The physical symbol grounding problem

Paul Vogt[a*†]

[a]Universiteit Maastricht, IKAT / Infonomics
P.O. Box 616, 6200 MD Maastricht, The Netherlands

This paper presents an approach to solve the symbol grounding problem within the framework of embodied cognitive science. It will be argued that symbolic structures can be used within the paradigm of embodied cognitive science by adopting an alternative definition of a symbol. In this alternative definition, the symbol may be viewed as a structural coupling between an agent's sensorimotor activations and its environment. A robotic experiment is presented in which mobile robots develop a symbolic structure from scratch by engaging in a series of language games. In this experiment it is shown that robots can develop a symbolic structure with which they can communicate the names of a few objects with a remarkable degree of success. It is further shown that, although the referents may be interpreted differently on different occasions, the objects are usually named with only one form.

1. Introduction

This paper tries to show how symbols can be redefined to describe cognitive functions within the paradigm of embodied cognition. Traditionally, cognitive scientists describe cognition in terms of symbol systems (Newell and Simon, 1976; Newell, 1980). This is very useful because they assume that cognitive agents manipulate symbols when they think, reason or use language. However, when explaining the processes underlying such (higher) cognitive functions in terms of symbol manipulation, at least two major fundamental problems arise: the *frame problem* (McCarthy and Hayes, 1969; Pylyshyn, 1987) and the *symbol grounding problem* (Harnad, 1990). These problems arise because symbols are defined as internal representations, which are supposed to relate to entities in the real world.

A recent approach in cognitive science tries to overcome these problems by describing cognition in terms of the dynamics of an agent's interaction with the world. This novel approach has been called *embodied cognitive science*, e.g., (Pfeifer and Scheier, 1999). In embodied cognitive science it is assumed that intelligence can be described in terms of an agent's bodily experiences that are acquired through its interaction with its environment. I.e., an agent's intelligence should be based on its past interactions with the physical world. Within this paradigm, the frame problem and the symbol grounding problem are avoided to some degree, because it is argued that symbol representations are no longer necessary to implement intelligent behaviors (Brooks, 1990).

But is this true? Are symbols no longer necessary? Indeed much can be explained without using symbolic descriptions, but most of these explanations only dealt with low-level reactive behaviors such as obstacle avoidance, phototaxis, simple forms of categorization and the like (Pfeifer and Scheier, 1999). Higher cognitive functions such as language processing have been modeled most successfully using symbolic representations. This can be inferred from the fact that most natural language processing applications use symbolic processing, either hand-coded or acquired statistically from large corpora, see, e.g., (ACL, 1997). It might therefore still be desirable to describe higher cognition in terms of symbol systems. However, to overcome the symbol grounding problem, the symbol system has to be embodied and situated (Pfeifer and Scheier, 1999). It has to be embodied in order to experience the world and it has to be situated so that

*The experimental work in this paper has been carried out at the Artificial Intelligence Laboratory of the Vrije Universiteit Brussel.
†*E-mail address:* p.vogt@cs.unimaas.nl

it can acquire its knowledge through interactions with the world, see, e.g., (Sun, 2000) for a discussion. This leads to the central question of this paper: Is it possible to define and develop an embodied and situated symbol system?

This paper argues that this could be possible by adopting Peirce's triadic definition of symbols. To distinguish this triadic definition from the traditional definition, these symbols will be called *semiotic symbols*. As will be shown these semiotic symbols are both embodied and situated. The proposed approach combines the paradigm of *physical grounding* (Brooks, 1990) with the *symbol grounding problem* (Harnad, 1990) such that semiotic symbols are grounded inherently and meaningfully from the physical interactions of robots with their environment. It will be argued that this approach reduces the symbol grounding problem into a technical problem which will be called the *physical symbol grounding problem* (Vogt, 2000b). To illustrate how a system of semiotic symbols can be constructed from scratch, a concrete robotic experiment will be presented.

A similar argumentation has recently been put forward in AI and cognitive science by Dorffner and colleagues. They apply the theory of semiotics for discussing the symbol grounding problem (Dorffner et al., 1993; Prem, 1995) and to illustrate their ideas they have simulated some aspects in a connectionist model of language acquisition (Dorffner, 1992). The major difference between their work and the current work is that Dorffner's model, besides being a connectionist model, has been tested only in simulations of language acquisition, whereas this work invokes a concrete robotic experiment that investigates language evolution. As the ideas behind their approach are very similar to the one presented below, this work will not be discussed further.

In the next section, the traditional cognitivist approach, the embodied cognitive science approach, and their problems in relation to symbols are shortly reviewed. This review paves the way for introducing an alternative interpretation of symbols. As will be argued, the newly defined semiotic symbols are meaningful within themselves and fit well within the embodied cognition paradigm. From there on, the article will present a model by which agents can acquire a set of semiotic symbols of which the meaning is grounded through the agents' interactions with their environment. The model (explained in section 3) is based on the *language game* model that has been proposed by Luc Steels to study the origins and evolution of language (Steels, 1996b). Section 4 will present the results of a concrete experiment in which robotic agents develop a set of semiotic symbols using the language game model. The results will be discussed in section 5. Finally, section 6 concludes.

2. Symbols

2.1. Symbols as internal representations

Traditionally, symbols are defined as internal representations with which computations can be carried out. As mentioned, this approach is subject to some fundamental problems such as the frame problem and the symbol grounding problem. In this section, a brief review of the classical *cognitivist* approach is given, together with a discussion of the symbol grounding problem. Although the frame problem is closely related, the discussion in this paper concentrates on the symbol grounding problem.

The discussion starts with the *physical symbol system hypothesis* as put forward by Newell and Simon (Newell and Simon, 1976; Newell, 1980). This hypothesis states that physical symbol systems are sufficient and necessary conditions for intelligence. Physical symbol systems (or symbol systems for short) are systems that can store, manipulate and interpret symbolic structures according to some specified rules.

In Newell and Simon's definition, symbols are considered to be patterns that provide distal access to some structure (Newell, 1990). These are internal representations that can be accessed from some external structure. (Where external is relative to the pattern, hence it could be some other pattern.) Although Newell (1990) admits that the relation between a symbol's meaning and the outside world is important, he leaves it unspecified. Other related conceptions of a symbol are also used in the cognitive science community. De Saussure, for instance, defines a sign as

a relation between a meaning and some arbitrary form/label, or more concretely as "a link between ... a concept and a sound pattern" (De Saussure, 1974, p. 66). Harnad defines a symbol as an "arbitrary category name" (Harnad, 1993). All these definitions assume that symbols are internal representations.

These notions of symbols fit well with the mind as a computer metaphor. However, symbols in computers only have a meaning when interpreted by an external observer; computers manipulate symbols without being aware of their meaning. Naturally, this is not the way human cognition works. Humans are very well capable of interpreting the symbols, which they manipulate for instance during thought or while using language; they need no external observer to do this. Therefore, one would like to have symbols that agents can interpret themselves.

This problem led Searle to formulate his famous Chinese Room argument (Searle, 1980), which will not be discussed here. It also led Harnad to formulate his *symbol grounding problem* (Harnad, 1990). As argued, symbolic manipulation should be *about* something and the symbols should *acquire* their meaning from reality. This is what Harnad calls the symbol grounding problem. According to Harnad, symbols should be grounded from the bottom-up by invariantly categorizing sensorimotor signals.

Harnad proposes that this should be done in three stages:

1. **Iconization** Analogue signals need to be transformed to *iconic representation* (or icons).

2. **Discrimination** "[The ability] to judge whether two inputs are the same or different, and, if different, how different they are."

3. **Identification** "[The ability] to be able to assign a unique (usually arbitrary) response – a 'name' – to a class of inputs, treating them all as equivalent or *invariant* in some respect." (Harnad, 1990, my italics)

Iconization and discrimination, according to Harnad yield sub-symbolic representations; symbols are the result of identification. Hence, identification is the goal of symbol grounding. The process of identification is task dependent (Sun, 2000). As will be argued in this paper, using language is a task that is particularly suited to do the identification. This is mainly because language through its conventions offers a basis for invariant labeling of the real world.

In Harnad's work symbols are still defined as names for categories of sensorimotor activity. As such the symbol grounding problem relates to the cognitivist paradigm, which concentrates on internal symbol processing (Ziemke, 1999). As will be argued, symbols could also be viewed as structural couplings between *reality* and sensorimotor activations of an agent that arises from the agent-environment interaction. This reality may be a real world object or some internal state. When symbols are structures that inherently relate reality with internal structures, they are already meaningful in some sense and the symbol grounding problem is not a fundamental problem anymore.

2.2. Symbols or no symbols?

To overcome the problems of the cognitivist approach, embodied cognitive science came around in the late 1980s and gained popularity ever since. The approach has strong roots in artificial intelligence, where it also became popular under the terms *nouvelle AI* and *behavior-based robotics*. Besides in AI, it also has many roots in other disciplines of cognitive science such as psychology (Gibson, 1979), linguistics (Lakoff, 1987), philosophy (Boden, 1996), and neuroscience (Edelman, 1987; Johnson, 1997).

The essence of this modern approach will be discussed here briefly in line with the argumentation brought by Brooks who introduced the *physical grounding hypothesis* (Brooks, 1990; Brooks, 1991). This hypothesis states that intelligence should be grounded in the interaction between a physical agent and its environment. Furthermore, according to this hypothesis, symbolic representations are no longer necessary. Intelligent behavior can be established by parallel operating sensorimotor couplings.

When, as Brooks argues, symbolic representations are no longer necessary, it could be argued that the symbol grounding problem is no longer relevant since there are no symbols (Clancey, 1997; Pfeifer and Scheier, 1999). Another important aspect of the physical grounding hypothesis is that intelligent behaviors are often *emergent* phenomena, e.g., (Pfeifer and Scheier, 1999). This means that intelligent behaviors may arise from mechanisms that appear not to be designed to perform the observed behavior. The physical grounding hypothesis lies at the heart of embodiment and situatedness. Intelligence is embodied through an agent's bodily experiences of its behavior and it is situated through the agent's interaction with the world.

Brooks and others showed that much of an agent's surprisingly intelligent behavior can be explained at the level of *sensorimotor* control, e.g., (Steels and Brooks, 1995; Arkin, 1998; Pfeifer and Scheier, 1999). Very simple mechanisms that connect agents' sensors with their motors can exhibit rather complex behavior without requiring symbolic representations (Braitenberg, 1984).

An example is the famous Cog experiment of Brooks and his colleagues, e.g., (Brooks et al., 1998). Cog is a humanoid robot that can mimic some human-like behaviors by connecting its sensory stimulation to some actuator response. Its behaviors are controlled according to the behavior-based paradigm. Several behaviors are modeled in a layered organization of sensorimotor couplings. These behavior modules are implemented as loosely coupled parallel processes and can be learned. Cog has learned, for instance, to detect human faces and gaze directions, to control hand-eye coordination, to saccade its eyes and to interact socially with humans (Brooks et al., 1998). All these behaviors are based on the physical grounding hypothesis.

Although many behaviors can be explained by the physical grounding hypothesis, the question still remains whether it is able to explain higher cognitive functions. The assumption taken in this paper is that intelligent behaviors such as thought and language do require some form of symbolic representation. The reason for this is twofold: First, as scientists like to describe overt behavior in terms of symbol manipulation, it is very useful to have a proper definition that fits well within the embodied cognition paradigm. This makes it easier to ascribe symbols to embodied cognitive agents from an observer's perspective. The second reason is that in order to facilitate higher cognitive functions such as language, agents might actually need symbols that they can manipulate. The question remains how are symbols represented?

2.3. Symbols as structural couplings

As already discussed, the traditional approach to cognitive science and AI is confronted with problems such as the *frame problem* (McCarthy and Hayes, 1969; Pylyshyn, 1987), the *symbol grounding problem* (Harnad, 1990) and the *Chinese Room* 'problem' (Searle, 1980). At the heart of these problems lies the fact that the traditional symbols are neither situated nor embodied, see, e.g., (Clancey, 1997; Pfeifer and Scheier, 1999) for broad discussions on these problems. As mentioned, the physical grounding hypothesis (Brooks, 1990) doubts the necessity of symbolic representations. But if they would be necessary they should be both situated and embodied (Clancey, 1997). In this section a definition of symbols will be given that is both situated and embodied from an agent's point of view.

2.3.1. Semiotic symbols

Various scientists from the embodied cognitive science field assume that when symbols should be necessary to describe cognition, they should be defined as structural couplings connecting objects to their categories based on their sensorimotor projections (Clancey, 1997; Maturana and Varela, 1992). There is, however, already a definition of a symbol that comes very close to such a structural coupling. This alternative definition stems from the work of Peirce (Peirce, 1931-1958).

Peirce's theory extends the semiotics of De Saussure. While De Saussure defines a sign as having a meaning (the *signified*) and a form (the *signifier*) (De Saussure, 1974), Peirce also includes its relation to a referent. A sign consists of what he calls a *representamen*, *interpretant* and

an *object*. These are defined as follows (Chandler, 1994):

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

Interpretant: ... the sense made of the sign.

Object: to which the sign refers.

According to Peirce, the sign is called a *symbol* if the representamen in relation to its interpretant is either arbitrary or conventionalized, so that the relationship must be learned. In this respect the representamen could be, for instance, a word-form as used in language. The interpretant could be viewed "as another representation which is referred to the same object" (Eco, 1976, p. 68). The object can be viewed as a physical object in the real world, but may also be an abstraction, an internal state or another sign. In the experiments described below, the objects will be physical objects and the interpretant will be represented by a category that is formed from the visual interaction of a robot with the real world.

Often the term symbol is used to denote the representamen, e.g., (Odgen and Richards, 1923; Harnad, 1993). Many scientists, *including Peirce*, tend to 'misuse' the term sign when referring to the representamen. However, the sign was originally defined by Peirce as the *triadic relation* (Chandler, 1994). In this paper *the triadic interpretation of the sign is adopted as the definition of the symbol*, provided that the representamen of the sign is either arbitrary or conventionalized. In order to distinguish this definition from the traditional interpretation, this alternative interpretation of the symbol shall be called a *semiotic symbol*.

Also a more familiar terminology is adopted. Following Steels, the representamen is called a *form*, the interpretant a *meaning* and the object a *referent* (Steels and Kaplan, 1999).

The sign (or semiotic symbol) is often illustrated as a *semiotic triangle* such as the one introduced by (Odgen and Richards, 1923). The triangle displayed here (figure 1) only differs from the original one in its terminology[3]. The dotted

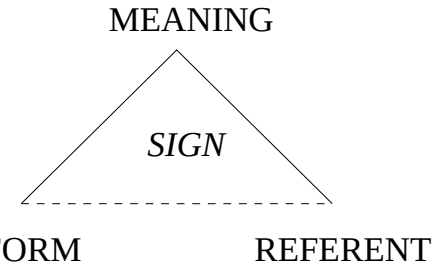


Figure 1. The semiotic triangle illustrates the relations that constitute a sign. When the form is either arbitrary or conventionalized, the sign can be interpreted as a symbol.

line in the figure indicates that the relation between form and referent is not always explicitly observable.

2.3.2. The meaning of meaning

The term *meaning* requires special attention. It has been (and still is) subject to much debate in philosophy and the cognitive sciences in general. This work tries to distill a definition of meaning that is suitable within the context of the current investigation.

According to Peirce, a semiotic symbol's meaning arises in its interpretation (Chandler, 1994). As such the meaning arises from the process of semiosis, which is the interaction between form, meaning and referent. This means that the meaning depends on how the semiotic symbol is constructed and with what function. This is comparable to the notion of meaning that is prominent in embodied cognitive science, where meaning depends on "the way we perceive the overall shape of things ... and by the way we interact with things with our bodies" (Lakoff, 1987, p. 292).

So, the meaning of semiotic symbols can be viewed as a functional relation between a form and a referent. This relation is based on an agent's bodily experience and interaction with a referent. The experience of an agent is based on its history of interactions. Each interaction be-

[3]In Odgen and Richards' original diagram, the term *symbol* was actually used instead of *form*. In addition, they call the meaning a *thought or reference*.

tween an agent and a referent can activate its past experiences bringing forth a new experience. The way these bodily experiences are represented and memorized form the internal representation of the meaning. The actual interaction between an agent and a referent 'defines' the functional relation.

In the experiment described below, robots develop a system of semiotic symbols through communicative interactions called language games. The robots have a very simple body and can only visually interact with objects and, in principle, point at them by orienting towards the objects. In the language games, communication is only used to name a visually detected referent (in case of the speaker) and to guess what visually detected referent the speaker names (in case of the hearer). The function (or *use* in Wittgenstein's terminology) of this naming is only to focus the hearer's attention towards a referent such that it, for instance, can go to or point at the referent. The meaning of the semiotic symbol in such a *guessing game* is conveyed by an agent's perception, categorization and naming of the referent, together with, in case of the hearer, an appropriate reaction. Such meanings may change dynamically over time as the robots (visually) interact more with the referents.

This use of meaning may not seem realistic, because the communication of semiotic symbols have no meaning with respect to the robots' survival as, for instance, pointed out by (Ziemke and Sharkey, 2001). But it is very much similar to the way infants seem to construct meanings upon their first visual encounter with some object. This is nicely illustrated by a series of experiments reported in (Tomasello and Barton, 1994). In these experiments, infants are shown a, for them, novel toy together with an invented name (i.e., a non-existing word), such as "toma". After that, the toy is hidden and the infants are requested to find the "toma". Even if they do not know where the toy is hidden, the infants are very successful in recognizing the "toma" when they find it. So, although the children are not yet able to grasp the entire meaning of the toy (they do not know, for instance, what they can or should do with it), they presumably do form some meaning for the object. This meaning is solely based on the infants' visual interaction with, and categorization of the toy. Naturally, when the infants further interact with the object, e.g., by playing with it, they expand their meaning of the object and they come to learn more of its functions.

As Ziemke and Sharkey do, one might still argue that robots cannot use semiotic symbols meaningfully, since they are not rooted in the robot as the robots are designed rather than shaped through evolution or physical growth (Ziemke and Sharkey, 2001; Ziemke, 1999). In that respect, they continue, today's robots do not use semiotic symbols meaningfully, since whatever task they might have stems from its designer or is in the head of a human observer. With this in mind, it will be assumed that robots, once they can construct semiotic symbols, they do so meaningfully. This assumption is made to illustrate how robots can construct semiotic symbols meaningfully. Thus, constructing semiotic symbols is considered as a necessary condition for meaningful symbol use, but no strong claims are being made on whether this is a sufficient condition for meaningful symbol use.

2.3.3. Physical symbol grounding

Adopting Peirce's triadic notion of a symbol has at least two advantages. One advantage is that one may argue that the semiotic symbol is *per definition* grounded, because the triadic relation (i.e., the semiotic symbol) already bears the symbol's meaning with respect to reality. As meaning is defined by the functional relation between the form, meaning and referent, one might argue that the use of such semiotic symbolic structures "are meaningful to begin with" (Lakoff, 1987, p. 372)[4]. Hence the symbol grounding problem is no longer relevant. Not because symbols do not exist anymore as argued by, e.g., Brooks, but because the semiotic symbols are already meaningful. This does not mean, however, that there is no problem anymore, the symbol grounding problem is no longer a *fundamental* problem of interpreting symbols meaning-

[4]Note that Lakoff does not explicitly apply this quote to semiotic symbols, but his argument is similar to the one expressed here.

fully. The problem is reduced to the process of semiosis, which is defined by Peirce as the interaction between the referent, meaning and form. The semiosis can be viewed as the process of constructing the semiotic triangle. Implementing this in autonomous systems, however, remains a hard problem. This problem shall be addressed as the *physical symbol grounding problem* (Vogt, 2000b) and will be treated as a technical problem.

The physical symbol grounding problem is a combination of the physical grounding problem and the symbol grounding problem. It is based on the idea that symbols should be grounded, cf. (Harnad, 1990) and the idea that they should be grounded by physical agents that interact with the real word, cf. (Brooks, 1990). To solve the physical symbol grounding problem, the three phases of symbol grounding identified by Harnad (1990) are still relevant. The next section will show how these phases (iconization, discrimination and identification) can be modeled in a concrete experiment.

The second advantage is that the semiotic symbol is situated and embodied. It should be acquired through some interaction between a physical agent and its environment. This allows to connect semiotic symbols with embodied cognition.

So constructing semiotic symbols is required for communication. Every time a semiotic symbol is used, it is (re-)constructed by its user. As a result, the semiotic symbols are not static, but may change dynamically whenever the agent-environment interaction requires so. In communication the form needs to be conventionalized (although still arbitrary to some extend). Establishing conventions about a semiotic symbol's form in relation to its referent is particularly useful for the invariant identification of the referent. As a referent is perceived differently under varying circumstances, it may be categorized differently. As will be shown, language use helps to identify the sensing of the referents invariantly. Therefore, it is assumed that meaning co-evolves with the language (Steels, 1997a), such that meaning can also be viewed as a cultural unit cf. (Eco, 1976). Similar arguments have been put forward to explain various aspects of language development, see, e.g., (Whorf, 1956; Lakoff, 1987).

2.3.4. Summary

In this subsection a definition of a semiotic symbol has been adopted that provides an alternative to the traditional definitions. As Clancey has argued, a symbol within the embodied cognition paradigm should be some structural coupling (Clancey, 1997). The semiotic definition provided by Peirce yields a structural coupling between the real world object and some internal representation (or a sensorimotor activation pattern), especially in the process of semiosis.

Since the semiotic symbol is defined by a relation between a form, meaning and referent, its meaning is an intrinsic property bearing the relation to the real world. Hence, it could be argued that the semiotic symbol is per definition grounded and the symbol grounding problem is not relevant anymore. However, there still remains the problem of constructing a semiotic symbol. Rather than a fundamental problem, this will be treated as a hard technical problem and it is addressed as the *physical symbol grounding problem.* The semiotic symbols acquire their meaning through perception and categorization as a process that conveys the semiotic symbols' names to the referents they stand for.

As argued, and as the experimental results will reveal, semiotic symbols are constructed most efficiently in communication. Language development gives rise to the development of meaning and vice versa.

3. Adaptive language games

3.1. Synthetic modeling of language evolution

From the second half of the 1990s, research at the Vrije Universiteit Brussel and at the Sony Computer Science Laboratory in Paris is focussed on the study of language origins and evolution. These studies are based on the *language game* model as proposed by Steels (1996b). In this model language *use* is the central issue in language development as has been hypothesized by the 'father of language games' Wittgenstein (1958). Steels hypothesized three mechanisms for

language development: *cultural interaction*, *individual adaptation* and *self-organization.*

In this model, agents have a mechanism to exchange parts of their vocabulary with each other, called cultural interaction. When novel situations occur in a language game, agents can expand their lexicons. In addition, they evaluate each 'speech act' on their effectiveness which they use to strengthen or weaken used form-meaning associations. These mechanisms are called individual adaptation. Although agents have been implemented with mechanisms that model local behavior, such as communicating to another agent and individual adaptations, the iterative combination of these mechanisms yields the emergence of a global coherence in the agents' language. This process is very similar to the self-organizing phenomena that have been observed in many biological, physical and cultural systems (Prigogine and Strengers, 1984; Maynard-Smith and Szathmáry, 1995; Maturana and Varela, 1992). Self-organization is thought to be a basic mechanism of complex dynamical systems.

The above mechanisms for lexicon development have been tested extensively under different settings by the researchers in Brussels and Paris to investigate various aspects of lexicon origins, evolution and development. These aspects include *lexicon formation* (Steels, 1996b), *lexicon dynamics* (Steels and McIntyre, 1999), *multiple word games* (Van Looveren, 1999) and *stochasticity* (Steels and Kaplan, 1998). All these experiments reveal that the language game model (also called the *naming game* in relation to lexicon formation) is a strong model to explain lexicon formation in terms of the nurture paradigm. The model is very similar to those that have been successfully studied by, e.g., (Batali, 1998; Oliphant, 1999; Kirby and Hurford, 1997). It contrasts the nativist approach advocated by, e.g., (Chomsky, 1980; Pinker and Bloom, 1990; Bickerton, 1998).

The above mechanisms to explore lexicon formation have been extended to include meaning creation (hence modeling symbol grounding). The extention resulted in a model of *discrimination games* (Steels, 1996c). The mechanisms on which the discrimination games are based are very similar to those of the language games: *agent-environment interaction*, *individual adaptation* and *self-organization.* Agents are given a mechanism to perceive their environment. Another mechanism allows them to adapt their memories. They can construct new categories that form the basis of the meaning, and they can adapt them based on their use in grounded language games. The evolution of meaning using these mechanisms is an emergent property of the simple interaction and adaptations, so the term self-organization is in place. The discrimination game model has been studied separately in simulations (Steels, 1996c) and on mobile robots (Vogt, 1998b; De Jong and Vogt, 1998).

Other experiments have been done in which the discrimination game has been coupled to the naming game in simulations (De Jong, 2000; Belpaeme, 2001), on immobile robots called the 'Talking Heads' (Belpaeme et al., 1998; Steels and Kaplan, 1999) and on mobile robots (Steels and Vogt, 1997; Vogt, 2000b). Again, and as will be shown in the next section, these experiments revealed that the three mechanisms are very powerful to explain how lexicons can evolve in coevolution with the meaning without implementing any prior knowledge of the lexicon and meaning.

Another variant of the language game that is worthwhile mentioning is the *imitation game* that has been used to study the origins of human-like vowel systems (De Boer, 1997, 2000a). De Boer showed in his experiments, that agents can develop repertoires of vowels that are very similar to vowel systems observed in human languages. These systems emerged based on the three mechanisms proposed by Steels as mentioned above. A crucial factor in the experiments' success is the realistic simulation of the vocal tract and auditory system with which the agents are modeled. It is important to realize that in the experiments none of the vowels have been preprogrammed, neither are their features as the Chomsky and Halle theory proposes (Chomsky and Halle, 1968).

For an overview of the experiments that are done in Brussels and Paris, consult (Steels, 1997b).

3.2. The experiment

The development of the language games has resulted in a model that has been implemented on various robotic platforms: on mobile LEGO robots (Steels and Vogt, 1997), on the Talking Heads, which are immobile pan-tilt cameras (Belpaeme et al., 1998) and most recently on the AIBO, which is a four legged robot (Kaplan, 2000). The experiment reported here is based on the one reported in (Steels and Vogt, 1997). The goal of this experiment is that two mobile robots, given their bodies and interaction/learning mechanisms, 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 scratch with which the robots can successfully communicate the names of the objects.

Although not always modeling lexicon evolution, lexicon grounding on real robots is becoming increasingly more popular. Other research, however, does not investigate lexicon development from scratch, but assumes a part of the lexicon is given. Examples are the work of (Billard and Hayes, 1997; Yanco and Stein, 1993) in robot-robot communication, or (Roy, 2000; Sugita and Tani, 2000) in human-robot communication. It is beyond the scope of this paper to discuss these researches here, but all this work solves, to some extend, the physical symbol grounding problem.

In the experiments, the robots play a series of *guessing games* (Steels and Kaplan, 1999). The guessing game is a variant of a language game in which the hearer tries to guess what referent the speaker names. Much of the processing of the robots that is not directly involved with lexicon development, like sensing, turn-taking and evaluating the feedback has been preprogrammed in a behavior-based manner, see (Vogt, 2000b) for details. The same holds for the learning mechanisms. The reason for this is not to complicate the system too much, so that a working experiment could be developed to investigate how robots can develop a grounded lexicon, given the mechanisms explained below. Other research that investigates various aspects of the origins of language and communication tries to explain how such mechanisms may have evolved. Examples of such research investigate the origins of communication channels (Quinn, 2001), feature detectors (Belpaeme, 1999) and communication as such (Noble, 2000; De Jong, 2000).

The basic scenario of a guessing game is illustrated in what Steels calls the semiotic square (Steels and Kaplan, 1999), see figure 2. The 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. This way the robots acquire a context setting. The speaker selects the sensing of one referent as the topic; the hearer considers all detected referents as a potential topic. The robots categorize the preprocessed data, which results in a meaning if the categorization is used in the communication, i.e., when it is coupled to a form. After categorization, the speaker produces an utterance and the hearer tries to interpret this utterance. When the meaning of this utterance applies to the categorization of the sensed referents, the hearer can act to the appropriate referent, for instance, by pointing at it. The guessing game is successful if the hearer guessed the right topic from the speaker's utterance. The success is evaluated in the feedback, which is passed back to both robots so that they can adapt their ontology of categories (used as memorized representations of the meanings) and lexicon.

As can be seen in figure 2, when the segment-node is ignored (although it is a necessary intermediate step), the guessing game allows the robots to construct a semiotic symbol. The game is thus similar to the process of semiosis. In this paper it is assumed that meaning arises from the sensing, segmentation and categorization. Hence semiotic symbols are illustrated with the triangle (figure 1) rather than the square. When a guessing game is successful, the relations between form, meaning and referent are established properly and one could argue that the constructed semiotic symbol is used meaningful. Below follows a detailed description of the experimental setup and the guessing game model. Each basic part of the model is exemplified with an illustra-

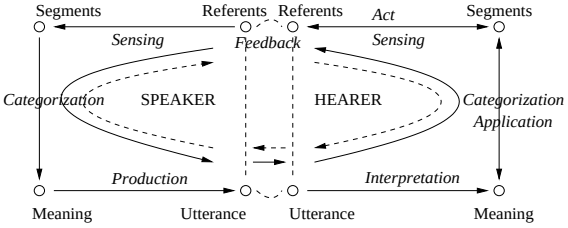


Figure 2. The semiotic square illustrates the guessing game scenario. The two squares show the processes of the two participating robots. This figure is adapted from (Steels and Kaplan, 1999).

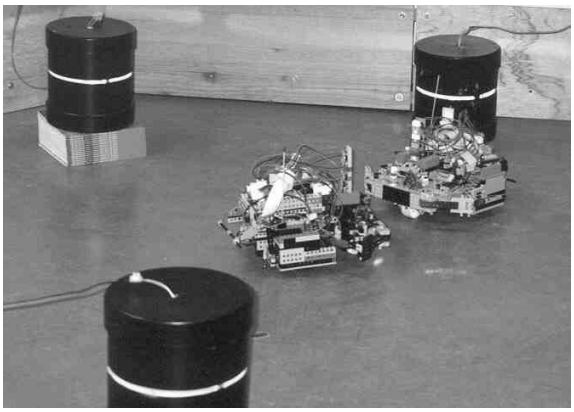


Figure 3. The LEGO robots in their environment.

tive example.

3.3. The experimental setup

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

The robots are situated in a small environment ($2.5 \times 2.5 m^2$) 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. Although this is not a realistic task for living organisms (or even for robots), complicating the task would not contribute much to the investigation of how robots can develop a lexicon to name things.

The reasons for designing the correspondence between light sources and sensors are twofold. First, it helps the observer in analyzing the experiments. It provides the observer with an easy tool to investigate what light source the robots saw based on the sensory data. Secondly, if the robots were to live in this environment and had to distinguish the light sources from each other in order to survive, evolution might have come up with a similar visual apparatus.

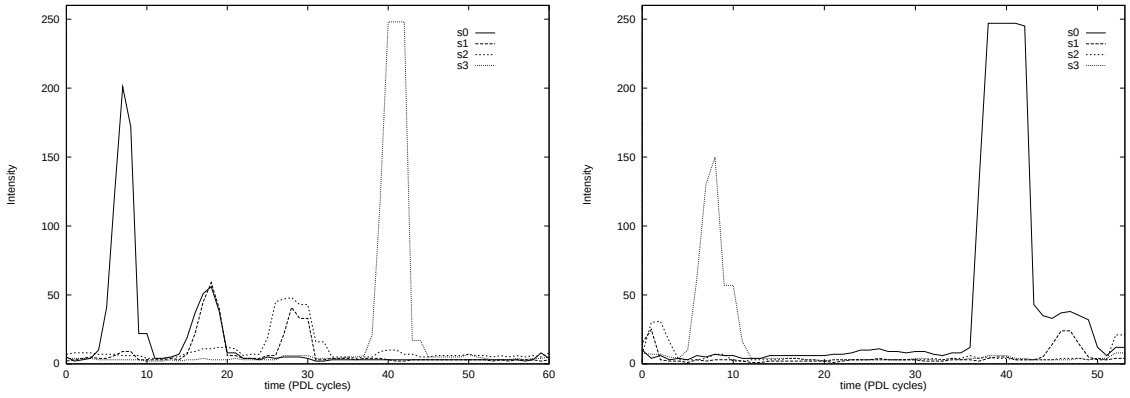
3.4. Sensing, segmentation and feature extraction

As a first step towards solving the symbol grounding problem, the robots have to construct what Harnad calls an iconic representation. Although Harnad refers to raw sensory data such as an image on the retina, 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. Below these three stages are explained in more detail.

3.4.1. 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



(a) r0 (b) r1

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 stand at different positions. 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 this sensing event took $1.5s$ for robot r0 and slightly less for robot r1. The robots have different rotation periods due to the noisy control of the robots.

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 of the four light sensors, see figure 4. 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, as an example, the sensing of two robots during a language game. Figure 4 (a) shows that robot r0 clearly detected the four light sources; there appears a peak for every light source. In each peak, the sensor that has the same height as the light source responsible for the peak (i.e., the corresponding sensor) has the highest intensity. The two robots do not sense the same view as can be seen in figure 4 (b). This is due to the fact that the robots are not located at the same place.

3.4.2. 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, between two peaks the sensory channels are noisy. The first step in the segmentation preprocesses the raw sensory data to remove this noise. The preprocessed sensory channel data $\tau_{i,t}$, for sensory channels $i = 0, \ldots, n$ on time steps t, is acquired by subtracting an upper noise level Θ_i from the raw sensory data $x_{i,t}$. This is expressed in the following equation:

$$\tau_{i,t} = \begin{cases} x_{i,t} - \Theta_i & \text{if } x_{i,t} - \Theta_i \geq 0 \\ 0 & \text{if } x_{i,t} - \Theta_i < 0 \end{cases} \quad (1)$$

The upper noise levels Θ_i for the different sensory channels have been obtained empirically. Applying the above equation to the middle part of the scene from figure 4 (a) results in the scene displayed in figure 5.

In this figure there are two connected regions that have positive sensory values for at least one

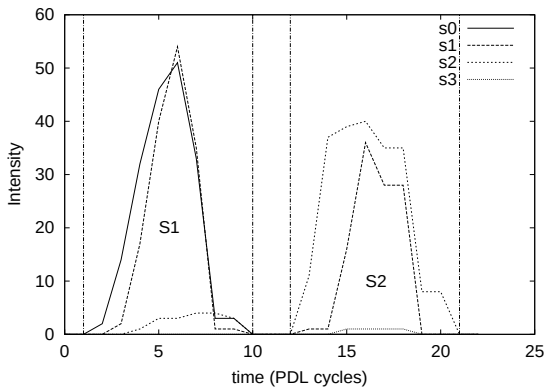


Figure 5. The preprocessed sensory channel data of a part of the sensed view from figure 4 (a). The vertical lines mark the boundaries of segments S_1 and S_2.

sensory channel. These regions, of which the boundaries are marked with vertical lines, relate to the sensing of a light source and form the segments. Note that there is no noise anymore between the segments.

The segmentation results in a set of segments $\{S_k\}$, where

$$S_k = \{\mathbf{s}_{k,0}, \ldots, \mathbf{s}_{k,n-1}\} \tag{2}$$

in which k is the number of the segment, $\mathbf{s}_{k,i} = (\tau_{k,i,0}, \ldots, \tau_{k,i,m})$ is the preprocessed sensory channel data, m is the length of the segment and n is the number of sensory channels (4 in this case). The following holds for each segment S_k: for all $j = 0, \ldots, m$ there exist at least one sensory channel i for which $\tau_{k,i,j} > 0$. This means for every segment that for every observation j there is at least one preprocessed sensory channel with a positive value. Applying this to the data shown in figure 5 yields the segments given in table 1.

It is assumed that each segment relates to the detection of one light source. Due to the noisy control and sensing of the robots, this is not always the case. The set of segments constitute what is called the *context* of the guessing game, i.e., $\text{Cxt} = \{S_1, \ldots, S_N\}$, 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.

3.4.3. Feature extraction

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:

$$\varphi(\mathbf{s}_{k,i}) = \frac{\max_{\mathbf{s}_{k,i}}(\tau_{k,i,j})}{\max_{S_k}(\max_{\mathbf{s}_{k,i}}(\tau_{k,i,j}))} \tag{3}$$

The result of applying a feature extraction to the data of sensory channel i will be called a feature $f_{k,i}$, so $f_{k,i} = \varphi(\mathbf{s}_{k,i})$. Segment S_k can now be related to a feature vector $\mathbf{f}_k = (f_{k,0}, \ldots, f_{k,n-1})$, 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.

Applying this feature extraction to the segments from table 1 goes as follows. First the maximum values for each sensory channel (the numerator in eq. 3) have to be identified. In segment S_1 are: 51, 54, 4 and 0 for sensory channels $\mathbf{s}_{1,0}, \ldots, \mathbf{s}_{1,3}$ resp.. The maximum of these maxima (the denominator in eq. 3) is in sensory channel $\mathbf{s}_{1,1}$. The features are extracted by normalizing each maximum value by the maximum value of $\mathbf{s}_{1,1}$. The resulting feature vector for segment S_1 is thus $\mathbf{f}_1 = (0.94, 1.00, 0.06, 0.00)$.

Similarly, segment S_2 has maximum values of 0, 36, 40 and 1 for sensory channels 0 to 3. Normalizing each value to 40 yields feature vector $\mathbf{f}_2 = (0.00, 0.90, 1.00, 0.03)$. The context of this robots could be described in terms of the feature vectors as $\text{Cxt} = \{\mathbf{f}_1, \mathbf{f}_2\}$.

The fact that the sensor that reads the highest intensity during the sensing of a light source is mounted at the same height as the light is an invariant property of the sensing. The feature extraction from eq. 3 has been designed to extract this property. In a feature vector there is one feature with value 1, the others have lower val-

Table 1
The segments S_1 and S_2 from figure 5. The preprocessed sensory channel data $\mathbf{s}_{k,i}$ is given in the columns. Note that for every segment each row has at least one positive value. Further note that here both segments have equal length. In general this is not the case.

	S_1				S_2			
	$\mathbf{s}_{1,0}$	$\mathbf{s}_{1,1}$	$\mathbf{s}_{1,2}$	$\mathbf{s}_{1,3}$	$\mathbf{s}_{2,0}$	$\mathbf{s}_{2,1}$	$\mathbf{s}_{2,2}$	$\mathbf{s}_{2,3}$
$\tau_{k,i,0}$	2	0	0	0	0	1	11	0
$\tau_{k,i,1}$	14	2	0	0	0	1	37	0
$\tau_{k,i,2}$	32	17	1	0	0	16	39	1
$\tau_{k,i,3}$	46	40	3	0	0	36	40	1
$\tau_{k,i,4}$	51	54	3	0	0	28	35	1
$\tau_{k,i,5}$	33	35	4	0	0	28	35	1
$\tau_{k,i,6}$	3	1	4	0	0	0	8	0
$\tau_{k,i,7}$	3	1	3	0	0	0	8	0

ues. This feature indicates to which light source it corresponds. Because in the example feature $f_{1,1}$ of feature vector $\mathbf{f}_1$ has a value 1, one can infer that this feature vector corresponds to the light source that is at the same height as sensor s1[5]. Similarly, $\mathbf{f}_2$ corresponds to the light source at the same height as sensor 2.

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. It is easier to implement categorization when its input has a fixed format. As the segments vary in length, they have no fixed format. 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. It should be noted that such feature extraction functions could well have been learned or evolved as, for instance, shown by (De Jong and Steels, 1999; Belpaeme, 1999).

3.5. Meaning formation

In order to form a semiotic symbol, the robots have to categorize the sensing of a referent so that it is distinctive from the categories relating to the other referents. This way the category can be used as the memorized representation of the meaning of this referent. Although a category should not be equaled with a meaning, it is labeled as such when used in communication, because it forms the memorized representation of a semiotic symbol's meaning.

The 'meaning' formation is modeled with the so-called *discrimination game* (Steels, 1996c). The discrimination game models the discrimination phase in symbol grounding as proposed by (Harnad, 1990) and it searches for distinctive categories in three steps: 1) The feature vectors from the context are categorized. 2) Categories that distinguish a topic from the other segments in the context are identified. And 3) the ontology of categories are adapted.

Each robot plays a discrimination game for the (potential) *topic(s)* individually. 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*.

3.5.1. Categorization

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 $\mathbf{c} = (x_0, \ldots, x_{n-1})$, i.e. a point in the n dimensional feature space $\mathcal{F}$, and ν, ρ and

[5]Note that the sensors are numbered from 0 to 3 and that sensor s0 is the lowest in height.

κ are scores. The category is the region in $\mathcal{F}$ in which all 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). This algorithm returns the category of which the prototype has the smallest Eucledian distance to $\mathbf{f}$. Each robot categorizes all segments this way.

Consider the context of feature vectors from the example derived in the previous section. Further suppose that the robot has two categories in its ontology c_1 and c_2, which are represented by the prototypes $\mathbf{c}_1 = (0.50, 0.95, 0.30, 0.00)$ and $\mathbf{c}_2 = (0.50, 0.95, 1.00, 0.00)$. Then the Eucledian distance between $\mathbf{f}_1$ and $\mathbf{c}_1$ is 0.50, and between $\mathbf{f}_1$ and $\mathbf{c}_2$ this is 1.04. Since the distance between $\mathbf{f}_1$ and $\mathbf{c}_1$ is smallest, $\mathbf{f}_1$ is categorized with $\mathbf{c}_1$. Likewise, $\mathbf{f}_2$ is categorized with $\mathbf{c}_2$.

In order to allow generalization and specialization in the categories, different versions of the feature space are available to a robot. These different feature spaces, indicated by $\mathcal{F}_\lambda$, allow different levels of generality specified by $\lambda = 0, \ldots, \lambda_{\max}$, where $\lambda_{\max}$ is the level in which maximum specialization can be reached. This maximum is introduced because the sensors have a limited resolution and to prevent unnecessary specialization, which makes the discrimination game computationally very inefficient. 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 higher λ is, the more dense the distribution of categories in feature space $\mathcal{F}_\lambda$ can be and the less general the categories in that feature space are. The categories of all feature spaces together constitute an agent's ontology.

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 $m \leq \lambda_{\max}$. So, assuming only one feature space, the two feature vectors from the example are categorized with the following sets of categories: $C_1 = \{c_1\}$ and $C_2 = \{c_2\}$.

In the discrimination games, a prototypical representation has been used rather than the binary representation used in (Steels, 1996c; Steels and Vogt, 1997) or the subspace representation used by De Jong (De Jong and Vogt, 1998), which all yield more or less similar results (De Jong and Vogt, 1998; Vogt, 2000b). The reason for adopting a prototype representation is its biological plausibility as inferred by for instance psychologists (Rosch et al., 1976) and physiologists (Churchland, 1989).

3.5.2. Discrimination

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

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

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. I.e., those categories that distinguish the topic from all other segments in the context.

3.5.3. Adaptation

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 (see also the example in section 3.5.4):

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 $\kappa = 1 - \frac{\lambda}{\lambda_{\max}}$.

The score ν indicates the effectiveness of a category in discrimination game, 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). Score κ is a constant, based on the feature space $\mathcal{F}_\lambda$ and the feature space that has the highest resolution possible (i.e., $\lambda = \lambda_{\max}$). This score implements a bias towards selecting the most general category. The other scores are updated as in reinforcement learning. It is beyond the scope of this paper to give exact details of the update functions, see (Vogt, 2000b) for these details.

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 during the construction of new prototypes, rather than the complete feature vector is to speed up the construction.

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:

$$\mathbf{c} := \mathbf{c} + \epsilon \cdot (\mathbf{f} - \mathbf{c}) \tag{4}$$

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. This update is similar to on-line k means clustering (MacQueen, 1967).

3.5.4. An example

To continue the example from above, recall that the two feature vectors in the context are categorized with the following sets of categories: $C_1 = \{c_1\}$ and $C_2 = \{c_2\}$. Suppose that S_1 is the topic of the discrimination, the distinctive category set is $DC = \{c_1\}$, because $c_1 \notin C_2$. Because $DC \neq \emptyset$, the discrimination game is a success. If c_1 is further successfully used in naming the referent, its prototype shifts towards $\mathbf{f}_1$ by applying equation (4). If $\epsilon = 0.1$, which is the case in the experiment, then $\mathbf{c}_1$ becomes $\mathbf{c}'_1 = (0.54, 0.96, 0.28, 0.00)$.

Now suppose that the context did not consists of two feature vectors, but three: Cxt $= \{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3\}$, where $\mathbf{f}_1$ and $\mathbf{f}_2$ are as before and $\mathbf{f}_3 = (1.00, 0.90, 0.05, 0.00)$. Further suppose that the robot has the same ontology as before, then $\mathbf{f}_3$ is categorized with the category set $C_3 = \{c_1\}$, since c_1 is the closest to $\mathbf{f}_3$. When S_1 is the topic, the distinctive category set is empty, since $C_1 = \{c_1\}$ and $c_1 \in C_3$. The discrimination game fails and two new categories are formed of which the prototypes are as follows when feature $f_{1,0}$ is selected as an exemplar of topic S_1 (note that only one feature space $\mathcal{F}$ is considered in this example): $\mathbf{c}_3 = (1.00, 0.95, 0.30, 0.00)$ and $\mathbf{c}_4 = (1.00, 0.95, 1.00, 0.00)$. Feature $f_{1,0}$ is copied to the new prototypes, together with the other features of already existing prototypes, compare $\mathbf{c}_3$ and $\mathbf{c}_4$ with $\mathbf{c}_1$ and $\mathbf{c}_2$.

3.6. Naming

After both robots have obtained distinctive categories of the (potential) topic(s) from the discrimination game as explained above, the *naming game* (Steels, 1996b) 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.

3.6.1. The lexicon

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 (memorized part of a) meaning represented by a category, and σ is the association score that indicates the effectiveness of the lexical entry in the language use.

3.6.2. Production

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

If it fails to do so, the speaker will first consider the next distinctive category from the 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.

3.6.3. Interpretation

On receipt of the utterance, the hearer searches its lexicon for entries for which the form matches the utterance, and for which 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 relates to this 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. A topic relates to an entry when its distinctive category matches the entry's meaning.

3.6.4. Feedback

In the feedback, the outcome of the guessing game is evaluated. It is important to note that in this paper, the term feedback is only used to indicate the process of evaluating a game's success by verifying whether both robots communicated about the same referent. As mentioned, the guessing game is successful when both robots communicated about the same referent. The feedback is established by comparing the feature vectors of the two robots relating to the topics. If these feature vectors correspond to each other, i.e., they both have a value of 1 in the same dimension, the robots have identified the same topic, cf. the invariance criterion mentioned in section 3.4.3. The outcome of the feedback is known to both robots.

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

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.

The feedback is evaluated rather artificial and not very realistic since robots normally have no access to each other's internal states. However, previous attempts to implement feedback physically have failed (Vogt, 1998a). In these attempts the hearer pointed at the topic so that the speaker could verify whether the hearer identified the same topic. This attempt, however, failed because the speaker was not able to verify reliably at what object the hearer pointed. This pointing strategy has been successfully implemented in the Talking Heads experiment (Steels and Kaplan, 1999). To overcome this technical problem, it is assumed that the robots can do this. Naturally, this problem needs to be solved in the future.

3.6.5. 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 of $P_s = 0.1$.

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 randomly selected segment from its context.

Table 2
The lexicon of the speaker and hearer used in the example. Each agent has associated three meanings (in the rows) with two forms (in the columns). Real-valued numbers indicate the association scores and a dash indicates there is no association.

Speaker			Hearer		
M_s	**tyfo**	**labo**	M_h	**tyfo**	**labo**
c_1	0.20	0.25	c'_1	0.10	-
c_2	-	0.65	c'_2	-	0.75
c_3	0.95	-	c'_3	0.70	0.05

3. There was a mismatch in referent: Both robots lower 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 (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[6].

3.6.6. An example

The following example illustrates the naming phase of the guessing game. Suppose the speaker and the hearer have a lexicon as given in table 2. Each agent has three private meanings $c_i(c'_i)$ associated with two different forms **tyfo** and **labo**. In the cells of the table the association scores are given, a dash indicates that there is no association between that meaning and form. Further suppose that both robots each have detected three light sources L1, L2 and L3.

Suppose that the speaker selected the segment relating to L3 as the topic, which it categorized distinctively with c_3. This category is only associated with the form **tyfo**, which it utters. Upon receiving the form **tyfo**, the hearer tries to interpret it. The hearer has two meanings associated with **tyfo**: c'_1 and c'_3. As the association score relating **tyfo** with c'_3 is highest, the hearer selects the segment belonging to c'_3 as the topic[7]. If this segment relates to L3, the guessing game is successful and the scores are updated as follows: The speaker and the hearer both increase the association score between c_3 (or c'_3) with **tyfo**, which become 0.955 and 0.73 respectively (recall that $\eta = 0.9$). Competing associations are laterally inhibited: the association score of $\langle c_1, \textbf{tyfo} \rangle$ becomes 0.18, of $\langle c'_1, \textbf{tyfo} \rangle$ becomes 0.09 and of $\langle c'_3, \textbf{labo} \rangle$ becomes 0.045.

Reconsider the lexicon in table 2, but now the speaker selected L1 that it categorized with c_1 as the topic. In this case, the speaker will select **labo** to name L1, because this has the highest association score. The hearer interprets **labo** with c'_2, which unfortunately has been categorized for L2. There is a mismatch in referent. The association scores of the used association scores are lowered. So, the association score of $\langle c_1, \textbf{labo} \rangle$ becomes 0.225 and of $\langle c'_2, \textbf{labo} \rangle$ becomes 0.675.

If the speaker has categorized the topic, e.g., the segment relating to L1, with a category that is not yet in its lexicon, say c_4, then it may invent a new form. This new form, for instance, **gufi** is then uttered to the hearer. The hearer, however, does not know the form yet and associates it with the categorization(s) of a randomly selected segment and the game is considered as a failure.

3.7. Summary

This section presented the guessing game model by which the experiments are done. Two mobile robots try to construct a lexicon with which they can solve the physical symbol grounding problem. In each guessing game, the robots try to name one of the light sensors that are in their surroundings. They do so by taking the following steps:

1. Sensing, segmentation and feature extraction.

[6] Independent of each other, Oliphant (1999), De Jong (2000) and Kaplan (2001) have shown that lateral inhibition is crucial for convergence in lexicon development.

[7] Note that the meaning score μ is discarded to simplify the example.

2. Topic selection.

3. Discrimination games: (a) categorization, (b) discrimination and (c) ontology adaptation.

4. Naming: (a) production, (b) interpretation, (c) evaluating feedback and (d) lexicon adaptation.

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

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 physical 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. Identification is done at the naming level and it is successful when the guessing game is successful. It is important to realize that internal representations stored in the robots' memories as such are not semiotic symbols; they only constitute part of a semiotic symbol *when* used in a language game. Only then the relation with a referent is assured.

4. The experiments

Using the defined model (and some variations of it), a series of experiments have been done to investigate various aspects of the model. These experiments are all reported in (Vogt, 2000b). One of them is reported in this section.

This section is organized as follows: Before reporting the experiment some expectations of the experiment's success are stated according to some statistics calculated from the recorded sensory data. The measures with which the experiment is analyzed are specified in section 4.2. Section 4.3 presents the results.

4.1. The sensory data

As mentioned, the guessing games are for a large part processed off-line on a PC. Only the sensing is done on-board. The recorded sensory data of the sensing is re-used to do multiple experiments using the same data, but also to process more games than have been recorded. The most important reason for the off-line processing has to do with time efficiency. Conducting a complete experiment on-board would take at least one week of full-time experimenting. Another advantage of processing off-board is that one can do multiple experiments in which various methods and parameter settings can be compared reliably.

The data that has been recorded for the experiments reported here, consists (after preprocessing the raw sensory data) of approximately 1,000 context settings. These context settings are used to experiment 10 runs of 10,000 guessing games. Hence in each run the context settings are used approximately 10 times. Statistics on the data set revealed that the average context size is about 3.5 segments per robot. In each game one robot is selected randomly to be the speaker, who arbitrarily selects one feature vector as the topic. Therefore it takes, in principle, approximately 7,000 games until a particular situation re-occurs.

Other statistics on the preprocessed data set revealed that the a priori chance for success is around 23.5 %. This means that when both the speaker and the hearer select a topic at random, in about 23.5 % they will select the same topic. This a priori chance is calculated from the average context size (3.5 segments) and the *potential understandability*. Since both robots not always detect the same surroundings (figure 4), the possibility exists that the speaker selects a topic that the hearer did not detect. Although in natural communication the hearer might try to find the missing information, this is not done here for practical reasons. Besides, humans communication is not always perfect, because they fail to construct a shared context. So, there is a maximum in the success to be expected, which has

been coined the potential understandability. The potential understandability has been calculated to lie around 80 %. For details on the calculation and other information about the sensory data see (Vogt, 2000b).

4.2. Measures

The experiments are investigated using six different measures. Two of these (the discriminative- and communicative success) measure the success rate of the discrimination games and the guessing games. The others measures indicate the quality of the system that emerges. These measures, which are based on the entropy measure taken from information theory (Shannon, 1948), are developed by Edwin De Jong (De Jong, 2000). They are called distinctiveness, parsimony, specificity and consistency, and are calculated every 200 guessing games. Below follows a description of these measures.

Discriminative success (Steels, 1996c) measures the number of successful discrimination games averaged over the past 100 guessing games.

Distinctiveness "Intuitively, distinctiveness expresses to what degree a meaning identifies the referent" (De Jong, 2000, p. 76). For this one can measure how the entropy of a meaning in relation to a certain referent $H(\rho|\mu_i)$ decreases the uncertainty about the referent $H(\rho)$. To do this, one can calculate the difference between $H(\rho)$ and $H(\rho|\mu_i)$. Here ρ are the referents $\rho_1, \ldots, \rho_n$ and μ_i relates to one of the meanings $\mu_1, \ldots, \mu_m$ for robot R. The distinctiveness D_R can now be defined as follows:

$$\begin{aligned} H(\rho|\mu_i) &= \sum_{j=1}^{n} -P(\rho_j|\mu_i) \cdot \log P(\rho_j|\mu_i) \\ \text{dist}(\mu_i) &= \frac{H(\rho) - H(\rho|\mu_i)}{H(\rho)} \\ &= 1 - \frac{H(\rho|\mu_i)}{H(\rho)} \\ D_R &= \frac{\sum_{i=1}^{m} P_o(\mu_i) \cdot \text{dist}(\mu_i)}{m} \end{aligned} \quad (5)$$

where $H(\rho) = \log n$ and $P_o(\mu_i)$ is the occurrence probability of meaning μ_i. The use of $P_o(\mu_i)$ as a weighting factor is to scale the importance of such a meaning to its occurrence.

Parsimony The parsimony P_R is calculated similar to the distinctiveness:

$$\begin{aligned} H(\mu|\rho_i) &= \sum_{j=1}^{m} -P(\mu_j|\rho_i) \cdot \log P(\mu_j|\rho_i) \\ \text{pars}(\rho_i) &= 1 - \frac{H(\mu|\rho_i)}{H(\mu)} \\ P_R &= \frac{\sum_{i=1}^{n} P_o(\rho_i) \cdot \text{pars}(\rho_i)}{n} \end{aligned} \quad (6)$$

with $H(\mu) = \log m$. Parsimony thus calculates to what degree a referent gives rise to a unique meaning.

Communicative success (Steels, 1996b) measures the number of successful guessing games averaged over the past 100 guessing games.

Specificity "The specificity of a word[-form] is ... defined as the relative decrease of uncertainty in determining the referent given a word that was produced" (De Jong, 2000, p. 115). It thus is a measure to indicate how well a word-form can identify a referent. It is calculated analogous to the distinctiveness and parsimony. For a set of word-forms $\sigma_1, \ldots, \sigma_q$, the specificity is defined as follows:

$$\begin{aligned} H(\rho|\sigma_i) &= \sum_{j=1}^{n} -P(\rho_j|\sigma_i) \cdot \log P(\rho_j|\sigma_i) \\ \text{spec}(\sigma_i) &= 1 - \frac{H(\rho|\sigma_i)}{H(\rho)} \\ S_R &= \frac{\sum_{i=1}^{q} P_o(\sigma_i) \cdot \text{spec}(\sigma_i)}{q} \end{aligned} \quad (7)$$

where $H(\rho) = \log n$ is defined as before and P_o is the occurrence probability of encountering word-form σ_i.

Consistency Consistency measures how consistent a referent is named by a certain word-form. It is calculated as follows:

$$\begin{aligned} H(\sigma|\rho_i) &= \sum_{j=1}^{q} -P(\sigma_j|\rho_i) \cdot \log P(\sigma_j|\rho_i) \\ \text{cons}(\rho_i) &= 1 - \frac{H(\sigma|\rho_i)}{H(\sigma)} \\ C_R &= \frac{\sum_{i=1}^{n} P_o(\rho_i) \cdot \text{cons}(\rho_i)}{n} \end{aligned} \quad (8)$$

where $H(\sigma) = \log q$ and $P_o(\rho_i)$ is defined as before.

The four entropy-based measures specify whether or not there is order in the system. When either measure has the value of 1, there is order in the system. When a measure has value 0, there is disorder. All these entropy measures are calculated per robot every 200 games within one run; the other measures are calculated after every single game. When presented, all measures are averaged over the 10 runs that are done each experiment.

4.3. The results

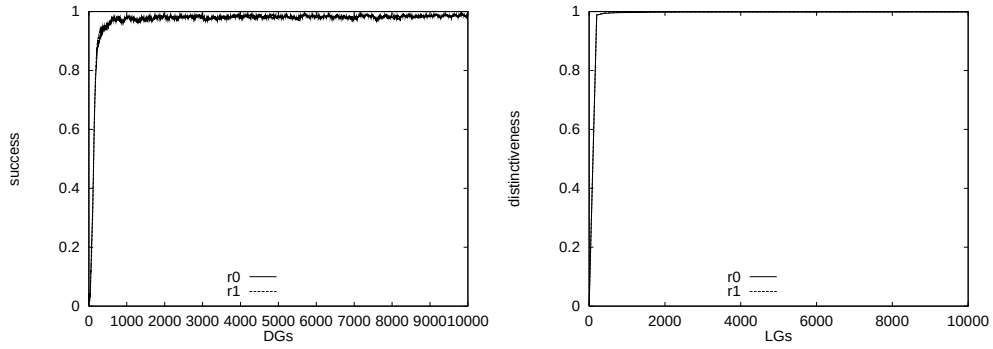
The experiment is done with 10 runs of 10,000 guessing games. Figure 6 shows the evolution of the different measures. The discriminative success approaches 100 % early in the experiment, figure 6 (a). This indicates that the discrimination game is a very efficient model for categorizing different sensings. Similar results are confirmed in other experiments using different representations of categories and varying numbers of objects, e.g., (Steels, 1996c; De Jong and Vogt, 1998). As the distinctiveness shows in figure 6 (b), when a categorization (or meaning) is used, it usually stands for one referent only. I.e., there is a one-to-one relation between meaning and referent. This, however, does not imply that there is a one-to-one relation between referent and meaning. The lower parsimony shows this, see figure 6 (c).

The communicative success (figure 6 (d)) approaches the potential understandability of 80 %. After 10,000 games, the communicative success is approximately 75 %. Hence the robots are fairly well capable of constructing a shared lexicon. The specificity, shown in figure 6 (e), increases to a value slightly above 0.9. It shows that when a form is used, it is mainly used to name one referent. Hence there is little polysemy. The consistency (figure 6 (f)) is lower than the specificity. This means that a referent is not always named with the same form. As will be shown later, this does not mean that the lexicon is inefficient, it rather means that the system bears some synonymy.

It should be noted that although the parsimony is almost as high as the consistency, this does not mean that the number of meanings used is close to the number of forms. It merely indicates that the inconsistent use of forms happen about as often as the inconsistent use of meanings.

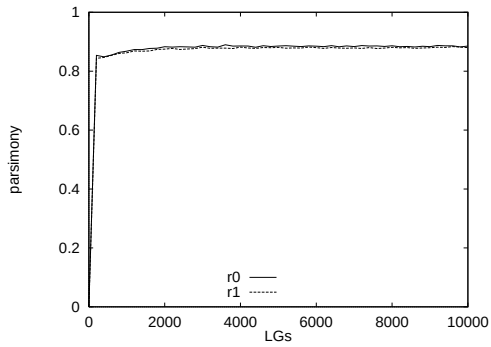
As can be seen in figures 6 (e) and (f) the specificity and consistency rise very rapidly. To understand this, it should be realized that these measures are calculated relative to the successful use of forms (in case of specificity) or to the successful naming of referents (in case of consistency). Specificity and consistency are not relative to the number of successful guessing games. So, this means that, whenever the robots communicate successfully, the used semiotic symbols reveal order in the lexicon. Similar arguments hold for the distinctiveness and parsimony, figures 6 (b) and (c).

Although still rather fast, the communicative success rises slower, see figure 6 (d). Because the agents tend to categorize the referents very differently on different occasions, there arises much synonymy and polysemy in the system. Too much synonymy and polysemy cause many confusions, so many guessing games fail. Hence the robots must disambiguate the synonymy and polysemy, which takes some time. Nevertheless, the agents perform better than chance already after a few hundred games. Kaplan has shown that the speed of convergence in communicative success depends, amongst others, on the number of meanings/referents, the number of agents and noise in

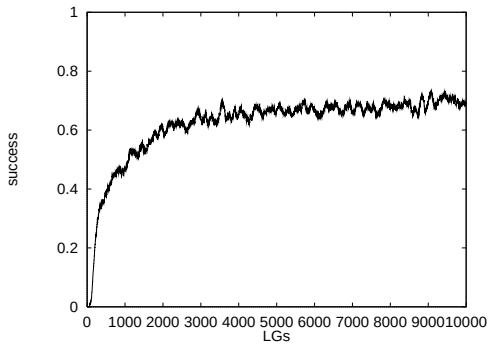


(a) Discriminative success

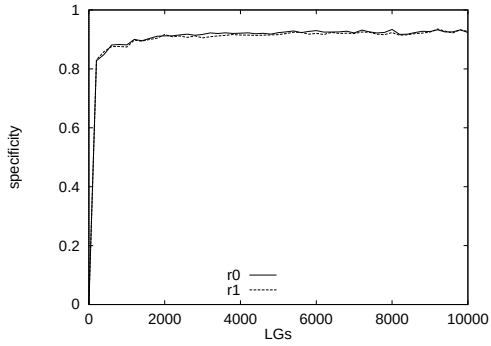
(b) Distinctiveness



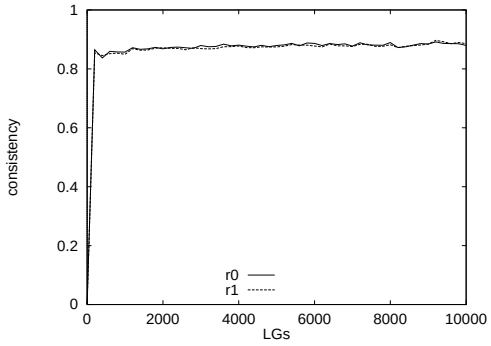
(c) Parsimony



(d) Communicative success



(e) Specificity



(f) Consistency

Figure 6. The evolution of the measures during the experiment.

transmission (Kaplan, 2001).

The run that will be discussed in more detail below resulted in the lexicon that is displayed in the semiotic landscape shown in figure 7. This figure shows the associations between referent, meaning and form for both robots with a strength that represents the relative occurrence frequency of connections that are successfully used over 10,000 guessing games. Ideally, the connections between referent-form-referent would be orthogonal. I.e., the couplings between a referent and its form should not cross-connect with other form-referent relations.

This orthogonality criterion is achieved for **mety**, **luvu** and possibly **zigi**. The word-forms **kate** and **demi** have cross-connections, but these are relatively unimportant because they occur with very low frequencies. More polysemy is found for **sema** and **tyfo**. As will be shown below, **tyfo** gets well established to name L1 almost unambiguously. The form **sema** however, provides some instability in the system.

Figure 8 shows various competition diagrams of robot r0, relating to referent L1 in one of the runs of the experiment. A competition diagram displays the relative co-occurrence frequency in time of, for instance, forms in relation to a referent (Steels and Kaplan, 1999). Figure 8 (a) shows the referent-form competition. This figure shows the successful co-occurrence of referent and form, where the occurrence of the form is calculated relative to the occurrence of the referent. Very infrequent occurring elements are left out for clarity. Figure 8 (a) shows that form **tyfo** clearly wins the competition and is nearly used uniquely to name light source L1. Hence L1 has very little synonymy. Vice-versa, the form-referent diagram shows that when the form **tyfo** is used, it is used mostly to name light source L1, see figure 8 (b). This, however, happens after game 3,000. Before this, the form **tyfo** shows quite some polysemy.

Figure 8 (c) shows that, throughout the run, L1 is categorized with more than one meaning of which two are used most frequently. A similar competition can be seen in figure 8 (d) where various meanings compete to be the meaning of **tyfo**. Apparently, these competitions compensate each other such that competitions as in figures 8 (a) and (b) emerge.

That the competition is not always running so smooth is shown in figure 9. Here there are two forms strongly competing for naming light source L2. In most cases, the forms are used to name only one referent in the end as figure 9 (b) shows. But sometimes this could fail. Nevertheless, the overall picture is that all referents are mostly named by only one form and all forms are mostly used to name only one referent. Such relations are much less frequently observed when investigating co-occurrences of referent and meaning or form and meaning.

5. Discussion

5.1. Meeting the limits

The results make clear that the robots construct a communication system that meets its limits. The communicative success is in the end nearly as high as the potential understandability.

Both the discriminative success and distinctiveness are very close to 1, and the specificity is also close to 1. So, when a robot uses a semiotic symbol successfully, it almost always refers to the same referent. This means that there is hardly any polysemy. The parsimony and consistency are somewhat lower than the distinctiveness and specificity. Hence, there are some one-to-many relations between referent and meaning and between referent and form in the system. The semiotic landscape (figure 7) already showed that most of the synonymy does not necessarily mean that the communication is difficult. Usually, the hearer can rather easily interpret any speaker's utterance.

The landscape also shows that a one-to-many relationship between form and meaning does not necessarily mean polysemy. In fact, it is beneficial, since it cancels out the one-to-many mapping of referent to meaning for a great deal. Hence semiotic symbols may have different meanings. Referents are interpreted differently when observed under different circumstances. Yet the referents can be named invariantly to a high degree. One may argue that the robots use different semiotic symbols when they use different

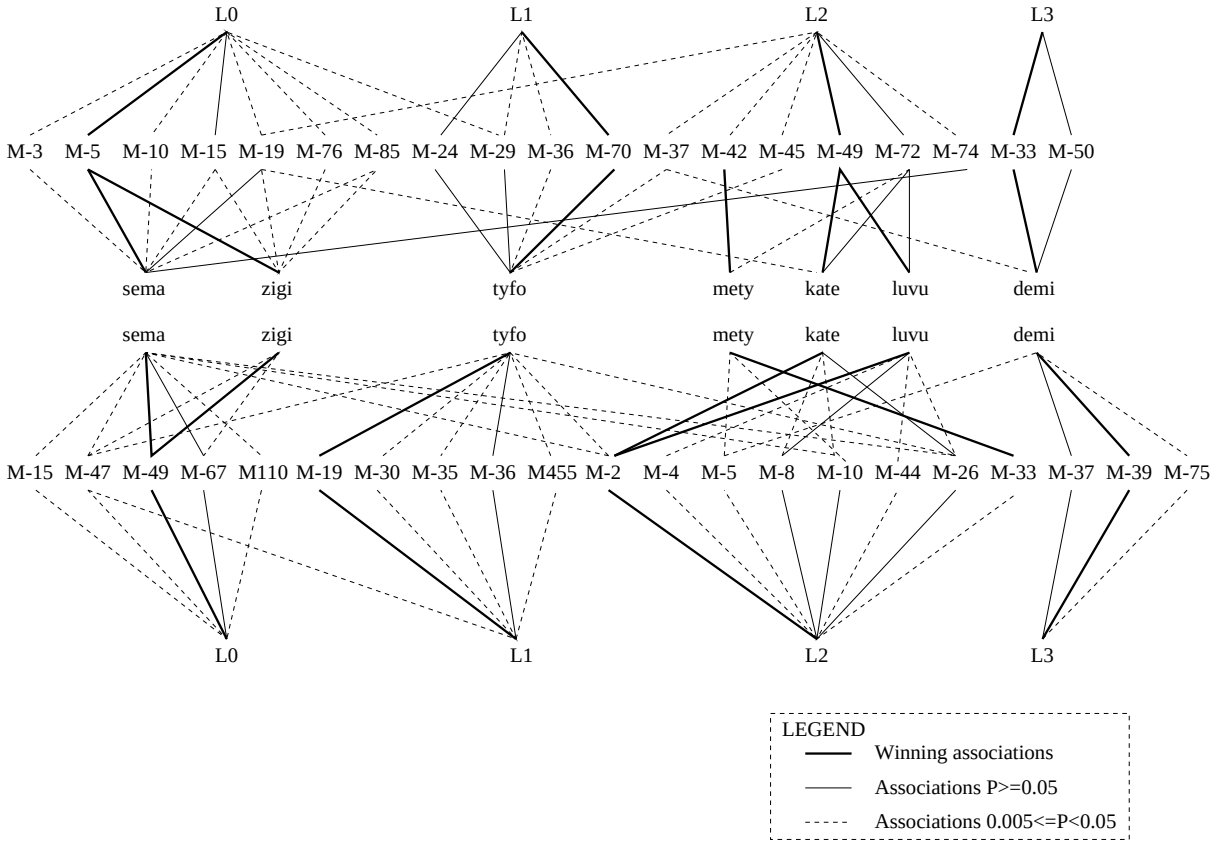
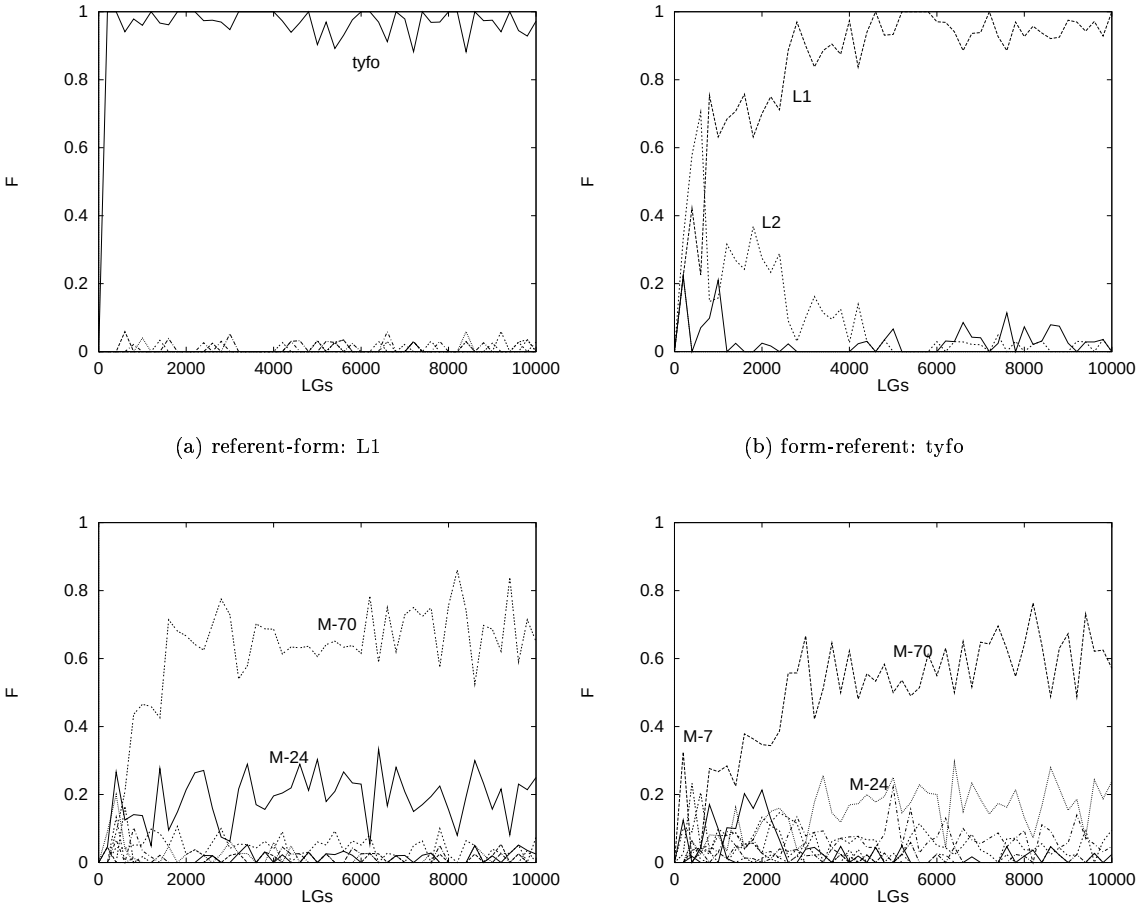


Figure 7. The semiotic landscape of the experiment. A semiotic landscape provides a way to illustrate how the semiotic symbols of the two robots are related. It illustrates the connections between referent/light source L, meaning M and forms such as **sema**, **zigi** etc.. The upper half of the graph shows the lexicon of robot r0, the lower half the lexicon of r1. The connections drawn indicate the relative occurrence frequencies (P) of referents, meanings and forms. The relations between referent and meaning are relative to the occurrence of the referent. The relations between meaning and form are relative to the occurrence of the form. Associations with an occurrence frequency of $P < 0.005$ are left out for clarity.



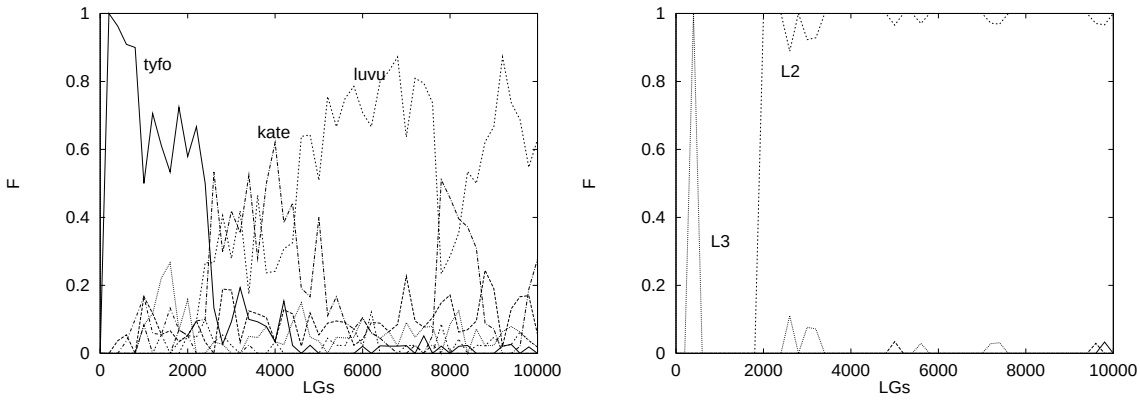
(a) referent-form: L1

(b) form-referent: tyfo

(c) referent-meaning: L1

(d) form-meaning: tyfo

Figure 8. Some competition diagrams of robot r0 in one run of the experiment. Figure (a) shows a referent-form competitions for light source L1. The y-axis shows the co-occurrence frequencies of form and referent relative to the occurrence of the referent over the past 200 guessing games. The x-axis shows the number of guessing games. Figure (b) shows the form-referent competition for **tyfo**. Again the y-axis shows the co-occurrence frequencies, but now of the referent relative to the occurrence of the form. Figure (c) shows the referent-meaning competition for L1 and (d) shows the form-meaning competition for **tyfo**.



(a) referent-form: L2

(b) form-referent: luvu

Figure 9. (a) A referent-form competition for light source L2 and (b) a form-referent competition for the form **luvu**.

meanings and sometimes this is even appropriate. However, when the semiotic symbols have the same form and relate to the same referent, attributing them to a single semiotic symbol can be useful. Especially when the semiotic symbols are only used to *name* a referent. Therefore, one could also argue that the meaning of a semiotic symbol changes dynamically over time, depending on the situation the robots are in. Note, by the way, that the meanings also change dynamically over time by the prototype shift towards detected referents, cf. eq. 4.

In what respect did the robots acquire meaningful symbols? In section 2.3.2, meaning has been defined as a functional relation between form and referent. The robots do not acquire meaning in the sense that they use the communicated symbols to fulfill a "life-task", neither is it rooted in the sense that the robots' bodies, their interaction and learning mechanisms are designed, see (Ziemke, 1999). But given the robots' bodies, interaction mechanisms and learning mechanisms, the semiotic symbols are meaningful in that they are used by the robots to name referents. They could also be used to perform simple actions, such as pointing at the referent, as shown in (Vogt, 1998a) and in the Talking Heads experiment (Steels and Kaplan, 1999). However, because this pointing could in the current experiments not be used to evaluate feedback, it has not been implemented in the current experiment. In more realistic experiments, the robots should use the communication to fulfill some task. Fulfilling tasks could then be used to evaluate the language game's success as is the case in, e.g., (De Jong, 2000; Vogt, 2000a; Kaplan, 2000). Further research is currently in progress where the robots use the lexicon to improve their learned capability to sustain their energy-level in order to 'survive'. This experiment combines the language game model with the 'viability' experiments previously done at the AI Lab in Brussels, see, e.g., (Steels, 1996a; Birk, 1998). The meaning of the semiotic symbols that are constructed in such an experiment is than based on the robots' activity to remain 'viable' over extended period of time. But even then, as argued in, e.g., (Ziemke, 1999; Ziemke and Sharkey, 2001), the semiotic symbols will only be really meaningful to an agent when the agent is completely rooted by, for instance, evolution (Ziemke, 1999). Current studies in ALife focus on how robotic agents may evolve, for

instance, their bodies (Lund et al., 1997), their control architecture (Nolfi and Floreano, 2000), their sensors (Jung et al., 2001) and communication channels (Quinn, 2001). These researches might help to explain how robots can become 'rooted' in their environment. For a broad discussion on these issues, see, e.g., (Ziemke and Sharkey, 2001).

In the current experiment the forms are transmitted as strings in the off-line processing on the PC. Previously, this has been done using radio communication (Steels and Vogt, 1997). However, in a more realistic setting, the transmission should occur in phonetic strings. In such a case, the phonetic strings are also physical objects and must be processed and categorized. Both processing and categorization could, for instance, be done in a similar way as modeled by De Boer (1997), see section 3.1. Using De Boer's model, vowels, or even more complex utterances (De Boer, 2000b; Oudeyer, 2001), could be developed in a similar way as the lexicon is developed. One non-trivial problem remains to be solved. This problem has to do with distinguishing utterances as forms from physical objects. How this problem can be solved is unclear, but it may depend on the context setting, and for that the agents need to develop more sophisticated means of recognizing a context. But perhaps it could also be solved by the evolution of communication channels as, for instance, is being investigated in (Quinn, 2001).

5.2. The dynamics of the system

What can be said about the dynamics of the system? Figure 6 (d) showed that the communicative success shows a rapid increase during the first 1,000 games, after which the success grows less rapidly. Furthermore, the figure shows a slower increase than, e.g., the discriminative success shown in figure 6 (a). So, what happens in the beginning? In the first few hundred games, when the robots try to name the various referents, there is a lot of confusion what the speaker intends to communicate. The hearer of a game adopts forms that may already exists for the robot, but that are not applicable in the current context (i.e. the meaning does not apply to any distinctive category). Consequently, a lot of variety is introduced in the lexicon. As a result of adapting the association scores, the robots will tend to select effective elements appropriately more often. This induces a repetitive cycle that strengthens the effective elements more and meanwhile weakens the ineffective ones. In the beginning this adaptation is more flexible than later on when the association scores become stronger. When the association scores are strong, it is more difficult for other associations to influence the communicative success. This, however, becomes less important as the communicative success rises to a satisfactory degree.

The dynamics of the system allows an important conclusion to be drawn, namely that the selection criteria and the dynamics of association scores cause a self-organizing convergence of the lexicon. The conclusion that this is an emergent property can, amongst others, be drawn from the fact that lexicon dynamics is controlled at the form-meaning level and not at the form-referent level. I.e., adaptations in the lexicon occur only at the form-meaning level. A referent is categorized differently in different situations, see, e.g., figure 8 (c). At the same time, these different categories may be named by only one name (or more concretely, one name may have different meanings in different situations) as figure 8 (d) depicts. Figures 8 (a) and (b) show that the one-to-many relations at the referent-meaning and form-meaning levels cancel each other out at the referent-form level in both directions. This happens despite the fact that when a form-meaning association is used successfully, the strength of competing form-meaning associations are laterally inhibited. Although this lateral inhibition helps to decrease polysemy and synonymy at the referent-form level, it is also an antagonizing force at the form-meaning level when the meaning is used to stand for the same referent. This antagonizing force, however, is not problematic due to the context dependence of the guessing games. Selected lexical elements must be applicable within the context of a particular game. This is a nice consequence of the pragmatic approach. Furthermore, the feedback signals that operate at the form-referent level contribute largely to the

convergence of the lexicon. It has been shown in (Vogt, 2000b) that leaving out the feedback without alternative ways of knowing the topic, does not lead to convergence in this experimental setup. This does not mean, however, that leaving out the feedback does not work in general. It may well be that not using such feedback, or any other non-verbal means of exchanging topic knowledge, might lead to convergence in a more rich environment, cf. the results of simulations reported in (Smith, 2001). This is currently being investigated. Another strategy that could be beneficial in such cases is to use a cooperative rather than a competitive selection and adaptation scheme as demonstrated in (Kaplan, 2001).

5.3. The relation to other work

The experiment presented here is unique in its modeling the development of a lexicon grounded in reality *from scratch* using mobile robots. Although the Talking Heads experiment (Belpaeme et al., 1998; Steels and Kaplan, 1999) also models lexicon development from scratch and is also a grounded experiment, the Talking Heads are immobile (they can only move pan-tilt from a fixed position). This immobility helped to evaluate the feedback on guessing games, because the Talking Heads used calibrated knowledge about their environment to evaluate feedback. In addition, the different sensings of a referent are more similar on different occasions than in the mobile robots as the Talking Heads sense their environment from a fixed position. In controlled experiments, this allowed to the Talking Heads to speed up the lexicon development (Steels and Kaplan, 1999). Nevertheless, the overall success on both platforms is more or less comparable.

Similar findings are found when comparing the work of this paper with the work of Billard and her colleagues on mobile robots (Billard and Hayes, 1997; Billard and Dautenhahn, 1998). In Billard's experiments, a student robot learns a grounded lexicon about its environment by imitating a teacher robot who has been preprogrammed with the lexicon. The overall results are similar to the results presented here, although the lexicon acquisition is much faster in (Billard and Dautenhahn, 1998). The latter result is presumably due to the fact that one of Billard's robots already knows the lexicon, while in the experiments here none of the robots has been preprogrammed with the lexicon.

Although not situated and embodied, the simulations of Oliphant, too, are relevant. He showed that populations of agents could rather easily develop highly efficient lexicons from scratch by associating given meanings with given forms (Oliphant, 1999). Also relevant is the work of De Jong who experimented with language games, of which the meanings were grounded in simulations (De Jong, 2000). In addition, De Jong's agents tried to improve their (task-oriented) behavior using the lexicon developed. Both Oliphant and De Jong showed that agents could develop a coherent lexicon without using feedback on the effect of the game, provided the agents have access to both the form and meaning during a language game. Although in this paper the robots did use such feedback, robotic experiments have confirmed Oliphant and De Jong's results (Vogt, 2000b, Vogt, 2001). In these experiments both robots had access to both the form and the topic by means of 'pointing', so that the hearer knows the topic in advance.

6. Conclusions

In order to overcome fundamental problems that exist in the cognitivist approach towards cognition, and to allow describing cognition in terms of symbols within the paradigm of embodied cognition, this paper proposes an alternative definition of symbols. This definition is not novel, but is adopted from Peirce's definition of symbols as the relation between a form, a meaning and a referent. The relation as such is not meaningful, but arises from its active construction and use. This process is called semiosis and has been modeled in robotic agents through adaptive language games.

As a result of the semiotic definition of symbols, it could be argued that the symbols are per definition grounded, because semiotic symbols have intrinsic meaning in relation to the real world, cf. (Lakoff, 1987). Hence the symbol grounding problem is no longer a fundamental problem, since a semiotic symbol is a relation between a

form, meaning and referent, and the way this relation is formed and used specifies its meaning. There, however, remains the problem of constructing the semiotic symbols, but this problem can be viewed as a technical problem. This problem is called the *physical symbol grounding problem.*

The experiment reported shows how robotic agents can develop a structured set of semiotic symbols which they can use to name various real world objects invariantly. The semiotic symbols are constructed through the robots' interactions with their environment (including themselves), individual adaptations and self-organization. These three mechanisms, hypothesized by Luc Steels (1996a), are based on the core principles of embodied cognitive science: embodiment, situatedness and emergence. The semiotic symbols are structural couplings that are formed through some interaction of an agent with the real world as proposed for instance by (Clancey, 1997).

An important result that the experiment reveals is that semiotic symbols need not be categorized the same under different circumstances. As the semiotic landscape shows, there is no one-to-one-to-one relation between a referent, meaning and form; this relation is rather one-to-many-to-one. In different situations, the robots detect the referents differently. Yet they are able to identify them invariantly at the form level. In the process of arriving at such invariant identification, which is the most important aspect of symbol grounding (Harnad, 1990), the co-evolution of form and meaning reveals to be extremely important.

The experiment shows that the physical symbol grounding problem can be solved in the simple experimental setup, given the language game model, the designed robots and under the assumption that feedback can be established by, for instance, using pointing. These given assumptions make that the physical symbol grounding problem is not entirely solved, because for this the language game model, the robots and other assumptions should be rooted by, e.g., evolution (Ziemke, 1999; Ziemke and Sharkey, 2001). The experiment nevertheless illustrates how semiotic symbols can be constructed and used and is thus an important step towards solving the physical symbol grounding problem, or at least, in our understanding of cognition.

Although the guessing game works well in the current experimental setup, it should be realized that this setup is rather simplistic. Future work should confirm the scalability of the model in more realistic and more complex environments using more complex robots. Another improvement that is currently under investigation, is that the communication system is used to perform concrete "life-tasks" rather than just using and developing a lexicon. This would make the approach more realistic since in natural systems communication is usually used to guide (task-oriented) behavior, such as coordinating each other's actions.

Acknowledgements

The author wishes to thank Ida Sprinkhuizen-Kuyper, Edwin De Jong and Eric Postma for proofreading earlier versions of this paper. Georg Dorffner, Erik Myin, Tom Ziemke and two anonymous reviewers are thanked for various useful comments that helped improve this paper a lot.

References

ACL (1997). *Proceedings of the Fifth Conference on Applied Natural Language Processing*, Washington, DC. Menlo Park: Association for Computational Linguistics.

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

Batali, J. (1998). Computational simulations of the emergence of grammar. In Hurford, J. R., Studdert-Kennedy, M., and Knight, C., editors, *Approaches to the Evolution of Language*, Cambridge, UK. Cambridge University Press.

Belpaeme, T. (1999). Evolution of visual feature detectors. In *Evolutionary Computation in Image Analysis and Signal Processing and Telecommunications First European Workshops, EvoIASP99 and EuroEcTel99 Joint*

Proceedings. Göteborg, Sweden. LNCS 1596, Berlin. Springer-Verlag.

Belpaeme, T. (2001). Simulating the formation of color categories. In *Proceedings of the International Joint Conference on Artificial Intelligence 2001 (IJCAI'01)*, Seattle, WA.

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.

Bickerton, D. (1998). Catastrophic evolution: the case for a single step from protolanguage to full human language. In Hurford, J., Knight, C., and Studdert-Kennedy, M., editors, *Approaches to the evolution of language*, pages 341–358, Cambridge. Cambridge University Press.

Billard, A. and Dautenhahn, K. (1998). Grounding communication in autonomous robots: an experimental study. *Robotics and Autonomous Systems*, 24(1-2):71–79.

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.

Birk, A. (1998). Robot learning and self-sufficiency: What the energy-level can tell us about a robot's performance. In Birk, A. and Demiris, J., editors, *Learning Robots: Proceedings of EWLR-6, Lecture Notes on Artificial Intelligence 1545*, pages 109–125. Springer-Verlag.

Boden, M. A. (1996). *The Philosophy of Artificial Life*. Oxford University Press, Oxford.

Braitenberg, V. (1984). *Vehicles, Experiments in Synthetic Psychology*. The MIT Press, Cambridge MA.

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

Brooks, R. A. (1991). Intelligence without representation. *Artificial Intelligence*, 47:139–159.

Brooks, R. A., Breazeal (Ferrell), C., Irie, R., Kemp, C. C., Marjanović, M., Scassellati, B., and Williamson, M. M. (1998). Alternative essences of intelligence. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence*, Menlo Park Ca. AAAI Press.

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

Chomsky, N. (1980). Rules and representations. *The behavioral and brain sciences*, 3:1–61.

Chomsky, N. and Halle, M. (1968). *The sound pattern of English*. The MIT Press, Cambridge, MA.

Churchland, P. M. (1989). *A Neurocomputational Perspective*. The MIT Press.

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

De Boer, B. (1997). Generating vowels in a population of agents. In Husbands, C. and Harvey, I., editors, *Proceedings of the Fourth European Conference on Artificial Life*, pages 503–510, Cambridge Ma. and London. MIT Press.

De Boer, B. (2000a). Emergence of vowel systems through self-organisation. *AI Communications*, 13:27–39.

De Boer, B. (2000b). Imitation games for complex utterances. In Van den Bosch, A. and Weigand, H., editors, *Proceedings of the Belgian-Netherlands Artificial Intelligence Conference*, pages 173–182.

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

De Jong, E. D. and Steels, L. (1999). Generation and selection of sensory channels. In *Evolutionary Computation in Image Analysis and Signal Processing and*

Telecommunications First European Workshops, EvoIASP99 and EuroEcTel99 Joint Proceedings. Göteborg, Sweden. LNCS 1596, Berlin. Springer-Verlag.

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., editors, *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.* Fontana, New York.

Dorffner, G. (1992). Taxonomies and part-whole hierarchies in the acquisition of word meaning - a connectionist model. In *Proceedings of 14th Annual Conference of the Cognitive Science Society*, pages 803–808, New Haven/Hillsdale/Hove. Lawrence Erlbaum Associates.

Dorffner, G., Prem, E., and Trost, H. (1993). Words, symbols, and symbol grounding. Technical Report TR-93-30, Oesterreichisches Forschungsinstitut fuer Artificial Intelligence, Wien. Available online at ftp://ftp.ai.univie.ac.at/papers/oefai-tr-93-30.ps.gz.

Eco, U. (1976). *A theory of semiotics.* Indiana University Press, Bloomington.

Edelman, G. M. (1987). *Neural Darwinism.* Basic Books Inc., New York.

Fu, K. S., editor (1976). *Digital Pattern Recognition*, Berlin. Springer-Verlag.

Gibson, J. J. (1979). *The Ecological Approach to Visual Perception.* Houghton-Mifflin, Boston.

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.

Johnson, M. H. (1997). *Developmental Cognitive Science.* Blackwell Publishers, Oxford, UK.

Jung, T., Dauscher, P., and Uthmann, T. (2001). Some effects of individual learning on the evolution of sensors. In Kelemen, J. and Sosík, P., editors, *Proceeding of the 6th European Conference on Artificial Life, ECAL 2001*, LNAI 2159, pages 432–435, Berlin Heidelberg. Springer-Verlag.

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

Kaplan, F. (2001). *La naissance d' une langue chez les robots.* Hermes Science.

Kirby, S. and Hurford, J. (1997). Learning, culture and evolution in the origin of linguistic constraints. In Husbands, C. and Harvey, I., editors, *Proceedings of the Fourth European Conference on Artificial Life*, Cambridge Ma. and London. MIT Press.

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

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

MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297.

Maturana, H. R. and Varela, F. R. (1992). *The tree of knowledge: the biological roots of human understanding.* Shambhala, Boston.

Maynard-Smith, J. and Szathmáry, E. (1995). *The Major Transitions in Evolution.* W. H. Freeman and Company Ltd., Oxford.

McCarthy, J. and Hayes, P. J. (1969). Some philosophical problems from the standpoint of artificial intelligence. *Machine Intelligence*, 4:463–502.

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

Newell, A. (1990). *Unified Theories of Cognition*. Harvard University Press, Cambridge, Ma.

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

Noble, J. (2000). Talk is cheap: Evolved strategies for communication and action in asymmetrical animal contests. In Meyer, J.-A., Berthoz, A., Floreano, D., Roitblat, H., and Wilson, S. W., editors, *From Animals to Animats 6: Proceedings of the Sixth International Conference on Simulation of Adaptive Behavior*, pages 481–490, Cambridge, MA. MIT Press.

Nolfi, S. and Floreano, F. (2000). *Evolutionary Robotics: The Biology, Intelligence, and Technology of Self-Organizing Machines*. The MIT Press, Cambridge, MA. / London.

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

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

Oudeyer, P.-Y. (2001). The origins of syllable systems : an operational model. In *Proceedings of the International Conference on Cognitive science, COGSCI'2001*, Edinburgh.

Peirce, C. S. (1931-1958). *Collected Papers*, volume I-VIII. Harvard University Press, Cambridge Ma.

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

Pinker, S. and Bloom, P. (1990). Natural language and natural selecion. *Behavioral and brain sciences*, 13:707–789.

Prem, E. (1995). Symbol grounding and transcedental logic. In Niklasson, L. F. and Bodén, M. B., editors, *Current Trends in Connectionism: proceedings of the Swedish Conference on Connectionism*, pages 271–282, New Haven/Hillsdale/Hove. Lawrence Erlbaum.

Prigogine, I. and Strengers, I. (1984). *Order out of Chaos*. Bantam Books, New York.

Pylyshyn, Z. W., editor (1987). *The Robot's Dilemma*, New Jersey. Ablex Publishing Corporation.

Quinn, M. (2001). Evolving communication without dedicated communication channels. In Kelemen, J. and Sosík, P., editors, *Proceeding of the 6th European Conference on Artificial Life, ECAL 2001*, LNAI 2159, pages 357–366, Berlin Heidelberg. Springer-Verlag.

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

Roy, D. (2000). A computational model of word learning from multimodal sensory input. In *International conference of cognitive modeling*, Groningen, The Netherlands.

Searle, J. R. (1980). Minds, brains and programs. *Behavioral and Brain Sciences*, 3:417–457.

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 Kelemen, J. and Sosík, P., editors, *Proceeding of the 6th European Conference on Artificial Life, ECAL 2001*, LNAI 2159, pages 381–390, Berlin Heidelberg. Springer-Verlag.

Steels, L. (1996a). Discovering the competitors. *Adaptive Behavior*, 4(2):173–199.

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

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

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

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

Steels, L. and Kaplan, F. (1998). Stochasticity as a source of innovation in language games. In *Proceedings of Alive VI.*

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

Steels, L. and McIntyre, A. (1999). Spatially distributed naming games. *Advances in complex systems*, 1(4).

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.

Sun, R. (2000). Symbol grounding: A new look at an old idea. *Philosophical Psychology*, 13(2):149–172.

Tomasello, M. and Barton, M. (1994). Learning words in nonostensive contexts. *Developmental Psychology*, 30(5):639–650.

Van Looveren, J. (1999). Multiple word naming games. In Postma, E. and Gyssens, M., editors, *Proceedings of the Eleventh Belgium-Netherlands Conference on Artificial Intelligence.* University of Maastricht.

Vogt, P. (1998a). The evolution of a lexicon and meaning in robotic agents through self-organization. In La Poutré, H. and van den Herik, J., editors, *Proceedings of the Netherlands-Belgium Artificial Intelligence Conference*, Amsterdam. CWI Amsterdam.

Vogt, P. (1998b). Perceptual grounding in robots. In Birk, A. and Demiris, J., editors, *Learning Robots, Proceedings of the EWLR-6, Lecture Notes on Artificial Intelligence 1545.* Springer-Verlag.

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

Vogt, P. (2000b). *Lexicon Grounding on Mobile Robots.* PhD 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.*

Whorf, B. L. (1956). *Language, Thought, and Reality.* MIT Press, Cambridge Ma.

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

Yanco, H. and Stein, L. (1993). An adaptive communication protocol for cooperating mobile robots. In Meyer, J.-A., Roitblat, H. L., 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.

Ziemke, T. (1999). Rethinking grounding. In Riegler, A., Peschl, M., and von Stein, A., editors, *Understanding Representation in the Cognitive Sciences: Does Representation Need Reality?*, New York. Plenum Press.

Ziemke, T. and Sharkey, N. E. (2001). A stroll through the worlds of robots and animals: Applying Jakob von Uexküll's theory of meaning to adaptive robots and artificial life. *Semiotica*, 134(1-4):701–746.