Adaptive grounding of lexicons on mobile robots

Vogt, P.A.

Published in:
Robo Sapiens

Publication date:
2001

Adaptive Grounding of Lexicons on Mobile Robots

Paul Vogt
IKAT Universiteit Maastricht
paul@cs.unimaas.nl

Abstract

This paper reports on experiments done on mobile robots. In the experiments the robots develop a lexicon of which the meaning is grounded in the real world. The experiments are based on an approach that is in line with the embodied cognition paradigm. Robots construct a lexicon by engaging in a series of language games. In a language game, two robots try to name an object they can perceive in their environment. When they fail to communicate successfully, they adapt their memories in order to improve performance in future games. This way a shared lexicon is formed. The experiments showed that the robots can do this fairly well within a simple experimental setup.

1 Introduction

Symbol grounding is one of the hardest problems in robotics research. How can robots develop and use symbols that have their meaning grounded in the real world? Or in other words: how do seemingly meaningless symbols acquire their meaning in relation to the real world? This is the main question of what is known as the *symbol grounding problem* [Harnad, 1990].

Several attempts have been made in the past to solve this problem. Some of these investigated the problem in relation to language acquisition, e.g. [Billard and Hayes, 1997; Sugita and Tani, 2000; Yanco and Stein, 1993]. In these experiments, however, (parts) of the language has been predefined. In the work of Billard and Hayes, for instance, a teacher robot has been preprogrammed with the language and a student robot learns it by imitation. Jun Tani provides the language by human-robot interaction. And robots in Yanco and Stein's work have also been preprogrammed with the signals that exist in their language and meaning. In their work, the robots learn to associate the proper signals with the proper meanings by means of reinforcement learning where feedback about the robots' performance is provided by a human instructor. So, all these investigations require some input from a human.

In the past few years, research at the AI Lab of the Vrije Universiteit Brussel has been focussed on the origins of lexicons on physical robots, see e.g. [Belpaeme et al., 1998; Steels, 1997; Steels and Vogt, 1997]. In these experiments real robots have been developed that interact with their environment, including each other. The interactions are modeled by so-called language games [Steels, 1996a]. In a language game, two robots try to communicate the name of some object they observed in their environment. They first sense what is in their environment. The sensing is then preprocessed and categorized. Then the speaker of the game (one of the two robots) tries to name the categorization of one of the objects. In turn, the hearer (the other robot) tries to interpret this name in relation to one of the categorized objects. Afterwards, feedback is provided whether or not the two robots communicated the same object. Depending on the outcome, the lexicon is adapted to improve their future performance.

The development of the lexicon is based on three mechanisms that define an *adaptive complex dynamical system*. These mechanisms are: (cultural) interaction, individual adaptation and self-organization [Steels, 1996a]. As a result of these mechanisms a shared lexicon emerges in a similar way as ant-paths are formed. The lexicon is an attractor of the adaptive complex dynamical system.

This paper is organized as follows: A brief discussion of the symbol grounding problem, its relation to embodied cognition and some adopted definitions are presented in the following section. Section 3 very briefly introduces the language game model. Some experimental results are provided in section 4. And finally, section 5 concludes.

2 The Symbol Grounding Problem

As mentioned, the symbol grounding problem [Harnad, 1990] questions how symbols acquire their meaning in relation to the real world. In Harnad's work, symbols are defined in the classical way. I.e. they are defined in terms of *physical symbol systems* as proposed by [Newell, 1980]. In Newell's definition, symbols are patterns that provide access to some distal structure. They are completely represented inside an agent's brain and have in some

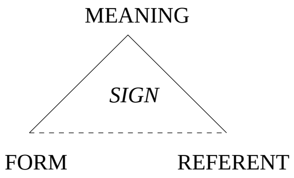


Figure 1: The semiotic triangle illustrates the relations between referent, form and meaning that constitute a symbol.

way a relation to the real world. The latter notion is subject of the symbol grounding problem.

Also around 1990 Rodney Brooks, proposed his *physical grounding hypothesis* [Brooks, 1990]. This hypothesis states that agents should have their intelligence grounded in the real world, but that they do not need symbolic representations. The intelligence could be represented by sensorimotor couplings and part of it is *situated* in reality. Hence it is an agent's interaction with reality from which intelligence emerges. Brooks' hypothesis is the foundation on which the embodied cognition paradigm is built.

So, one may argue that agents need not solve the symbol grounding problem as there are no symbols, see e.g. [Pfeifer and Scheier, 1999]. However, humans seem to use symbols. But if one comes up with an alternative definition of a symbol, symbols might still fit within the embodied cognition paradigm [Clancey, 1997]. Such an alternative is provided by the theory of semiotics. In semiotics the notion of a *sign* is central. A sign is defined by a relation between a *form*, a *meaning* and a *referent* [Chandler, 1994]:

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".

If the form in relation to the meaning is either arbitrary or conventionalized so that the relationship must be learned, the sign is called a symbol [Peirce, 1931]. The relation between the three elements are usually illustrated by a semiotic triangle, see figure 1.

An advantage of this definition of the symbol is that it is *per definition* grounded, since the referent (a real world object) is an intrinsic part of the symbol. Yet the problem remains that the semiotic triangle still has to be constructed. This is, like the symbol grounding problem a very hard problem. This is so, because when a robot detects a referent under different condition (e.g. from different positions), its sensory stimulation differs extremely. To solve the symbol grounding problem, the robot must identify these different sensings invariantly.

The fact that the semiotic definition provides a structural coupling between referent, meaning and form makes the symbol fit within the embodied cognition paradigm. It is situated and embodied. The latter means that it depends on the an agents physical interaction with the real world. The language games that have been implemented on the robots implement the interaction and the construction of the semiotic symbols.

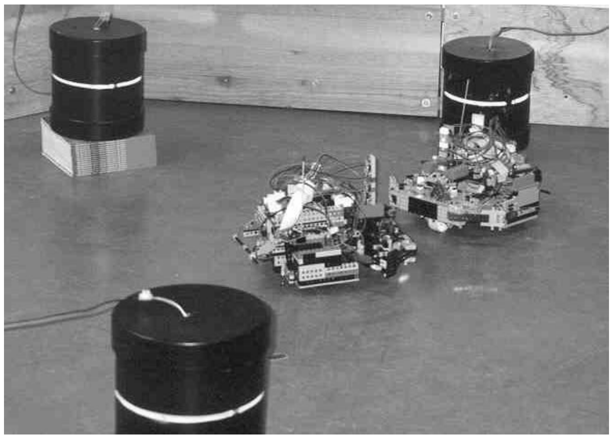


Figure 2: The robots situated in their environment as used in the experiments.

3 The Language Game Model

The experiments are done using two LEGO robots that are situated in a small environment consisting of four light sources, see figure 2. Each light source is placed at a different height and the robots are equipped with four light sensors, each mounted at a corresponding height. The aim of the experiment is that the robots develop a shared lexicon of which the meaning is grounded in their world. The lexicon is a set of form-meaning associations that relate, ideally, to some light source. Prior to the experiments, the robots have no categories and their lexicon is empty.

The lexicon development is guided by a series of *guessing games*. A guessing game is a variant of a language game in which the hearer tries to guess what light source the speaker is naming. Below follows a very brief step-by-step description of the guessing game. The interested reader is referred to [Vogt, 2000] for the details.

1. The guessing game starts when both robots are close to each other with their backs oriented towards each other.

2. One by one, the robots rotate 720^o in order to detect their environment. The sensing results in a spatial view of the middle 360^o. The sensing is described by a set of raw sensory data which is sent to a PC for further processing.

3. The sensing is preprocessed by segmentation and feature extraction processes. This results in a description of the sensing, now called a *context* of feature vectors. A feature vector is a 4 dimensional vector describing some properties of the sensing. Each feature vector is supposed to relate to the detection

of one light source. It is important to note that not always all four light sources are detected.

4. After the context is constructed, the speaker chooses one arbitrary feature vector as the topic. The hearer considers all feature vectors as a potential topic; it has to guess which feature vector is the real topic.
5. Both robots individually try to find a distinctive categorization for the (potential) topic(s). This is modeled by so-called *discrimination games* [Steels, 1996b]. The aim of a discrimination game is to find a categorization of a topic that distinguishes the topic from all other feature vectors in the context. A category is represented by a prototype (i.e. a point) in the feature space that is spanned from all possible feature vectors. When no distinctive category can be found by the discrimination game, new categories can be created for which the feature vector of the topic is used as an exemplar. If the discrimination game succeeds, the distinctive categories are forwarded to the naming part of the guessing game. If the guessing game succeeds, the prototypical category is moved towards the feature vector of the topic. This way the category becomes are representative example of the feature vectors it categorizes.
6. After both robots thus acquired distinctive categories that relate to the (potential) topic(s), the speaker tries to produce an utterance that names the distinctive category of its topic. It does so by searching its lexicon for elements of which the meaning matches the distinctive category. If there are more than one, it selects the one that has the highest association score and the associated form is uttered. The association score indicates the element's past effectiveness in the communication. If the speaker does not find such an element, it creates a new form, associates this with the distinctive category (which now becomes a meaning) and adds the new element to its private lexicon.
7. When the hearer receives an utterance, it tries to interpret it. It searches its own lexicon for elements of which the form match the utterance, *and* of which the meaning matches a distinctive category of one of the potential topics. If there are more than one such elements, the hearer selects the one that has the highest association score. The feature vector to which the matching distinctive category relates is then chosen as the hearer's topic of the game. I.e. this feature vector is what the hearer guessed the speaker has named.
8. Feedback is provided on whether the hearer found a lexical element and if so, whether both robots communicated the same topic.
9. Depending on the outcome (provided by the feedback) the lexicon is adapted. Three possible outcomes / adaptations remain:
   (a) The hearer has not found an element in its lexicon that matches the received utterance and of which the meaning is consistent in the game's context. In this case, the hearer adopts the uttered form and associates it with the distinctive categorization(s) of one arbitrarily selected feature vector. The speaker decreases the association score of the used form-meaning association. The guessing game fails.
   (b) The hearer has found a matching element, but the selected topic is not consistent with the speaker's topic. In this case, the hearer again adopts the uttered form and associates it with the distinctive categorization of an arbitrarily selected feature vector. Both robots decrease the association score of the used form-meaning association. The game fails.
   (c) Both robots have selected a lexical element in relation to a consistent topic. The hearer thus guessed right and the guessing game succeeds. Both robots increase association score of the used element and competing elements are *laterally inhibited*. An element is competing when the form matches the communicated form, but its meaning is inconsistent. It is also competing when the meaning matches the meaning of the used element, but its form is inconsistent.

The guessing game thus models a communication act. The mechanisms on which the game is based guide the formation of a shared global lexicon. The lexicon is constructed locally inside each robot and is spread through the population through cultural interactions. Score adaptations and selection drive the self-organization of the global lexicon.

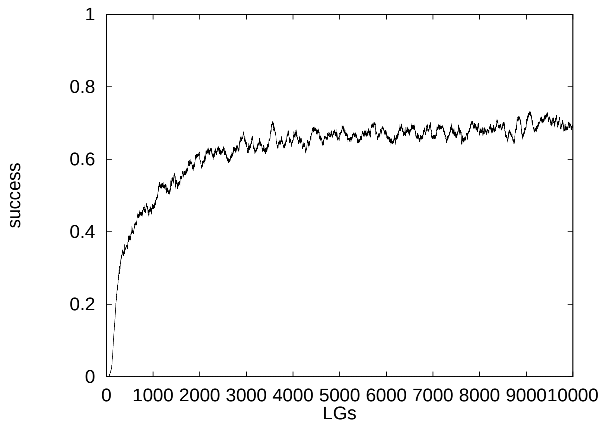
4 Experimental Results

Given the above description of the guessing games a large series of experiments have been done. The results of these experiments are reported in [Vogt, 2000]. In this section results of one such experiment is presented.

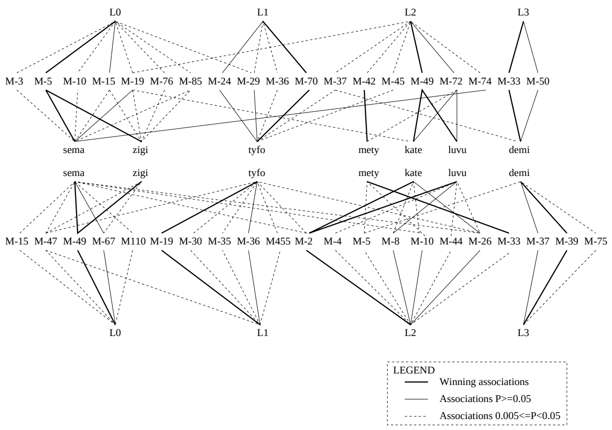
Prior to the experiments a data set of raw sensory data from the sensing of 1,000 situations is recorded. One situation consists of the sensing as if both robots play a guessing game. With this data set, a series of 10 runs have been done each experiment. One run consists of 10,000 guessing games. Each game one of both robots is arbitrary selected as the speaker, which arbitrarily selects one feature vector as the topic.

From the sensory data set, some expectations can be extracted. For instance the a priori chance that the robots choose the same topic when each selects a topic arbitrarily. This a priori chance has been calculated to be 23 %. Furthermore, since the robots not always detect all light sources, they may not have acquired a coherent context. This yield a maximum in their potential understandability. I.e. there is a maximum in the expected communicative success. This potential understandability has been calculated to be 80 %.

Figure 3 (a) shows the evolution of the communicative success. The communicative success is the success of the



(a) CS



(b) Lexicon

Figure 3: The results of the experiment. Figure (a) shows the communicative success and figure (b) shows an instance of an emerged lexicon.

guessing games averaged over the past 100 games and averaged over the 10 runs. As can be seen, the success increases towards a value around 75 %, which is much larger than the a priori chance for success and slightly lower than the potential understandability. Hence the robots seem to be pretty well capable of developing a shared lexicon.

An instance of such a lexicon is shown in figure 3 (b). This figure shows a semiotic landscape that displays the connections between referent (or light source) L, the meanings M and forms like **tyfo**. This is for both robots r0 (upper half) and r1 (lower half). The connections are weighted by their relative co-occurrence frequencies during one entire run. The connections between referent and meaning are relative to the occurrence of the referent and the connections between form and meaning are relative to the occurrence of the form.

Ideally, the graph is orthogonal with respected to the referents. This means that, ideally, the connections between referent, meaning and form for both robots should not mix. A referent may be related with several meanings and several forms, as long as there is no relation to another referent. Furthermore, each referent is ideally associated with only one form. Figure 3 (b) shows that in most cases the graph almost meets the orthogonality criterion. Connections that mix up are connections that have low co-occurrence frequencies. The figure also shows that, although there are some synonymous relations between referent and form, these are far less than the one-to-many relations between referent and meaning. Moreover, the polysemous relations between form and referent are even lower. The latter means that when a robot uses some form, it almost uniquely determines the relation to some referent. The former means that when a robot tries to name some referent, it only uses a few forms to do so.

5 Conclusions

This paper showed an experiment in which two mobile robots solved the symbol grounding problem. The robots do so by engaging in a series of guessing games in which they try to communicate a referent. While they do so, they construct a lexicon which is shared through their communication and of which the meaning is grounded in the real world. The lexicon consists of arbitrary, though conventionalized forms that are related with one or more meanings. The meanings are categorizations that the robots constructed from their sensing of the real world. This is what constitutes a situated (or semiotic) symbol, i.e. the relation between referent, meaning and form.

Given local mechanisms like categorization, naming and adaptation, the robots develop a shared lexicon. This is an emergent phenomenon. The goal of constructing a shared lexicon has not been preprogrammed, neither is the structure of it. It emerges through the self-organizing effect implemented by the local mechanisms.

As the semiotic landscape showed (figure 3 (b)), the lexicon is formed of symbols that have one-to-many relations between referent and meaning and one-to-many relations between form and meaning. The result is that the one-to-many relations between referent and meaning are canceled out by the one-to-many relations between form and meaning, yielding one-to-one or one-to-few relations between referent and form. So, the symbols are identified more or less invariantly only at the naming level. This provides a strong argument in favor of a co-evolution between language and meaning, rather than two separate evolutions as is for instance hypothesized by [Deacon, 1997].

Acknowledgments

This work has been done at the Vrije Universiteit Brussel leading to a Ph. D. thesis under the supervision of Luc Steels. The author wishes to thank Luc Steels for providing an excellent research environment and many fruitful discussions. In addition all colleagues from the VUB AI Lab are thanked for their contributions in one way or the other.

References

Belpaeme, T., Steels, L., and van Looveren, J. (1998). The construction and acquisition of visual categories. In Birk, A. and Demiris, J., editors, *Learning Robots, Proceedings of the EWLR-6, Lecture Notes on Artificial Intelligence 1545*. Springer.

Billard, A. and Hayes, G. (1997). Robot's first steps, robot's first words ... In Sorace and Heycock, editors, *Proceedings of the GALA '97 Conference on Language Acquisition - Edingburgh*, Human Communication Research Centre. University of Edinburgh.

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

Chandler, D. (1994). Semiotics for beginners. [WWW document] **URL** http://www.aber.ac.uk/media/Documents/S4B/semiotic.html.

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

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

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

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

Peirce, C. (1931). *Collected Papers*, volume I-VIII. Harvard University Press, Cambridge Ma.

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

Steels, L. (1996a). Emergent adaptive lexicons. In Maes, P., editor, *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., editor, *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 Vogt, P. (1997). Grounding adaptive language games in robotic agents. In Husbands, C. and Harvey, I., editors, *Proceedings of the Fourth European Conference on Artificial Life*, Cambridge Ma. and London. MIT Press.

Sugita, Y. and Tani, J. (2000). A connectionist model which unifies the behavioral and the linguistic processes: Results from robot learning experiments. Technical Report SCSL-TR-00-001, Sony CSL.

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

Yanco, H. and Stein, L. (1993). An adaptive communication protocol for cooperating mobile robots. In Meyer, J.-A., Roitblat, H., and Wilson, S., editors, *From Animals to Animats 2. Proceedings of the Second International Conference on Simulation of Adaptive Behavior*, pages 478–485, Cambridge Ma. The MIT Press.