interaction (Kaplan 2000). To allow proper human-robot interaction, the robots should be equipped with a sensor modality that has similarities to the human sensory system so that the robot can develop human-like categories. Further work that investigates the scalability of the model is currently in progress.

Acknowledgments

The research has been carried out at the AI Lab of the Vrije Universiteit Brussel resulting in a Ph. D. thesis under the supervision of Luc Steels. Luc Steels is thanked for his expert advice and inspiration. Also many thanks to all other colleagues of the lab for their support in way or another.

References

Billard, A., and Hayes, G. 1998. Transmitting communication skills through imitation in autonomous robots. In Birk, A., and Demiris, J., eds., *Learning Robots, Proceedings of the EWLR-6, Lecture Notes on Artificial Intelligence 1545*. Springer-Verlag.

Chandler, D. 1994. Semiotics for beginners. http://www.aber.ac.uk/media/Documents/S4B/semiotic.html.

Coradeschi, S., and Saffiotti, A. 2000. Anchoring symbols to sensor data: preliminary report. In *Proceedings of the Seventeenth National Conference on Artificial Intelligence (AAAI-2000)*, 129–135.

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

Fu, K., ed. 1976. *Digital Pattern Recognition*. Berlin: Springer-Verlag.

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.

Kaplan, F. 2000. Talking aibo : First experimentation of verbal interactions with an autonomous four-legged robot. In Nijholt, A.; Heylen, D.; and Jokinen, K., eds., *Learning to Behave: Interacting agents CELE-TWENTE Workshop on Language Technology*.

Newell, A. 1980. Physical symbol systems. *Cognitive Science* 4:135–183.

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

Rosenstein, M., and Cohen, P. R. 1998. Concepts from time series. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence*. Menlo Park Ca.: AAAI Press.

Steels, L., and Kaplan, F. 1999. Situated grounded word semantics. In *Proceedings of IJCAI 99*. Morgan Kaufmann.

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

Steels, L. 1996a. Emergent adaptive lexicons. In Maes, P., 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 Tokoro, M., ed., *Proceedings of the International Conference on Multi-Agent Systems*. Menlo Park Ca.: AAAI Press.

Steels, L. 1997. Synthesising the origins of language and meaning using co-evolution, self-organisation and level formation. In Hurford, J.; Knight, C.; and Studdert-Kennedy, M., eds., *Approaches to the evolution of language*. Cambridge: Cambridge University Press.

Vogt, P. 2000. *Lexicon Grounding on Mobile Robots*. Ph.D. Dissertation, Vrije Universiteit Brussel. http://www.cs.unimaas.nl/p.vogt/thesis.html.

Wittgenstein, L. 1958. *Philosophical Investigations*. Oxford, UK: Basil Blackwell.