How mobile robots can self-organise a vocabulary

Paul Vogt

Computational Models of Language Evolution 2

language science press

Computational Models of Language Evolution

Editors: Luc Steels, Remi van Trijp

In this series:

1. Steels, Luc. The Talking Heads Experiment: Origins of words and meanings.

2. Vogt, Paul. How mobile robots can self-organize a vocabulary.

3. Bleys, Joris. Language strategies for the domain of colour.

4. van Trijp, Remi. The evolution of case grammar.

5. Spranger, Michael. The evolution of grounded spatial language.

ISSN: 2364-7809

How mobile robots can self-organise a vocabulary

Paul Vogt

language
science
press

Paul Vogt. 2015. *How mobile robots can self-organise a vocabulary* (Computational Models of Language Evolution 2). Berlin: Language Science Press.

This title can be downloaded at:
http://langsci-press.org/catalog/book/50

ISBN: 978-3-944675-43-5 (Digital)
978-3-946234-00-5 (Hardcover)
978-3-946234-01-2 (Softcover)
ISSN: 2364-7809

Cover and concept of design: Ulrike Harbort
Fonts: Linux Libertine, Arimo
Typesetting software: XeLaTeX

Language Science Press
Habelschwerdter Allee 45
14195 Berlin, Germany
langsci-press.org

Storage and cataloguing done by FU Berlin

Freie Universität Berlin

Preface

You are currently reading the book version of my doctoral dissertation which I successfully defended at the Vrije Universiteit Brussel on 10^{th} November 2000, slightly more than 13 years ago at the time of writing this preface. I feel privileged 15? to have been the very first to implement Luc Steels' language game paradigm on a robotic platform. As you will read, the robots I used at that moment were very limited in their sensing, computational ressources and motor control. Moreover, I spent much time repairing the robots, as they were built from LEGO parts (not LEGO Mindstorms, which was not yet available at the start of my research) and a homemade sensorimotor board. As a result, the experimental setup and the evolved lexicons were also very limited. Nevertheless, the process of implementing the model, carrying out the experiments and analysing these, has provided a wealth of insights and knowledge on lexicon grounding in an evolutionary context, which, I believe, are still relevant today.

Much progress has been made since the writing of this dissertation. First, the language game paradigm has been implemented in more advanced robots, starting with the Talking Heads (Steels et al. 2002) and the Sony Aibo (Steels & Kaplan 2000), which emerged while I was struggling with the LEGO robots, then soon followed by various humanoid platforms, such as Sony's Qrio (see, e.g. Steels 2012 and this book series). Second, the cognitive architecture has become much more advanced through the development of FCG (Steels & Beule 2006), which allowed for more complex languages to emerge, resembling more closely natural languages. Third, the underlying processes of language games, in particular of the naming game, and the resulting dynamics in an evolutionary context have been widely studied using methods stemming from statistical mechanics (e.g., Baronchelli et al. 2006).

During the first years after the completion of my dissertation, I have published various studies from this book as journal articles (Vogt 2000a, 2002, 2003a). A broader review of using robots in studies of language evolution has appeared in Vogt 2006. Building further on the work presented in this book, I formulated the PHYSICAL SYMBOL GROUNDING HYPOTHESIS (Vogt 2002). This hypothesis essentially states that Harnad's (1990) symbol grounding problem is not a philosophical problem, but a technical problem that needs to be addressed by (virtual) robotic agents situated in a (virtual) environment, provided we adopt Peirce's semiotics, because according to this view, symbols have per definition meaning. As physical symbol grounding can in principle be achieved by individual agents, the ability to develop a shared symbolic communication system is a (much) harder challenge. This challenge, which I have called SOCIAL SYMBOL GROUNDING (Vogt & Divina 2007), has remained my primary research focus.

The fact that I worked in a lab without robotic platforms forced me to continue my research in simulations. Although simulations move away from the advantages of studying physically situated language development, it allowed me to scale up and, not unimportantly, speed up the research. Together with Hans Coumans, I reimplemented the three types of language games studied in this book (the observational game, the guessing game and what I then called the selfish game) in a simulation to demonstrate that the selfish game can work properly (Vogt & Coumans 2003), despite the results presented in this book. In my dissertation, the term "selfish game" was used to indicate that the hearer had to interpret an utterance solely based on the utterance and the context without receiving additional cues through joint attention or feedback. I later discovered that the statistical learning method I implemented is known as CROSS-SITUATIONAL LEARNING (Pinker 1984; Siskind 1996). As I have worked a lot on cross-situational learning (XSL) over the past decade, I have decided to change the term "selfish game" into XSL game. Apart from a few small typos, this is the only change made with respect to the original dissertation.

Over the years, I have become convinced that XSL is the basic learning mechanism that humans use to learn word-meaning mappings. XSL learning allows the learner to infer the meaning of a word by using the covariation of meanings that occur in the contexts of different situations. In Smith et al. (2006), we have shown that XSL can be highly robust under large amounts of referential uncertainty (i.e. a lexicon can be learned well even when an agent hears a word in contexts containing many possible meanings). However, this was shown using a mathematical model containing many unrealistic assumptions. When relaxing such assumptions, such as using a robot (cf. this book), having many agents in the population (Vogt & Coumans 2003) or assuming that words and meanings occur following a Zipfian distribution (Vogt 2012), XSL is no longer that powerful. To resolve this, a learner requires additional cues to learn a human-size lexicon, such as joint attention or corrective feedback.

These ideas were further elaborated in the EU funded New Ties project (Gilbert et al. 2006), in which we aimed to set up a large scale ALife simulation, containing thousands of agents who "lived" in a complex environment containing all sorts of objects, who could move around, and who would face all sorts of challenges in order to survive. The agents would learn to survive through evolutionary adaptation based on genetic transmission, individual learning and social learning of skills and language. Although we only succeeded partially, an interesting modification of the language game was implemented. In this implementation, agents could engage in a more dialogue-like interaction requesting additional cues or testing learnt vocabulary. They could also pass on learnt skills to other agents using the evolved language. The interactions could involve both joint attention and corrective feedback to reduce referential uncertainty, while learning was achieved through XSL (Vogt & Divina 2007; Vogt & Haasdijk 2010).

Another line of research that I have carried out after writing this book, combined the language game paradigm with Kirby and Hurford's 2002 ITERATED LEARNING MODEL, studying the emergence of compositional structures in language (Vogt 2005a,b). This hybrid model, implemented in the simulation toolkit THSim (Vogt 2003b), simulates the Talking Heads experiment.[1] These studies have provided fundamental insights on how compositionality might have evolved through cultural evolution by means of social interactions, social learning and self-organisation. Population dynamics, transmission over generations, and the active acquisition of language and meaning were considered crucial ingredients of this model (for an overview of the results, see Vogt 2007).

While I was making good progress with all this modelling work, providing interesting and testable predictions on language evolution and language acquisition, I increasingly realised the importance of validating these predictions with empirical data from studies with humans (or other animals). Together with Bart de Boer, we organised a week-long meeting in which language evolution modellers working on various topics were coupled to researchers working on empirical data from various fields, such as child language acquisition, animal communication, cognitive linguistics, etc. In this workshop, novel approaches to compare our models as closely as possible to empirical findings were developed (Vogt & de Boer 2010).

As there is virtually no empirical data on the evolution of word-meaning mappings, the most straightforward comparison that could be made with my modelling research was to compare to child language acquisition (Vogt & Lieven 2010). Although there is a wealth of data on child language acquisition, none was found that captured the data needed to make a reliable comparison. Therefore, I decided to collect the data myself. This resulted in a project on which I have worked

[1] Downloadable from http://ilk.uvt.nl/~pvogt/thsim.html.

for over the past five years. Its aim is to develop longitudinal corpora of children's interactions with their (social) environment from different cultures (the Netherlands and Mozambique), together with parental estimates of the children's vocabulary size at different ages during children's second year of life. In these corpora, recordings of naturalistic observations are annotated based on the type of interactions (e.g. dyadic vs. triadic interactions), the use of gestures such as pointing, the use of feedback, the child-directed speech and the children's social network of interactions. The resulting corpora contain statistical descriptions of the types of interactions and stimuli which the children from the different cultures encounter. The idea is that these corpora can be used to set the parameters of language game simulations similar to the one described in Vogt & Haasdijk (2010). The aim is to simulate observed naturalistic interactions and to compare the lexicon development of the artificial agents with that of the simulated children. If the predictions from the simulations match the observed development of the children, then we may be confident that the model is an accurate (or at least highly plausible) theory of children's language acquisition. Development of the ultimate model, however, may take another 13 years. (For more details on this approach, consult Vogt & Mastin 2013.)

Now, let us move on to where it all started for me. Before going there, however, I would like to apologise for any mistake that you may encounter, or visions I may no longer adhere to, and which could easily have been repaired if I would have had the time. Enjoy the rest of the journey.

Paul Vogt
Tilburg, November 2013.

Acknowledgments

In 1989 I started to study physics at the University of Groningen, because at that time it seemed to me that the working of the brain could best be explained with a physics background. Human intelligence has always fascinated me, and I wanted to understand how our brains could establish such a wonderful feature of our species. After a few years I got disappointed in the narrow specialisation of a physicist. In addition, it did not provide me the answers to the question I had. Fortunately, the student advisor of physics, Professor Hein Rood introduced to me a new study, which would start in 1993 at the University of Groningen (RuG). This study was called "cognitive science and engineering", which included all I was interested in. Cognitive science and engineering combined physics (in particular biophysics), artificial intelligence, psychology, linguistics, philosophy and neuroscience in an technical study in intelligence. I would like to thank Professor Rood very much for that.

This changed my life. After a few years of study, I became interested in robotics, especially the field of robotics that Luc Steels was working on at the AI Lab of the Free University of Brussels. In my last year I had to do a research project of six months resulting in a Master's thesis. I was pleased to be able to do this at Luc Steels' AI Lab. Together we worked on our first steps towards grounding language on mobile robots, which formed the basis of the current PhD thesis. After receiving my MSc degree (*doctoraal* in Dutch) in cognitive science and engineering, Luc Steels gave me the opportunity to start my PhD research in 1997.

I would like to thank Luc Steels very much for giving me the opportunity to work in his laboratory. He gave me the chance to work in an extremely motivating research environment on the top floor of a university building with a wide view over the city of Brussels and with great research facilities. In addition, his ideas and our fruitful discussions showed me the way to go and inspired me to express my creativity.

Many thanks for their co-operation, useful discussions and many laughs to my friends and (ex-)colleagues at the AI Lab Tony Belpaeme, Karina Bergen, Andreas Birk, Bart de Boer, Sabine Geldof, Edwin de Jong, Holger Kenn, Dominique Osier, Peter Stuer, Joris Van Looveren, Dany Vereertbrugghen, Thomas Walle and all those who have worked here for some time during my stay. I cannot forget to thank my colleagues at the Sony CSL in Paris for providing me with a lot of interesting ideas and the time spent during the inspiring off-site meetings: Frédéric Kaplan, Angus McIntyre, Pierre-Yves Oudeyer, Gert Westermann and Jelle Zuidema.

Students Björn Van Dooren and Michael Uyttersprot are thanked for their very helpful assistance during some of the experiments. Haoguang Zhu is thanked for translating the title of this thesis into Chinese.

The teaching staff of cognitive science and engineering have been very helpful for giving me feedback during my study and my PhD research, especially thanks to Tjeerd Andringa, Petra Hendriks, Henk Mastebroek, Ben Mulder, Niels Taatgen and Floris Takens. Furthermore, some of my former fellow students from Groningen had a great influence on my work through our many lively discussions about cognition: Erwin Drenth, Hans Jongbloed, Mick Kappenburg, Rens Kortmann and Lennart Quispel. Also many thanks to my colleagues from other universities that have provided me with many new insights along the way: Ruth Aylett, Dave Barnes, Aude Billard, Axel Cleeremans, Jim Hurford, Simon Kirby, Daniel Livingstone, Will Lowe, Tim Oates, Michael Rosenstein, Jun Tani and those many others who gave me a lot of useful feedback.

Thankfully I also have some friends who reminded me that there was more in life than work alone. For that I would like to thank Wiard, Chris and Marcella, Hilde and Gerard, Herman and Xandra and all the others who somehow brought lots of fun in my social life.

I would like to thank my parents very much for their support and attention throughout my research. Many thanks to my brother and sisters and inlaws for being there for me always. And thanks to my nieces and nephews for being a joy in my life

Finally, I would like to express my deepest gratitude to Miranda Brouwer for bringing so much more in my life than I could imagine. I thank her for the patience and trust during some hard times while I was working at a distance. I dedicate this work to you.

Brussels, November 2000

Contents

Preface **iii**

Acknowledgments **vii**

1 Introduction **1**
1.1 Symbol grounding problem . . . 3
1.1.1 Language of thought . . . 3
1.1.2 Understanding Chinese . . . 5
1.1.3 Symbol grounding: philosophical or technical? . . . 8
1.1.4 Grounding symbols in language . . . 11
1.1.5 Physical grounding hypothesis . . . 12
1.1.6 Physical symbol grounding . . . 14
1.2 Language origins . . . 18
1.2.1 Computational approaches to language evolution . . . 19
1.2.2 Steels' approach . . . 20
1.3 Language acquisition . . . 24
1.4 Setting up the goals . . . 26
1.5 Contributions . . . 28
1.6 The book's outline . . . 31

2 The sensorimotor component **33**
2.1 The environment . . . 33
2.2 The robots . . . 34
2.2.1 The sensors and actuators . . . 35
2.2.2 Sensor-motor board II . . . 41
2.3 The Process Description Language . . . 42
2.4 Cognitive architecture in PDL . . . 48
2.5 Summary . . . 52

Indexes **53**
Name index . . . 53
Language index . . . 53

Subject index 53

Bibliography **53**

1 Introduction

L'intelligence est une adaptation.
(Piaget 1996)

One of the hardest problems in artificial intelligence and robotics is what has been called the SYMBOL GROUNDING PROBLEM (Harnad 1990). The question how "seemingly meaningless symbols become meaningful" (Harnad 1990) is a question that also holds grip of many philosophers for already more than a century, e.g. Bretano (1874), Searle (1980) and Dennett (1991).[1] With the rise of artificial intelligence (AI), the question has become very actual, especially within the symbolic paradigm (Newell 1990).[2] The symbol grounding problem is still a very hard problem in AI and especially in robotics (Pfeifer & Scheier 1999).

The problem is that an agent, be it a robot or a human, perceives the world in analogue signals. Yet humans have the ability to categorise the world in symbols that they, for instance may use for language. The perception of something, like e.g. the colour red, may vary a lot when observed under different circumstances. Nevertheless, humans are very good at recognising and naming this colour under these different conditions. For robots, however, this is extremely difficult. In many applications the robots try to recognise such perceptions based on the rules that are pre-programmed. But there are no singular rules that guide the conceptualisation of red. The same argument holds for many, if not all perceptions. A lot of solutions to the symbol grounding problem have been proposed, but there are still many limitations on these solutions.

Intelligent systems or, as Newell (1980) called them, "physical symbol systems" should amongst others be able to use symbols, abstractions and language. These symbols, abstractions and language are always about something. But how do they become that way? There is something going on in the brains of language users that give meaning to these symbols. What is going on is not clear. It is clear from neuroscience that active neuronal pathways in the brain activate mental

[1] In philosophy the problem is usually addressed with the term "intentionality", introduced by Bretano (1874).

[2] In the classical and symbolic AI the problem has also been addressed in what is known as the "frame problem" (Pylyshyn 1987).

states. But how does this relate to objects and other things in the real world? According to Maturana & Varela (1992) there is a structural coupling between the things in the world and an organism's active pathways. Wittgenstein (1958) stresses the importance of how language is used to make a relation with language and its meaning. The context of what he called a language game and the purpose of the language game establishes the meaning of it. According to these views, the meaning of symbols is established for a great deal by the interaction of an agent with its environment and is context dependent. A view that has been adopted in the field of pragmatics and situated cognition (Clancey 1997).

In traditional AI and robotics the meaning of symbols was predefined by the programmer of the system. Besides that these systems have no knowledge about the meaning of these symbols, the symbols' meanings were very static and could not deal with different contexts or varying environments. Early computer programs that modelled natural language, notably SHRDLU (Winograd 1972) were completely pre-programmed, and hence could not handle the complete scope of a natural language. It could only handle that part of the language that was pre-programmed. SHRDLU has been programmed *as if* it were a robot with an eye and arm that was operating in a blocks world. Within certain constrictions, SHRDLU could manipulate English input such that it could plan particular goals. However, the symbols that SHRDLU was manipulating had no meaning for the virtual robot. Shakey, a real robot operating in a blocks world, did solve the grounding problem. But Shakey was limited to the knowledge that had been pre-programmed.

Later approaches to solve the grounding problem on real world multi-agent systems involving language have been investigated by Yanco & Stein (1993) and Billard & Hayes (1997). In the work of Yanco and Stein the robots learned to communicate about actions. These actions, however, were pre-programmed and limited, and are therefore limited to the meanings that the robots had. In Billard & Hayes (1997) one robot had pre-programmed meanings of actions, which were represented in a neural network architecture. A student robot had to learn couplings between communicated words and actions it did to follow the first robot. In this work the student robot learned to ground the meaning of its actions symbolically by associating behavioural activation with words. However, the language of the teacher robot was pre-programmed and hence the student could only learn what the teacher knows.

In the work of Billard and Hayes, the meaning is grounded in a situated experiment. So, a part of the meaning is situated in the context in which it is used. However, the learned representation of the meaning is developed through bodily experiences. This is conform with the principle of EMBODIMENT (Lakoff 1987), in which the meaning of something is represented according to bodily experiences. The meaning represented in someone's (or something's) brain depends on previous experiences of interactions with such meanings. The language that emerges is therefore dependent on the body of the system that experiences. This principle is made clear very elegantly by Thomas Nagel in his famous article *What is it like to be a bat?* (Nagel 1974). In this article Nagel argues that it is impossible to understand what a bat is experiencing because it has a different body with different sensing capabilities (a bat uses echolocation to navigate). A bat approaching a wall must experience different meanings (if it has any) than humans would have when approaching a wall. Thus a robot that has a different body than humans will have different meanings. Moreover, different humans have different meaning representations because they encountered different experiences.

This book presents a series of experiments in which two robots try to solve the symbol grounding problem. The experiments are based on a recent approach in AI and the study of language origins, proposed by Luc Steels (1996b). In this new approach, behaviour-based AI (Steels & Brooks 1995) is combined with new computational approaches to the language origins and multi-agent technology. The ideas of Steels have been implemented on real mobile robots so that they can develop a grounded lexicon about objects they can detect in their real world, as reported first in Steels & Vogt 1997. This work differs from the work of Yanco & Stein (1993) and Billard & Hayes (1997) in that no part of the lexicon and its meaning has been programmed. Hence their representation is not limited due to pre-programmed relations.

The next section introduces the symbol grounding problem in more detail. This section first discusses some theoretical background on the meaning of symbols after which some practical issues on symbol grounding are discussed. The experiments are carried out within a broader research on the origins of language, which is presented in §1.2. A little background on human language acquisition is given in §1.3. The research goals of this book are defined in §1.4. The final section of this chapter presents the outline of this book.

1.1 Symbol grounding problem

1.1.1 Language of thought

Already for more than a century philosophers ask themselves how is it possible that we seem to think in terms of symbols which are *about* something that is in the real world. So, if one manipulates symbols as a mental process, one could ask what is the symbol (manipulation) about? Most explanations in the literature are however in terms of symbols that again are about something, as in folk-psychology intentionality is often explained in terms of beliefs, desires etc. For instance, according to Jerry Fodor (1975), every concept is a propositional attitude. Fodor hypothesises a "Language of Thought" to explain why humans tend to think in a *mental* language rather than in natural language alone.

Fodor argues that concepts can be described by symbols that represent propositions towards which attitudes (like beliefs or desires) can be attributed. Fodor calls these symbols "propositional attitudes". If *P* is a proposition, then the phrase "I belief that *P*" is a propositional attitude. According to Fodor, all mental states can be described as propositional attitudes, so a mental state is a belief or desire *about* something. This *something*, however, is a proposition, which according to Fodor is *in the head*. But mental states should be about something that is in the real world. That is the essence of the symbol grounding problem. The propositions are symbol structures that are represented in the brain, sometimes called "mental representations". In addition, the brain consists of rules that describe how these representations can be manipulated. The language of thought, according to Fodor, is constituted by symbols which can be manipulated by applying existing rules. Fodor further argues that the language of thought is innate, and thus resembles Chomsky's universal grammar very well.

Concepts are in this Computational Theory of Mind (as Fodor's theory sometimes is called) constructed from a set of propositions. The language of thought (and with that concepts) can, however, not be learned according to Fodor, who denies:

> [r]oughly, that one can learn a language whose expressive power is greater than that of a language that one already knows. Less roughly, that one can learn a language whose predicates express extensions not expressible by those of a previously available representational system. Still less roughly, that one can learn a language whose predicates express extensions not expressible by predicates of the representational system *whose employment mediates the learning*. (Fodor 1975: 86, Fodor's italics)

According to this, the process of concept learning is the testing of hypotheses that are already available at birth. Likewise, Fodor argues that perception is again the formulating and testing of hypotheses, which are already available to the agent. So, Fodor argues that, since one cannot learn a concept if one does not have the conceptual building blocks of this concept, and since perception needs such building blocks as well, concept learning does not exist and therefore concepts must be innate. This is a remarkable finding, since it roughly implies that all that we know is actual innate knowledge. Fodor called this innate inner language "Mentalese". It must be clear that it is impossible to have such a language. As Patricia S. Churchland puts it:

> [The Mentalese hypothesis] entails the ostensibly new concepts evolving in the course of scientific innovation – concepts such as atom, force field, quark, electrical charge, and gene – are lying ready-made in the language of thought, even of a prehistoric hunter-gatherer… The concepts of modern science are defined in terms of the theories that embed them, not in terms of a set of "primitive conceptual atoms," whatever those may be. (Churchland 1986: 389)

Although the Computational Theory of Mind is controversial, there are still many scientist who adheres to this theory and not the least many AI researchers. This is not surprising, since the theory tries to model cognition computationally, which of course is a nice property since computers are computational devices. It will be shown however that Fodor's Computational Theory of Mind is not necessary for concept and language learning. In particular it will be shown that robots can be developed that can acquire, use and manipulate symbols which are about something that exists in the real world, *and* which are initially not available to the robots.

1.1.2 Understanding Chinese

This so-called symbol grounding problem was made clear excellently by John R. Searle with a gedankenexperiment called the "Chinese Room" (Searle 1980). In this experiment, Searle considers himself standing in a room in which there is a large data bank of Chinese symbols and a set of rules how to manipulate these symbols. Searle, while in the room receives symbols that represent a Chinese expression. Searle, who does not know any Chinese, manipulates these symbols according to the rules such that he can output (other) Chinese symbols as if it was responding correctly in a human like way, but only in Chinese. Moreover, this room passes the Turing test for speaking and understanding Chinese.

Searle claims that this room cannot understand Chinese because he himself does not. Therefore it is impossible to build a computer program that can have mental states and thus being what Searle calls a "strong AI".[3] It is because Searle

[3] It is not the purpose of this book to show that computer programs can have mental states, but to show that symbols in a robot can be about something.

inside the room does not know what the Chinese symbols are about that Searle concludes that the room does not understand Chinese. Searle argues with a logical structure by using some of the following premises (Searle 1984: 39):

(1) Brains cause minds.

(2) Syntax is not sufficient for semantics.

(3) Computer programs are entirely defined by their formal, or syntactical, structure.

(4) Minds have mental contents; specifically, they have semantic contents.

Searle draws his conclusions from these premises in a correct logical deduction, but, for instance, premise (1) seems incomplete. This premise is drawn from Searle's observation that:

> [A]ll mental phenomena [...] are caused by processes going on in the brain. (Searle 1984: 18)

One could argue in favour of this, but Searle does not mention what causes these brain processes. Besides metabolic and other biological processes that are ongoing in the brain, brain processes are caused by sensory stimulation and maybe even by *sensorimotor* activity as a whole. So, at least some mental phenomena are to some extent caused by an agent's[4] interaction with its environment.

Premise (3) states that computer programs are entirely defined by their formal structure, which is correct. Only Searle equates formal with syntactical, which is correct when syntactic means something like "manipulating symbols according to the rules of the structure". The appearance of *symbols* in this definition is crucial, since they are by definition about something. If the symbols in computer programs are about something, the programs are also defined by their semantic structure.

Although Searle does not discuss this, it may be well possible that he makes another big mistake in assuming that he (the central processing unit) is the part where all mental phenomena should come together. An assumption which is debatable (see, e.g. Dennett 1991; Edelman 1992). It is more likely that consciousness is more distributed. But it is not the purpose here to explain consciousness,

[4] I refer to an agent when I am talking about an autonomous agent in general, be it a human, animal, robot or something else.

instead the question is how are symbols about the world. The Chinese Room is presented to make clear what the problem is and how philosophers deal with it.

Obviously Searle's Chinese Room argument found a lot of opposition in the cognitive science community. The critique presented here is in line with what has been called the "system's reply" and to a certain extend the "robot's reply".[5] The system's reply holds that it is not the system who does not understand Chinese, but it is *Searle* who does not. The system as a whole does, since it passed the Turing test.

The robot's reply goes as follows: The Chinese Room as a system does not have any other input than the Chinese symbols. So the system is a very unlikely cognitive agent. Humans have perceptual systems that receive much more information than only linguistic information. Humans perceive visual, tactile, auditory, olfactory and many other information; the Chinese Room does, as it seems, not. So, what if we build a device that has such sensors and like humans has motor capacities? Could such a system with Searle inside understand Chinese?

According to Searle in his answer to both the system's as robot's reply (Searle 1984), his argument still holds. He argues that both the system's reply and the robot's reply do not solve the syntax vs. semantics argument (premise 2). But the mistake that Searle makes is that premise (3) does not hold, thus making premise (2) redundant. Furthermore, in relation to the robot's reply, Searle fails to notice the fact that brain processes are (partly) caused by sensory input and thus mental phenomena are indirectly caused by sensory stimulation.

And even if Searle's arguments are right, in his answer to the robot's reply he fails to understand that a robot is actually a *machine*. It is not just a computer that runs a computer program. And as Searle keeps on stressing:

> "Could a machine think?" Well, in one sense, of course, we are all machines. [...] [In the] sense in which a machine is just *a physical system which is capable of performing certain kinds of operations* in that sense we are all machines, and we can think. So, trivially there are machines that can think. (Searle 1984: 35, my italics)

The reason why the phrase "a physical system which is capable of performing certain kinds of operations" is emphasised is because it is exactly that what a robot is. A robot is more than a computer that runs a computer program.

A last point that is made in this section is that Searle does not speak about development. Could Searle learn to understand Chinese if it was in the room

[5] See for instance the critiques that appeared in the open peer commentary of Searle's 1980 article in the *Behavioural and Brain Sciences*.

from its birth and that he learned to interpret and manipulate the symbols that were presented to him? It is strange that a distinguished philosopher like Searle does not understand that it is possible to develop computer programs which can learn.

The Chinese Room introduced the symbol grounding problem as a thought experiment that inspired Stevan Harnad to define his version of the problem (Harnad 1990). Although controversial, the Chinese Room experiment showed that there are nontrivial problems arising when one builds a cognitive robot that should be able to acquire a meaningful language system. The arguments presented against the Chinese Room are the core of the argument why robots can ground language. As shall become clear, there's more to language than just symbol manipulation according to some rules.

1.1.3 Symbol grounding: philosophical or technical?

Although it might seem very philosophical up to now, this book in no way tries to solve the philosophical problem of what is meaning. In fact there is no attempt being made in solving any philosophical problem. The only thing that is done here is to translate a philosophical problem into a technical problem, which will be tackled in this work. The solution to the technical problem could then be the meat for the philosophers to solve their problem.

Before discussing the symbol grounding problem in more technical detail, it is useful to come up with a working definition of what is meant with a symbol. Harnad's definition of a symbol is very much in line with the standard definition used in artificial intelligence. This definition is primarily based on physical symbol systems introduced by Newell and Simon Newell (1980, 1990). According to Harnad symbols are basically a set of arbitrary tokens that can be manipulated by rules made of tokens; the tokens (either atomic or composite) are "semantically interpretable" (Harnad 1990).

In this book a definition taken from semiotics will be adopted. Following Charles Sanders Peirce and Umberto Eco (1976, 1986) a symbol will be equalled with a SIGN. Using a different, but more familiar terminology than Peirce Nöth (1990), a sign consists of three elements (Chandler 1994):[6]

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

[6] An instructive introduction into the theory of semiotics can be found on the world-wide web (Chandler 1994). The work of Peirce is collected in Peirce 1931–1958.

Interpretant The sense made of the sign.

Object To which the sign refers.

Rather than using Peirce's terms, the terms adopted in this book are FORM for representamen, MEANING for interpretant and REFERENT for object. The adopted terminology is in line with Steels' terminology (Steels 1999). It is also interesting to note that the Peircean sign is not the same as the Saussurean sign (de Saussure 1974). De Saussure does not discuss the notion of the referent. In de Saussure's terminology the form is called "signifier" and the meaning is called the "signified".

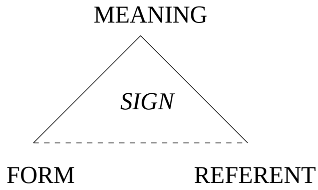


Figure 1.1: A semiotic triangle shows how a referent, meaning and form are related as a sign.

How the three units of the sign are combined is often illustrated with the semiotic triangle (Figure 1.1). According to Peirce, a sign becomes a "symbol" when its form, in relation to its meaning "is arbitrary or purely conventional – so that the relationship must be learnt" (Chandler 1994). The relation can be conventionalised in language. According to the semiotic triangle and the above, a symbol is *per definition* grounded.

In the experiments reported in this book, the robots try to develop a shared and grounded lexicon about the real world objects they can detect. They do so by communicating a name of the categorisation of a real world object. In line with the theory of semiotics, the following definitions are made:

Referent The referent is the real world object that is subject of the communication.

Meaning The meaning is the categorisation that is made of the real world object and that is used in the communication.

Form The form is the name that is communicated. In principle its shape is arbitrary, but in a shared lexicon it is conventionalised through language use.

Symbol A symbol is the relation between the referent, the meaning and the form as illustrated in the semiotic triangle.

This brings us to the technically hard part of the symbol grounding problem that remains to be solved: How can an agent construct the relations between a form, meaning and referent? In his article Harnad (1990) recognises three main tasks of grounding symbols:

Iconisation Analogue signals need to be transformed to ICONIC REPRESENTATION (or icons).[7]

Discrimination "[The ability] to judge whether two inputs are the same or different, and, if different, how different they are." Note that in Harnad's article, discrimination is already pursued at the perceptual level. In this book, discrimination is done at the categorical level.

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)

So, what is the problem? Analogue signals can be iconised (or recorded) rather simple with meaningless sub-symbolic structures. The ability to discriminate is easy to implement just by comparing two different sensory inputs. The ability to identify requires to find *invariant* properties of objects, events and state of affairs. Since finding distinctions is rather easy, the big problem in grounding actually reduces to identifying

> invariant features of the sensory projection that will reliably distinguish a member of a category from any non-members with which it could be confused. (Harnad 1990)

Although people might disagree, for the roboticists this is not more than a technical problem. The question is whether or not there exist real invariant features of a category in the world. This probably could be doubted quite seriously (see e.g.Harnad 1993). For the time being it is assumed that there are invariant properties in the world and it will be shown that these invariants can be found if an embodied agent is equipped with the right physical body and control. The latter inference is in line with the PHYSICAL GROUNDING HYPOTHESIS (Brooks 1990), which will be discussed below.

[7] The terms "icon" and "iconisation", as they are used by Harnad and will be adopted here, should not be confused with Peirce's notion of these terms.

Stevan Harnad proposes that the SGP for a robot could possibly be solved by invoking (hybrid) connectionist models with a serious interface to the outside world in the form of TRANSDUCERS (or sensors) (Harnad 1993). Harnad, however admits that the symbol grounding problem also might be solved with other than connectionist architectures.

1.1.4 Grounding symbols in language

In line with the work of Luc Steels the symbols are grounded in language (see e.g. Steels 1997b, 1999). Why grounding symbols in language directly and not ground the symbols first and develop a shared lexicon afterwards? Associating the grounded symbols with a lexicon is then a simple task, (see e.g. Oliphant 1997; Steels 1996b). However, as Wittgenstein (1958) pointed out, the meaning of something depends on how it is used in language. It is situated in the environment of an agent and depends on the bodily experience of it. Language use gives feedback on the appropriateness of the sense that is made of a referent. So, language gives rise to the construction of meanings and the construction of meaning gives rise to language development. Hence, meaning co-evolves with language.

That this approach seems natural can be illustrated with Roussau's paradox. Although for communication, categorisation of reality needs to be similar to different language users, different languages do not always employ the same categorisations. For instance, there are different referential frames to categorise spatial relations in different language communities. In English there are spatial relations like *left*, *right*, *front* and *back* relative to some axis. However in Tzetal, a Mayan language, this frame of reference is not used. The Tzetal speakers live in an area with mountains and their frame of reference is absolute in relation to the mountain they are on. The spatial relations in this language can be translated with 'uphill', 'downhill' and 'across'. If something is higher up the mountain in relation to the speaker, they can say "this something is uphill of me".

So, if a novel language user enters a language society, how would it know how to categorise such a spatial relation? To know this, the new language user has to learn how to categorise the reality in relation to the language that is used by the particular language society. Therefore it is thought to be necessary to ground meaning in language. How lexicon development interacts with the development of meaning will become clearer in the remainder of this book.

1.1.5 Physical grounding hypothesis

Another approach to grounding is physical grounding. In his article *Elephants Don't Play Chess*, Rodney Brooks (1990) proposed the physical grounding hypothesis as an additional constraint to the physical symbol system hypothesis.

> The physical grounding hypothesis states that to build a system that is intelligent it is necessary to have its representations grounded in the physical world. (Brooks 1990)

The advantage of the physical grounding hypothesis over physical symbol system hypothesis is that the system (or agent) is directly coupled to the real world through its set of sensors and actuators.

> Typed input and output are no longer of interest. They are not physically grounded. (Brooks 1990)

In Brooks' approach symbols are not a necessary condition for intelligent behaviour anymore (Brooks 1990, 1991). Intelligent behaviour can emerge from a set of simple couplings of an agent's sensors with its actuators,[8] as is also shown in e.g. Steels & Brooks 1995, Steels 1994c, and Steels 1996a. An example is "wall following". Suppose a robot has two simple behaviours: (1) the tendency to move towards the wall and (2) the tendency to move away from the wall. If the robot incorporates both behaviours at once, then the resulting *emergent* behaviour is wall following. Note that agents designed from this perspective have no cognitive abilities. They are reactive agents, like e.g. ants are, rather than cognitive agents that can manipulate symbolic meanings.

The argument that Brooks uses to propose the physical grounding hypothesis is that

> [evolution] suggests that problem solving behaviour, language, expert knowledge and application, and reason, are all rather simple once the essence of being and reacting are available. That essence is the ability to move around in a dynamic environment, sensing the surroundings to a degree sufficient to achieve the necessary maintenance of life and reproduction. (Brooks 1990)

This rapid evolution is illustrated in Figure 1.2. Brooks also uses this argument of the rapid evolution of human intelligence as opposed to the slow evolution of life on earth in relation to symbols.

[8] Note that Brooks' approach does not necessarily invoke connectionist models.

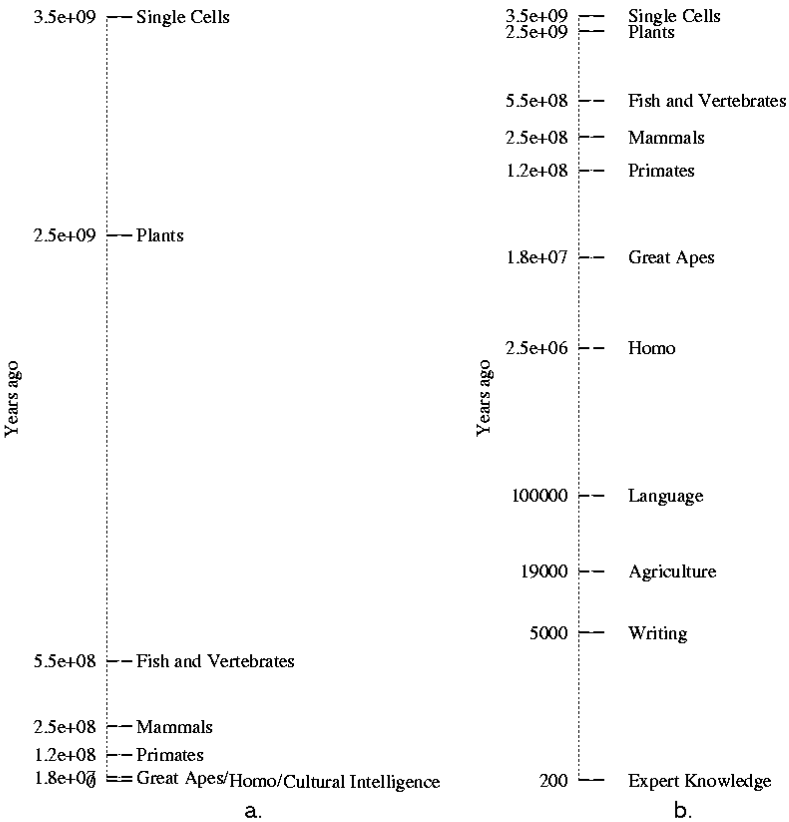


Figure 1.2: (a) The evolutionary time-scale of life and cognitive abilities on earth. After the entrance of the great apes, evolution of man went so fast that it cannot be shown on the same plot, unless it is shown in logarithmic scale (see b). It appears from the plot that cultural evolution works much faster than biological evolution. Time-scale is adapted from Brooks (1990).

> [O]nce evolution had symbols and representations things started moving rather quickly. Thus symbols are the key invention … Without a carefully built physical grounding any symbolic representation will be mismatched to its sensors and actuators. (Brooks 1990)

To explore the physical grounding hypothesis, Brooks and his co-workers at the MIT AI Lab developed a software architecture called the SUBSUMPTION ARCHITECTURE (Brooks 1986). This architecture is designed to connect a robot's sensors to its actuators so that it "embeds the robot correctly in the world" (Brooks 1990). The point made by Brooks is that intelligence can emerge from an agent's physical interactions with the world. So, the robot that needs to be built should be both embodied and situated. The approach proposed by Brooks is also known as "behaviour-based AI".

1.1.6 Physical symbol grounding

The physical grounding hypothesis (Brooks 1990) states that intelligent agents should be grounded in the real world. However, it also states that the intelligence need not to be represented with symbols. According to the physical symbol system hypothesis the thus physically grounded agents are no cognitive agents. The physical symbol system hypothesis (Newell 1980) states that cognitive agents are *physical symbol systems* with the following features (Newell 1990: 77):

Memory
- Contains structures that contain symbol tokens
- Independently modifiable at some grain size

Symbols
- Patterns that provide access to distal structures
- A symbol token is the occurrence of a pattern in a structure

Operations Processes that take symbol structures as input and produce symbol structures as output

Interpretation Processes that take symbol structures as input and execute operations

Capacities
- Sufficient memory and symbols
- Complete compositionality
- Complete interpretability

Clearly, an agent that uses language is a physical symbol system. It should have a memory to store an ontology and lexicon. It has symbols. The agent makes operations on the symbols and interprets them. Furthermore, it should have the capacity to do so. In this sense, the robots of this book are physical symbol systems.

A physical symbol system somehow has to represent the symbols. Hence the physical grounding hypothesis is not the best candidate. But since the definition of a symbol adopted in this book has an explicit relation to the referent, the complete symbol cannot be represented inside a robot. The only parts of the symbols that can be represented are the meaning and the form. Like in the physical grounding hypothesis, a part of the agent's knowledge is in the world. The problem is: how can the robot *ground* the relation between internal representations and the referent? Although Newell (1990) recognises the problem, he does not investigate a solution to it.

This problem is what Harnad (1990) called the symbol grounding problem. Because there is a strong relation between the physical grounding hypothesis (that the robot has its knowledge grounded in the real world) and the physical symbol system hypothesis (that cognitive agents are physical symbol systems) it is useful to rename the symbol grounding problem in the PHYSICAL SYMBOL GROUNDING PROBLEM.

The physical symbol grounding problem is very much related to the FRAME PROBLEM (Pylyshyn 1987). The frame problem deals with the question how a robot can represent things of the dynamically changing real world and operate in it. In order to do so, the robot needs to solve the symbol grounding problem.

As mentioned, this is a very hard problem. Why is the physical symbol grounding problem so hard? When sensing something in the real world under different circumstances, the physical sensing of this something is different as well. Humans are nevertheless very good at identifying this something under these different circumstances. For robots this is different. The one-to-many mappings of this something unto the different perceptions needs to be interpreted so that there is a more or less one-to-one mapping between this something and a symbol, i.e. the identification needs to be invariant. Studies have shown that this is an extremely difficult task for robots.

Already numerous systems have been physically grounded (see e.g. Brooks 1990; Steels 1994c; Barnes et al. 1997; Kröse et al. 1999; Tani & Nolfi 1998; Berthouze & Kuniyoshi 1998; Pfeifer & Scheier 1999; Billard & Hayes 1997; Rosenstein & Cohen 1998a; Yanco & Stein 1993 and many more). However, a lot of these systems do not ground symbolic structures because they have no form (or arbitrary label) attached. These applications ground "simple" physical behaviours in the Brooksean sense. Only a few physically grounded systems mentioned above grounded symbolic structures, for instance in the case of Yanco & Stein (1993), Billard & Hayes (1997), and Rosenstein & Cohen (1998a).

Yanco & Stein developed a troupe of two robots that could learn to associate certain actions with a pre-defined set of words. One robot would decide what action is to be taken and communicates a relating signal to the other robot. The learning strategy they used was reinforcement learning where the feedback in their task completion was provided by a human instructor. If both robots performed the same task, a positive reinforcement was given, and when both robots did not, the feedback consisted of a negative reinforcement.

The research was primarily focussed on the learning of associations between word and meaning on physical robots. No real solution was attempted to solve the grounding problem and only a limited set of word-meaning associations were pre-defined. In addition, the robots learned by means of supervised learning with a human instructor. Yanco & Stein showed, however, that a group of robots could converge in learning such a communication system.

In Billard & Hayes (1997) two robots grounded a language by means of imitation. The experiments consisted of a teacher robot, which had a pre-defined communication system, and a student robot, which had to learn the teacher's language by following it. The learning mechanism was provided by an associative neural network architecture called DRAMA. This neural network learned associations between communication signals and sensorimotor couplings. Feedback was provided by the student's evaluation if it was still following the teacher.

So, the language was grounded by the student using this neural network architecture, which is derived from Wilshaw networks. Associations for the teacher robot were pre-defined in their couplings and weights. The student could learn a limited amount of associations of actions and perceptions very rapidly (Billard 1998).

Rosenstein & Cohen (1998a) developed a robot that could ground time series by using the so-called method of delays, which is drawn from the theory of non-linear dynamics. The time series that the robots produce by interacting in their environment are categorised by comparing their delay vectors, which is a low-dimensional reconstruction of the original time series, with a set of prototypes. The concepts the robots thus ground could be used for grounding word-meanings (Rosenstein & Cohen 1998b).

The method proposed by Rosenstein & Cohen has been incorporated in a language experiment where two robots play FOLLOW ME GAMES to construct an ontology and lexicon to communicate their actions (Vogt 1999, 2000b). This was a preliminary experiment, but the results appear to be promising.

A similar experiment on language acquisition on mobile robots has been done by the same group of Rosenstein & Cohen at the University of Massachusetts (Oates et al. 1999). The time series of a robot's actions are categorised using a clustering method for distinctions (Oates 1999). Similarities between observed time series and prototypes are calculated using dynamical time warping. The thus conceptualised time series are then analysed in terms of human linguistic interactions, who describe what they see when watching a movie of the robot operating (Oates et al. 1999).

Other research propose simulated solutions to the symbol grounding problem, notably Cangelosi & Parisi (1998) and Greco et al. (1998). In his work Angelo Cangelosi created an ecology of edible and non-edible mushrooms. Agents that are provided with neural networks learn to categorise the mushrooms from "visible" features into the categories of edible and non-edible mushrooms.

A problem with simulations of grounding is that the problem cannot be solved in principle, because the agents that "ground" symbols do not do so in the *real world*. However, these simulations are useful in that they can learn us more about how categories and words could be grounded. One of the important findings of Cangelosi's research is that communication helps the agents to improve their categorisation abilities (Cangelosi et al. 2000).

Additional work can be found in *The grounding of word meaning: Data and models* (Gasser 1998), the proceedings of a joint workshop on the grounding of word meaning of the AAAI and Cognitive Science Society. In these proceedings, grounding of word meaning is discussed among computer scientists, linguistics and psychologists.

So, the problem that is tried to be solved in this book is what might be called the physical symbol grounding problem. This problem shall not be treated philosophically but technically. It will be shown that the quality of the physically grounded interaction is essential to the quality of the symbol grounding. This is in line with Brooks' observation that a.o. language is

> rather easy once the essence of being and reacting are available.
> (Brooks 1990)

Now that it is clear that the physical symbol grounding problem in this work is considered to be a technical problem, the question rises how it is solved. In 1996, Luc Steels published a series of papers in which some simple mechanisms were introduced by which autonomous agents could develop a "grounded" lexicon

(Steels 1996b,c,d,e, for an overview see Steels 1997c). Before this work is discussed, a brief introduction in the origins of language is given.

1.2 Language origins

Why is it that humans have language and other animals cannot? Until not very long ago, language has been ascribed as a creation of God. Modern science, however, assumes that life as it currently exists has evolved gradually. Most influencing in this view has been the book of Charles Darwin *The origins of species* (1968). In the beginning of the existence of life on earth, humans were not yet present. Modern humans evolved only about 100,000 to 200,000 years ago. With the arrival of homo sapiens, language is thought to have emerged. So, although life on earth is present for about 3.5 billion years, humans are on earth only a fraction of this time.

Language is exclusive to humans. Although other animals have communication systems, they do not use a complex communication system like humans do. At some point in evolution, humans must have developed language capabilities. These capabilities did not evolve in other animals. It is likely that these capabilities evolved biologically and are present in the human brain. But, what are these capabilities? They are likely to be the initial conditions from which language emerged. Some of them might have co-evolved with language, but most of them were likely to be present before language originated. This is likely because biological evolution is very slow, whereas language on the evolutionary time scale evolved very fast.

The capabilities include at least the following things: (1) The ability to associate meanings of things that exist in the world with arbitrary word-forms. (2) The ability to communicate these meaningful symbols to other language users. (3) The ability to vocalise such symbols. (4) The ability to map auditory stimuli of such vocalisations to the symbols. And (5) the ability to use grammatical structures. These abilities must have evolved somehow, because they are principle features of human language. There are probably more capabilities, but they serve to accomplish the five capabilities mentioned. In line with the symbol grounding problem this book concentrates on the first two principle capabilities.

Until the 1950s there was very little research going on about the evolution and origins of language. Since Noam Chomsky wrote his influential paper on syntactic structures (Chomsky 1956), linguistic research and research on the evolution of language boomed. It took until 1976 for the first conference on the origins and evolution of language to be held (Harnad et al. 1976). Most papers of this conference involved empirical research on ape studies, studies on gestural communication and theoretical and philosophical studies. Until very recently, many studies had a high level of speculation and some strange theories were proposed. For an overview of theories that were proposed on the origins and evolution of language until 1996, see Aitchison (1996).

1.2.1 Computational approaches to language evolution

With the rise of advanced computer techniques in artificial intelligence (AI) and Artificial Life (ALife), it became possible to study the origins and evolution of language computationally. In the 1990s many such studies were done. It is probably impossible to say with this approach exactly how language originated, but the same is probably true for all other investigations. The only contribution computer techniques can bring is a possible scenario of language evolution. Possible initial conditions and hypotheses can be validated using computer techniques, which may shed light on how language may have emerged. Furthermore, one can rule out some theories, because they do not work on a computer.

Many early (and still very popular) scenarios were investigated based on Chomsky's theory about a UNIVERSAL GRAMMAR, which are supposed to be innate.[9] According to Chomsky, the innate universal grammar codes "principles" and "parameters" that enable infants to learn any language. The principles encode universals of languages as they are found in the world. Depending on the language environment of a language learner, the parameters are set, which allows the principles of a particular language to become learnable. So, the quest for computer scientist is to use evolutionary computation techniques to come up with a genetic code of the universal grammar. That this is difficult can already be inferred from the fact that up to now not one non-trivial universal tendency of language is found which is valid for every language.

In the early nineties a different approach gained popularity. This approach is based on the paradigm that language is a complex dynamical adaptive system. Here it is believed that universal tendencies of language are learned and evolve culturally.

Agent based simulations were constructed in which the agents tried to develop (usually an aspect of) language. The agents are made adaptive using techniques taken from AI and adaptive behaviour (or ALife). The main approach taken is a bottom-up approach. In contrast to the top-down approach, where the intelli-

[9] One of the reasons why Chomsky's theory is still very popular amongst computational linguistics is that the theory has a computational approach.

gence is modelled and implemented in rules, the bottom-up approach starts with implementing simple *sensorimotor* interfaces and learning rules, and tries to increase the complexity of the intelligent agent step by step.

Various models have been built by a variety of computer scientists and computational linguists to investigate the evolution of language and communication (e.g. Cangelosi & Parisi 1998; Kirby & Hurford 1997; MacLennan 1991; Oliphant 1997; Werner & Dyer 1991). It goes beyond the scope of this book to discuss all this research, but there is one research that is of particular interest for this book, namely the work of Mike Oliphant (1997, 1998, 2000).

Oliphant simulates the learning of a symbolic communication system in which a fixed number of signals are matched with a fixed number of meanings. The number of signals that can be learned is equal to the number of meanings. Such a coherent mapping is called a Saussurean sign (de Saussure 1974) and is the idealisation of language. The learning paradigm of Oliphant is an *observational* one and he uses an associative network incorporating Hebbian learning. With observational is meant that the agents during a language game have access to both the linguistic signal and its meaning.

As long as the communicating agents are aware of the meaning they are signalling, the Saussurean sign can be learned (Oliphant 1997, 2000). The awareness of the meaning meant by the signal should be acquired by observation in the environment. Oliphant further argues that reinforcement types of learning as used by Yanco & Stein (1993) and Steels (1996b) are not necessary and unlikely (see also the discussion about the no negative feedback evidence in §1.3). But he does not say they are not a possible source of language learning (Oliphant 2000).

The claim Oliphant makes has implications on why only humans can learn language. According to Oliphant (1998), animals have difficulty in matching a signal to a meaning when it is not an innate feature of the animal. Although this is arguable (Oliphant refers here to e.g. Gardner & Gardner 1969 and Premack 1971), he observes the fact that in these animal learning the communication is explicitly taught by the researchers.

1.2.2 Steels' approach

This adaptive behaviour based approach has also been adopted by Luc Steels (e.g. Steels 1996b,c, 1997c). The work of Steels is based on the notion of LANGUAGE GAMES (Wittgenstein 1958). In language games agents construct a lexicon through cultural interaction, individual adaptation and self-organisation. The view of Wittgenstein is adopted that language gets its meaning through its use and should be investigated accordingly. The research presented in this book is in line with the work done by Luc Steels. This research is part of the ongoing research done at the Sony Computer Science Laboratory in Paris and at the Artificial Intelligence Laboratory of the Free University of Brussels (VUB), both directed by Luc Steels.

The investigation in Paris and Brussels is done on both simulations and grounded robots. It focuses on the origins of sound systems, in particular in the field of phonetics (De Boer 1997, 1999; Oudeyer 1999), the origins of meaning (Steels 1996c; Steels & Vogt 1997; De Jong & Vogt 1998; Vogt 1998b; De Jong & Steels 1999), the emergence of lexicons (Steels 1996b; Steels & Kaplan 1998; Kaplan 2000; Vogt 1998a; Van Looveren 1999), the origins of communication (De Jong 1999b, 2000) and the emergence of syntax (Steels 2000). Within these subjects various aspects of language like stochasticity (Steels & Kaplan 1998; Kaplan 2000), dynamic language change (Steels 1997a; Steels & McIntyre 1999; De Boer & Vogt 1999), multi-word utterances (Van Looveren 1999), situation concepts (De Jong 1999a) and grounding (Belpaeme et al. 1998; Steels & Vogt 1997; Steels 1999; Kaplan 2000) are investigated.

Bart de Boer of the VUB AI Lab has shown how agents can develop a human-like vowel system through self-organisation (De Boer 1997, 1999). These agents were modelled with a human like vocal tract and auditory system. Through cultural interactions and imitations the agents learned vowel systems as they are found prominently among human languages.

First in simulations (Steels 1996b,c) and later in grounded experiments on mobile robots (Steels & Vogt 1997; Vogt 1998b,a; De Jong & Vogt 1998) and on the Talking Heads (Belpaeme et al. 1998; Kaplan 2000; Steels 1999) the emergence of meaning and lexicons have been investigated. Since the mobile robots experiment is the issue of the current book, only the other work will be discussed briefly here.

The simulations began fairly simple by assuming a relative perfect world (Steels 1996b,c). Software agents played naming and discrimination games to create lexicons and meaning. The lexicons were formed to name predefined meanings and the meanings were created to discriminate predefined visual features. In later experiments more complexity was added to the experiments. From findings of the mobile robots experiments (Vogt 1998b) it was found that the ideal assumptions of the naming game, for instance, considering the topic to be known by the hearer, were not satisfied. Therefore a more sophisticated naming game was developed that could handle noise of the environment (Steels & Kaplan 1998).

For coupling the discrimination game to the naming game, which first has been done in Steels & Vogt (1997), a new software environment was created: the

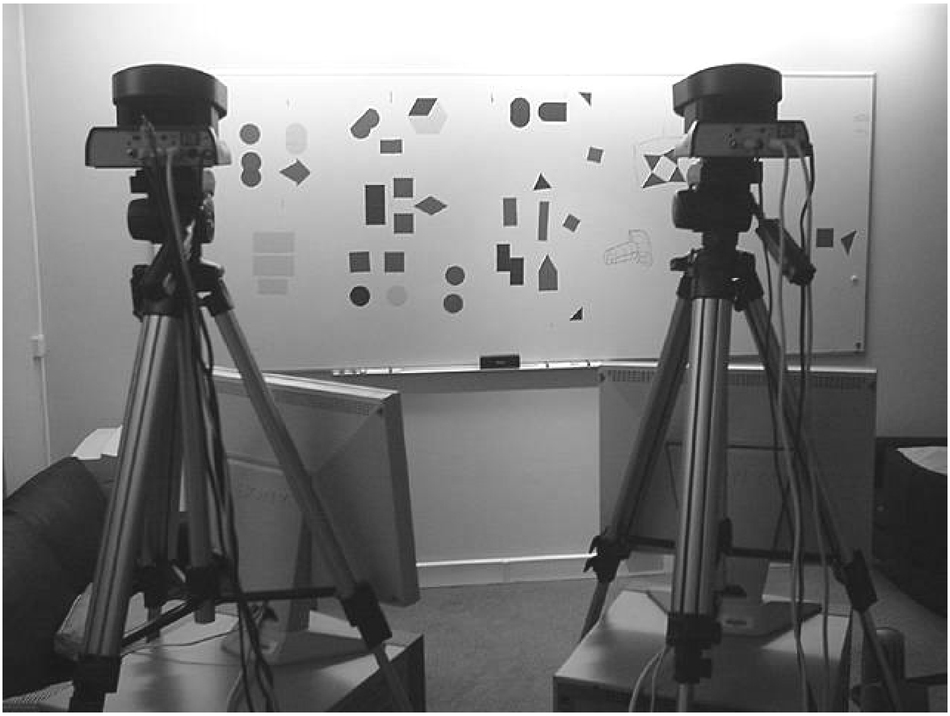


Figure 1.3: The Talking Heads as it is installed at Sony csl Paris.

geom world (Steels 1999). The geom world consisted of an environment in which geometric figures could be conceptualised through the discrimination game. The resulting representations could then be lexicalized using the naming game. The Talking Heads are also situated in a world of geometrical shapes that are pasted on a white board the cameras of the heads look at (Figure 1.3).

The Talking Heads consist of a couple of installations that are distributed around the world. Installations currently exist in Paris at the Sony csl, in Brussels at the vub ai Lab, in Amsterdam at the Intelligent Autonomous Systems laboratory of the University of Amsterdam. Temporal installations have been operational in Antwerp, Tokyo, Laussane, Cambridge, London and at another site in Paris. Agents can travel the world through the internet and embody themselves into a Talking Head. A Talking Head is a pan-tilt camera connected to a computer. The Talking Heads play language games with the cognitive capacities and memories that each agent has or has acquired. The language games are similar to the ones that are presented in the subsequent chapters. The main difference is that the

Talking Heads do not move from their place, which the mobile robots do. The Talking Heads have cameras as their primary sensory apparatus and there are some slight differences in the cognitive capabilities as will become clear in the rest of this book.

All these experiments show similar results. Label-representation (or form-meaning) pairs can be grounded in sensorimotor control, for which (cultural) interactions, individual adaptation and self-organisation are the key mechanisms. A similar conclusion will be drawn at the end of this book. The results of the experiments on mobile robots will be compared with the Talking Heads as reported mainly in Steels 1999. Other findings based on the different variations of the model, which inspects the different influences of the model will be compared with the PhD thesis of Frédéric Kaplan of Sony CSL in Paris (Kaplan 2000).[10]

A last set of experiments that will be brought to the reader's attention is the work done by Edwin de Jong of the VUB AI Lab. De Jong has done an interesting experiment in which he showed that the communication systems that emerged under the conditions by which language research is done in Paris and Brussels are indeed complex dynamical systems (De Jong 2000). The communication systems of his own experiments all evolved towards an attractor and he showed empirically that the system was a complex dynamical system.

Using simulations, De Jong studied the evolution of communication in experiments in which agents construct a communication system about situation concepts (De Jong 1999a). In his simulation, a population of agents were in some situation that required a response in the form of an action. I.e. if one of the agents observed something (e.g. a predator), all the agents needed to go in some save state. De Jong investigated if the agents could benefit from communication, by allowing the agents to develop a shared lexicon that is grounded in this simulated world. The agents were given a mechanism to evaluate, based on their previous experiences, whether to trust on their observations or on some communicated signal. The signal is communicated by one of the agents that had observed something.

While doing so, the agents developed an ontology of situation concepts and a lexicon in basically the same way as in the work of Luc Steels. This means that the robots play discrimination games to build up the ontology and naming games to develop a language. A major difference is that the experiments are situated in a task oriented approach. The agents have to respond correctly to some situation.

[10] Currently Frédéric Kaplan is working on human-machine interaction on the AIBO robot that looks like a dog and which has been developed by Sony CSL in Tokyo. Naturally, the AIBO learns language according to the same principles advocated by our labs.

To do so, the agents can evaluate their success based on the appropriateness of their actions. As will be discussed in Chapter ??, De Jong used a different method for categorisation, called the ADAPTIVE SUBSPACE METHOD (De Jong & Vogt 1998).

One interesting finding of De Jong was that it is not necessary that agents use feedback on the outcome of their linguistic interactions to construct a coherent lexicon, provided that the robots have access to the meaning of such an interaction and lateral inhibition was assured. Hence this confirms the findings of Mike Oliphant (1998). Questions about the feedback on language games are also issued in the field of human language acquisition.

1.3 Language acquisition

Although children learn an existing language, lessons from the language acquisition field may help to understand how humans acquire symbols. This knowledge may in turn help to build a physically grounded symbol system. In the experiments presented in the forthcoming, the robots develop only a lexicon by producing and understanding one word utterances. In the literature of language acquisition, this period is called EARLY LEXICON DEVELOPMENT. Infants need to learn how words are associated with meanings. How do they do that?

In early lexicon development it is important to identify what cues an infant receives of the language it is learning. These cues not only focus on the linguistic information, but also on the extra-linguistic information. It is not hard to imagine that when no linguistic knowledge is available about a language, it seems impossible to learn such a language without extra-linguistic cues such as pointing or feedback about whether one understands a word correctly. (Psycho-)linguists have not agreed upon what information is available to a child and to what extend.

The POVERTY OF THE STIMULUS argument led Chomsky to propose his linguistic theory. Although an adult language user can express an unlimited number of sentences, a language learner receives a limited amount of linguistic information to master the language. With this argument Chomsky concluded that linguistic structures must be innate. But perhaps there are other mechanisms that allow humans to learn language. Some might be learned and some might be innate.

A problem that occupies the nativist linguists is the so-called NO NEGATIVE FEEDBACK EVIDENCE (e.g. Bowerman 1988). The problem is that in the innate approach language can only be learned when both positive and negative feedback on language is available to a language learner. However, psychological research has shown that no negative feedback is provided by adult language users (Braine

1971). Demetras and colleagues, however showed that there is more negative feedback provided than assumed (Demetras et al. 1986). In addition, it is perhaps underestimated how much feedback a child can evaluate itself from its environment. Furthermore, feedback is thought to be an important principle in cognitive development (see e.g. Clancey 1997).

One alternative for the feedback, which is assumed to be provided after the linguistic act, is the establishment of joint attention *prior* to the linguistic communication. Do children really receive such input? Early studies of Tomasello showed that children can learn better when joint attention is established, as long as this is done spontaneously by the child (Tomasello et al. 1986, cited in Barrett 1995). Explicit drawing of attention seemed to have a negative side effect. Although it has been assumed that pointing was a frequently used method to draw a child's attention, later studies have argued against such this assumption. Tomasello reported in a later studies that pointing is not necessary for learning language, *provided* there is explicit feedback (Tomasello & Barton 1994).

In this article, Tomasello and Barton report on experiments where children learn novel words under two different conditions. In one condition, children do not receive extra-linguistic cues when the word-form is presented. There is a so-called "nonostensive" context. When at a later moment the corresponding referent is shown, a positive feedback is given if the child correctly relates the referent with given word-form. If the child relates the word-form to an incorrect referent, negative feedback is given. In the second condition, joint attention is established simultaneous with the presentation of the word-form. In this condition the child receives a so-called "ostensive" context. Tomasello & Barton (1994) showed in their experiments that children could equally well learn novel word-meaning relations in both condition.

Yet another strategy is proposed by Eve Clark (1993). She argues that children can fill in knowledge gaps when receiving novel language, provided the context was known.

So, a lot of strategies appear to be available to a language learner, and there may be more. It is not unlikely that a combination of the available strategies is used; perhaps some more frequent than others. A natural question rises: Which strategies work and which do not? In this book experiments are presented that investigate both the role of feedback and joint attention.

1.4 Setting up the goals

This book presents the development and results of a series of experiments where two mobile robots develop a grounded lexicon. The experiments are based on language games that have first been implemented on mobile robots in Steels & Vogt (1997) and Vogt (1997). The goal of the language games is to construct an ontology and lexicon about the objects the robots can detect in their environment.

The sensory equipment with which the robots detect their world is kept simple, namely sensors that can only detect light intensities. One of the goals was to develop the experiments without changing the simplicity of the robots very much and to keep the control architecture within the behaviour-based design. Luc Steels (1996b) hypothesises three basic mechanisms for language evolution, which have been introduced above: individual adaptation, cultural evolution and self-organisation.

In a language game, robots produce a sensorimotor behaviour to perceive their environment. The environment consists of a set of light sources, which are distinguishable in height. The raw sensory data that results from this sensing is segmented, yielding a set of segments of which each segment relates to the detection of a light source. These segments can be described by features, which are categorised by the individual robots. The categorisation is processed by so-called DISCRIMINATION GAMES (Steels 1996c). In this process the robots try to develop categories that discriminates one segment from another. The lexicon is formed based on an interaction and adaptation strategy modelled in what has been called NAMING GAMES (Steels 1996b). In a naming game one robot has the role of a speaker and the other robot has the role of the hearer. The speaker tries to name the categorisation (or meaning) of a segment it has chosen to be the topic. The hearer tries to identify the topic using both linguistic and extra-linguistic information when available.

The language game is adaptive in that the robots can adapt either their ontology or lexicon when they fail to categorise of name the topic. This way they may be successful in future games. In addition, the robots can adapt association strengths that they use to select elements of their ontology or lexicon. The selection principle is very much based on natural selection as proposed by Charles Darwin (1968), but the evolution is not spread over generations of organisms, but over "generations" of language games. The principle is that the most effective elements are selected more and ineffective ones are selected less frequently, or even not at all. This way the most effective elements of the language are spread in the language community, thus leading to a cultural evolution.

The idea of cultural evolution has best been described by Richard Dawkins in his book *The Selfish Gene* (1976). In this book Dawkins proposes the notion of MEMES. Memes are elements that carry the notion of ideas, like the idea of a wheel. Like genes, memes are generated as varieties of previous ideas and possibly as complete new ideas. The memes are spread in the society by cultural interactions. The evolution of memes is similar to that of genetic evolution and good memes survive, whereas bad memes do not. However, the cultural evolution is much faster than biological evolution and several generations of memes can occur in a society within the lifetime of an organism. When changing the notion of memes into language elements, a cultural evolution of language arrives. The emergence of language through cultural evolution is based on the same principle as biological evolution, namely self-organisation.

Three main research questions are raised in this book:

1. Can the symbol grounding problem be solved with these robots by constructing a lexicon through individual adaptation, (cultural) interaction and self-organisation? And if so, how is this accomplished?
2. What are the important types of extra-linguistic information that agents should share when developing a coherent communication system?
3. What is the influence of the physical conditions and interaction of the robots on developing a grounded lexicon?

The first question is an obvious one and can be answered with yes, but to a certain extend. As argued in §1.1.3, the symbol grounding problem is solved when the robots are able to construct a semiotic sign of which the form is either arbitrary or conventionalised. Since the robots try to ground a shared lexicon, the form has to be conventionalised. Therefore the robots solve the symbol grounding problem when they successfully play a language game. I.e. when both robots are able to identify a symbol with the same form that stands for the same referent.

Throughout the book the model that accomplishes the task is presented and revised to come up with two language game models that work best. Although the basics of the models, namely the discrimination- and naming game are very simple, the implementation on these simple robots has proven to be extremely difficult. Not all the designer's frustrations are made explicit in this book, but working with LEGO robots and "home-made" sensorimotor boards made life not easier. In order to concentrate on the grounding problem, some practical assumptions have been made leaving some unsolved technical problems.

The two models that are proposed at the end of the experimental results show different interaction strategies that answer the second question. Both feedback and joint-attention are important types of extra-linguistic information necessary for agents to develop a lexicon, although not necessarily used simultaneously. How feedback and joint attention can be established is left as an open question. Technical limitations drove to leave this question open as one of the remaining frustrations. Some of these limitations are the same that introduced the assumptions that have been made.

Although more difficult to show, the quality of physical interactions have an important influence on the robots' ability to ground a lexicon. When the robots are not well adapted to their environment (or vice versa) no meaningful lexicon can emerge. In addition, when the robots can co-ordinate their actions well to accomplish a certain (sub)task, they will be better in grounding a lexicon than when the co-ordination is weak.

1.5 Contributions

How does this book contribute to the field of artificial intelligence and cognitive science? The main contributions made in this book that there is an autonomous system that is grounded in the real world of which no parts of the ontology or lexicon is pre-defined. The categorisation is organised hierarchically by prototypical categories. In addition, the book investigates different types of extra-linguistic information that the robots can use to develop a shared lexicon. No single aspect is more or less unique. However, the combination of some aspects is.

Table 1.1 shows the contributions of research that is most relevant to this work. The table lists some aspects that the various researchers have contributed in their work. The aspects that are listed are thought to be most relevant to this work. Note that with Steels' work the Talking Heads experiments are meant. In the discussion at the end of this book, a more detailed comparison with the Talking Heads is made.

Of the related work, the work of Cangelosi & Parisi (1998), De Jong (2000), and Oliphant (1997) is not grounded in the real world. The work of Cangelosi et al. and De Jong is grounded only in simulations. This makes the grounding process relatively easy, because it avoids the problems that come about when categorising the real world. Oliphant does not ground meaning at all. The work of this book is grounded in the real world.

Some researchers, notably Billard & Hayes (1997), Cangelosi & Parisi (1998), and Yanco & Stein (1993), pre-define the language. I.e. they define how a word-

Table 1.1: Various aspects investigated by different researchers. Each column of the table is reserved for a particular research. The related work in this table is from (the group of): Billard (B), Cangelosi (C), De Jong (D), Oliphant (O), Rosenstein (R), Steels (S), Vogt (V), and Yanco & Stein (Y). The other symbols in the table stand for "yes" (+), "no" (–) and "not applicable" (·).

Aspect	B	C	D	O	R	S	V	Y
Grounded in real world	+	–	–	–	+	+	+	+
Language pre–defined	+	+	–	–	·	–	–	+
Meaning pre–defined	+/–	–	–	+	–	–	–	+
Prototypical categories	–	–	–	·	+	–	+	–
Hierarchical layering of categories	–	–	+	·	–	+	+	–
Nr. of meanings given	+/–	–	–	+	–	–	–	+
Nr. of forms given	+	+	–	+	·	–	–	+
Nr. of agents	2	≥ 2	≥ 2	≥ 2	1	≥ 2	2	≥ 2
Calibrated world	–	–	–	·	–	+	–	–
Mobile agents	+	+	+	·	+	–	+	+
Camera vision	–	·	·	·	–	+	–	–
Autonomous	+	+	+	+	+	+	+	–
Task oriented	+	+	+	–	–	–	–	+
Extra–linguistic	–	–	+	+	·	–	+	–

form relates to a behaviour or real world phenomenon. The pre-defined language in Billard and Hayes' experiments is only given to the teacher robot, the student robot has to learn the language. Although in the work of Yanco & Stein the robots learn the language, the researchers have pre-defined the language and they provide feedback whether the language is used successfully. Rosenstein & Cohen (1998a) do not model language yet. Hence the question if they pre-define the language is not applicable. In the work done at the VUB AI Lab no such relationships are given to the agents. This is also not given in the work of Mike Oliphant (1997). This means that the agents construct the language themselves.

Meaning is pre-defined if the agents have some representation of the meaning pre-programmed. This is done in the work of Billard & Hayes (1997), Oliphant (1997) and Yanco & Stein (1993). In the work of Billard & Hayes, the meaning is only given to the teacher robot. The student robot learns the representation of the meaning. Oliphant's agents only have abstract meanings that have no relation to the real world. In the work that is done in most of Steels' group the agents construct their own ontology of meanings.

Of the researchers that are compared with this work, only Rosenstein & Cohen (1998a) make use of prototypes as a way of defining categories. All other work makes use of some other definition. This does not mean that the use of prototypes is uncommon in artificial intelligence, but it is uncommon in the grounding of language community.

A hierarchical structuring of the categorisations is only done by the researchers of Steels' group, this book included. The advantage of hierarchical structuring of categories is that a distinction can be either more general or more specific.

Quite some researchers pre-define the number of meanings and/or forms that is, or should arise in the language (Billard & Hayes 1997; Cangelosi & Parisi 1998; Oliphant 1997; Yanco & Stein 1993). Naturally, language is not bound by the number of meanings and forms. Therefore, the number of meanings and forms is unbound in this book.

It may be useful if the position of the robot in relation to other robots and objects in their environment is known exactly. Especially for technical purposes, like pointing to an object. However, such information is not always known to the language users. In the Talking Heads experiment, the robots have calibrated knowledge about their own position (which is fixed) and the position of the other robot, and they can calculate the position of objects in their world. Such information is not available to the robots in this book. This is one of the main differences between the Talking Heads and the current experiments. Another difference with the Talking Heads is the use of camera vision, rather than low-level sensing. Still other differences are at the implementation of the model. These differences have been discussed above and will be discussed more in Chapter ??.

Not all experiments deal with robots that are mobile in their environment. In particular the Talking Heads are not mobile, at least not in the sense that they can move freely in their environment. The Talking Heads can only go from physical head to physical head. The locations of these heads are fixed.

Except the work of Yanco & Stein (1993), all experiments are autonomous, i.e. without the intervention of a human. Yanco & Stein give their robots feedback about the effect of their communication. This feedback is used to reinforce the connections between form and meaning. The system designed in this book is completely autonomous. The only intervention taken is to place the robots at a close distance rather than letting them find each other. This is done in order to speed up the experiments. In previous implementations, the robots did find each other themselves (Steels & Vogt 1997). There is no intervention at the grounding and learning level involved.

In most of the experiments mentioned, the agents have only one task: developing language. Some scientist argue that language should be developed in a task-oriented way, e.g. Billard & Hayes (1997), Cangelosi & Parisi (1998), De Jong (2000) and Yanco & Stein (1993). In particular, the task should have an ecological function. This seems natural and is probably true. However, in order to understand the mechanisms involved in lexicon development, it is useful to concentrate only on lexicon development. Besides, developing langauge is in some sense task-oriented.

As explained, one of the research goals is to investigate the importance of extra-linguistic information that guides the lexicon development. This has also been investigated by Oliphant (1997) and De Jong (2000).

So, in many respects the research that is presented in this book is unique. It takes on many aspects of a grounded language experiment that is not shared by other experiments. The experiment that comes closest is the Talking Heads experiment. The results of the experiments from this book will therefore be compared in more detail at the end of this book.

1.6 The book's outline

The book is basically divided in three parts. In the first part, the model by which the experiments are developed is introduced. Part two presents experimental results. And the final part is reserved for discussions and conclusions.

Chapter 2 introduces the experimental set-up. This includes the environment in which the robots behave and the technical set-up of the robots. This chapter explains the Process Description Language (PDL) in which the robots are programmed. PDL is for the purpose of these experiments extended from a behaviour-based architecture in a behaviour-based *cognitive* architecture. This is to enable better controllable planned behaviour. People not interested in the technical details of the robots may omit this chapter. For these people it is advisable to read §2.1 in which the environment is presented. In addition, the part on the white light sensors in §2.2.1 is important to follow some of the discussions.

The language game model is introduced in Chapter ??. It explains how the robots interact with each other and their environment. The interaction with their environment includes sensing the surroundings. The result of the sensing is preprocessed further to allow efficient categorisation. The discrimination game with which categorisation and ontological development is modelled is explained. After that, the naming game is presented, which models the naming part of the language game and the lexicon formation. The chapter ends with a presentation of how the discrimination game and the naming game are coupled to each other.

The experimental results are presented in Chapters ??, ?? and ??. Chapter ?? first introduces the measures by which the results are monitored. The first experiment that is presented is called the BASIC EXPERIMENT. A detailed analysis is made of what is going on during the experiment. As will become clear it still has a lot of discrepancies. These discrepancies are mostly identified in following chapters.

The experiments presented in Chapter ?? are all variants of the basic experiment. In each only one parameter or strategy has been changed. The experiments investigate the impact from various strategies for categorisation, physical interaction, joint attention and feedback. In addition, the influence of a few parameters that control adaptation are investigated. Each set of experiments is followed by a brief discussion.

The final series experiments are presented in Chapter ??. Two variants of the language games that have proven to be successful in previous chapters are investigated in more detail. Each of these experiments have a varying strategy of using extra-linguistic information and are additionally provided with parameter settings that appeared to yield the best results. The first experiment is the GUESSING GAME in which the hearer has to guess what light source the speaker tries to name, without previous knowledge about the topic. In the second experiment prior topic knowledge is provided by joint attention. No feedback on the outcome is provided in the second game, called the OBSERVATIONAL GAME.

Chapter ?? discusses the experimental results and presents the conclusions. The discussion is centred on the research questions posed in the previous section. Additional discussions centre on the similarities and differences with related work, in particular with the work done by other members of the VUB AI Lab and Sony CSL Paris. Finally some possible future directions are given.

2 The sensorimotor component

In this chapter the design and architecture of the robots is discussed. The experiments use two small LEGO vehicles, which are controlled by a small sensorimotor board. The robots, including their electronics, were designed at the VUB AI Lab. They were constructed such that the configuration of the robots can be changed easily. Sensors may be added or changed and the physical robustness of the robots has improved through time. In some experiments they *are* changed substantially, but in most experiments the robots remain the same.

The robots are controlled by a specialised sensorimotor board, the SMBII[1] (Vereertbrugghen 1996). The sensorimotor board connects the sensory equipment with the actuators in such a way that the actuators and sensor readings are updated 40 times per second. The actuators respond to sensory stimuli, where the response is calculated by a set of "parallel" processes. These processes are programmed in the *Process Description Language* (PDL), which has been developed at the VUB AI Lab as a software architecture to implement behaviour-oriented control (Steels 1994b).

The outline of the experiments is discussed in Chapter ??; this chapter is concentrated on the physical set-up of the robots and their environment in the different experiments. The robots' environment is presented in §2.1. §2.2 discusses the physical architecture of the robots. Section 2.3 discusses the Process Description Language.

2.1 The environment

The environment that has been used for the experiments in the past varied across some of the experiments. The environment in early experiments (Steels & Vogt 1997; Vogt 1998a,b) had different light sources than the current environment. Furthermore, the size of the environment shrinked from $5 \cdot 5m^2$ to $2.5 \cdot 2.5m^2$. In the current environment there are four different white light sources, each placed at a different height (Figure 2.1).

[1] Read as SMB-2.

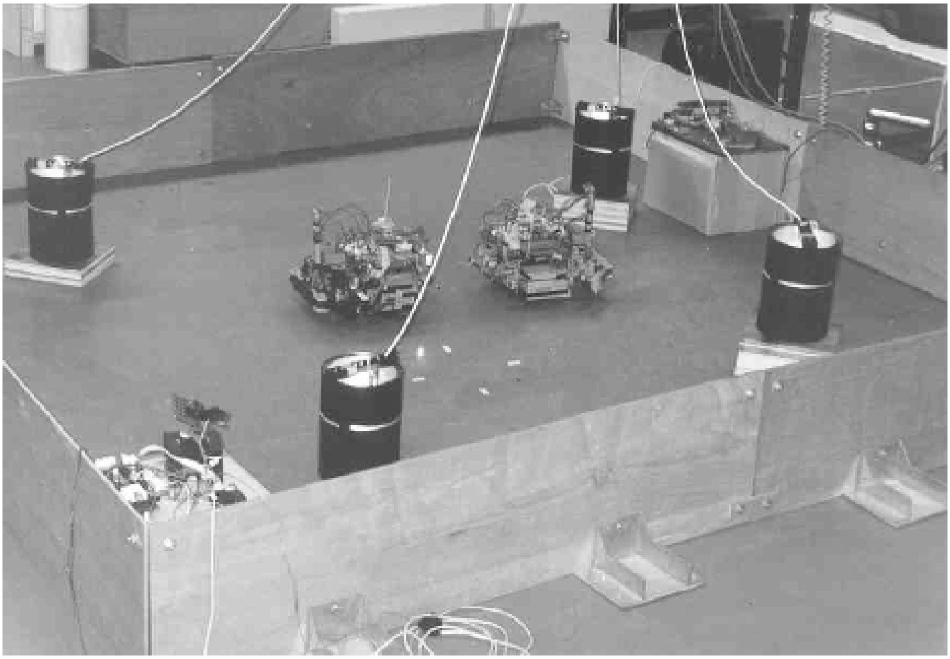


Figure 2.1: The robots in the environment as is used in the experiments.

These white light (WL) sources (or light sources for short) all emit their light from black cylindrical boxes with small slits. The light sources are halogen lights and each box now has a height of 22 cm, a diameter of 16 cm and 3 horizontal slits. Each slit has its centre at a height of 13 cm (measured from the bottom of the box) and is 0.8 cm wide. Although the different slits are intersected by a bar, they can be approximated to be one slit.

The boxes are placed such that the height of the slit varied per light source. The four different heights are distributed with a vertical distance of 3.9 cm. In one experiment the difference in height was changed to 2.9 cm. The robots were adjusted to this environment (or vice versa) so that the light sensors were placed at the same height as the centre of the slits.

2.2 The robots

In the experiments two LEGO robots as in Figure 2.2 are used. Each robot has a set of sensors to observe the world. These sensors are low-level. They can only detect the intensity of light in a particular frequency domain. Other low-level sensors

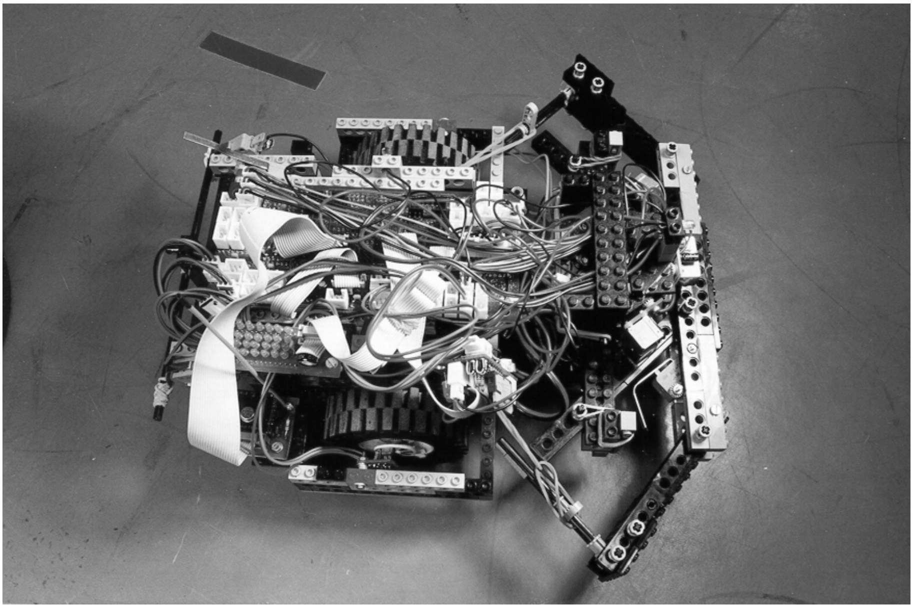


Figure 2.2: One of the LEGO robots used in the experiments.

are used to control the robots in their movement. The sensors are connected to a dedicated sensorimotor board, the so-called SMBII. On the SMBII all sensors are read at a rate of 40 Hz. The sensor readings are processed according to the software, written in PDL (see next section). After the sensors have been processed the SMBII outputs the actuator commands and sends its appropriate signals to the actuators. The robots are powered by a re-chargeable nickel–cadmium battery pack as used in portable computers.

In this section the set-up of the sensors and actuators of the robots are discussed first. Secondly the architecture of the SMBII is discussed briefly.

2.2.1 The sensors and actuators

The robots in all experiments have a set-up like shown schematically in Figure 2.3. The sensory equipment consists of four binary bumpers, three infrared (IR) sensors and a radio link *receiver*. The radio link is a module that also has a radio link *transmitter*, which is classified as an actuator. The infrared sensors are part of the infrared module, which also consists of an actuator: the infrared transmitter. Two independent motors complete the actuator set-up. All sensors and actuators are connected to the SMBII, which is powered by a battery-pack. The battery-pack

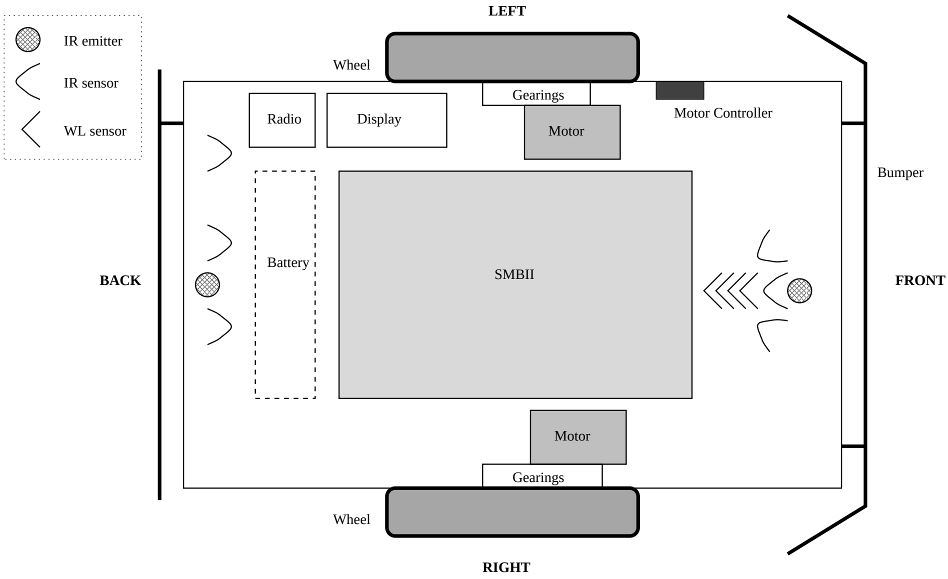


Figure 2.3: A schematic overview of the basic set-up of the robots that are used in the experiments.

also powers the motor-controller. The motor-controller, controlled by the SMBII controls the motors. The motors are connected to the wheels via a set of gears. Finally there are four white light sensors that are responsible for the perception.

Below a more detailed description of the most important sensors and actuators are given.

The bumpers The robots have four bumpers that are used for touch based obstacle avoidance; two on the front and two on the back of the robot, both left and right. Each bumper is a binary switch: when it is pressed it returns 1, else it returns 0. The bumpers have a spanning construction of LEGO (see Figure 2.4(a) and 2.4(b)). If a robot bumps with this construction into an obstacle. The program can then react on the sensed collision.

The infrared module Whereas the bumpers are simple binary sensors, the infrared module Figure 2.4(a) is more complex. The infrared module consists of infrared emitters and sensors. The emitters are light emitting diodes emitting infrared. The infrared sensors themselves are sensors that can be found in e.g. television sets. They detect light at infrared wavelengths and send a signal to the SMBII that is proportional to the intensity of the infrared. The sensors are mounted such that they can discriminate infrared coming from the left, centre and right sides in front of the robot. The sensors are not calibrated in the sense that one can calculate the exact angle from where the infrared is coming or from what distance. Also the positions of the sensors are not exactly symmetric, due to some physical limitations of the sensors and the LEGO construction. Vogt (1997) discusses some practical problems concerning the modulation and characteristics of the infrared module in detail.

The radio link The radio link module is a transmitter/receiver device designed to connect with the SMBII (see Figure 2.4(b)). The module is a Radiometrix BM-433F module with RX (receive) and TX (transmission) connections. The module can send up to 40 Kbit/s, but is used at 9.6 Kbit/s.

Every clock cycle of the SMBII a packet of messages can be sent. A packet can consist of a maximum of 31 messages each up to 127 bytes long. A message has a transmission ID and an destination address, which define the sender and receiver(s) of the message. It also has a bit defining the reliability of the transmission; this bit has to be set to *unreliable*, i.e. to 0, because the reliability protocol has not been implemented in the radio link kernel. This has the consequence that if a message is sent, it is not sure if the message arrives at its destination. But *when* it arrives, the message arrives error-less. About 5 % of the messages sent do not arrive at their destination.

This unreliability has some technical impacts on the experiments. Since data logging, recording and communication passes through the radio link, not all information is received. Filters had to written to find out whether all data was logged and if not, part of the data would be unreliable and should therefore be discarded. It is beyond the scope of this dissertation to go into the details of such filters here. For the purpose of the book it is assumed that the radio transmission is reliable.

The motor controller and the motors The motor controller is a device that transforms and controls motor commands coming from the SMBII into signals that are sent to the standard LEGO DC motors. Each robot has two independent motors. So, in order to steer the robot, one has to send a (possibly different) signal to each motor.

Gearing The motors are not directly connected to the wheels. They are connected to the wheels with a set of gears (see Figure 2.4(c)). The wheels are placed such that they form an axis approximately through the centre of the robot so that it can rotate around this point. A third small caster-wheel is used to stabilise the robot.

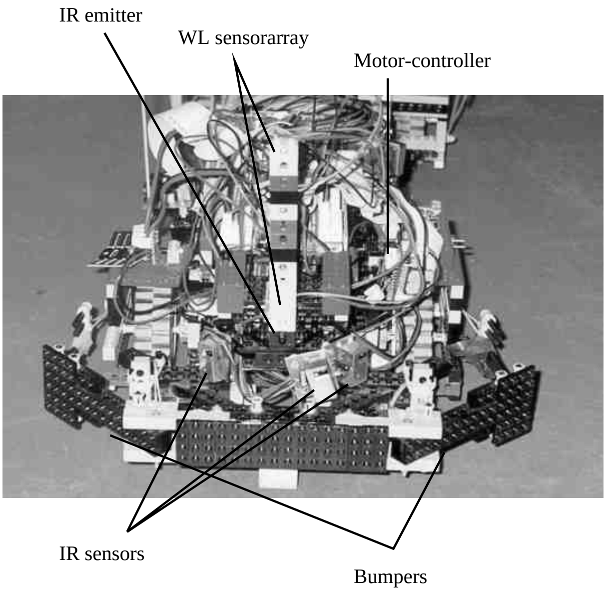
The light sensors The white light sensors are the most crucial sensors in the experiments. This is because they are used for the perception of the analogue signals that the robots are supposed to ground. Each robot has four white light sensors stacked on top of each other. The sensors have a vertical distance of 3.9 cm between each other. Each sensor is at the same height as a light source (Figure 2.4(a)).

The light sensors were calibrated such that the characteristics of all sensors are roughly the same. Figure 2.5 shows the characteristics of the calibrated light sensors as empirically measured for the experimental set-up. On the x-axis of each plot the distance of the robot to the light source is given in centimetres; the y-axis shows the intensity of the light in PDL values. PDL scales the light sensors between 0 and 255, where 0 means no detection of light and 255 means maximum intensity. The calibration of each sensor is done while it was exposed to a corresponding light source. A sensor is said to *correspond* with a light source when it has the same height. The complete figure shows the characteristics of the two robots $r0$ and $r1$, each with four sensors ($s0 \dots s3$).

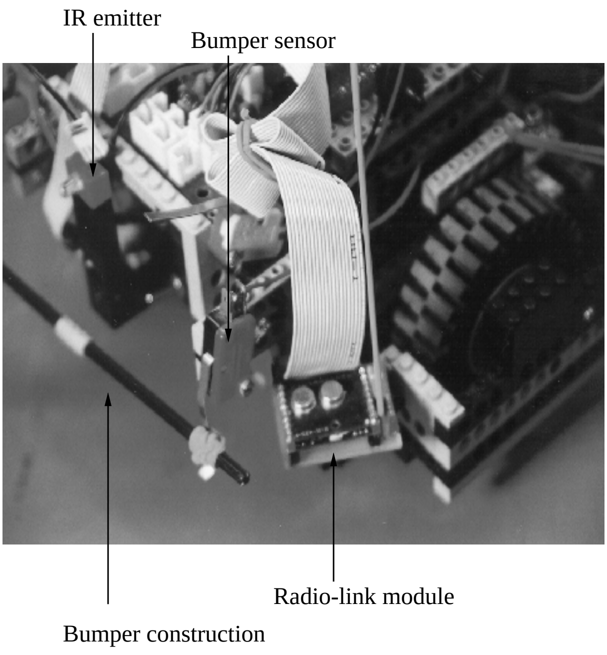
It is notable that for light source $L0$ the characteristics of sensor $s3$ is high at the beginning (Figure 2.5 (a) and (e)). This is because for the lowest light source $L0$, sensor $s3$ is higher than the top of the box, which is open. At a larger distance the light coming from the top of this box cannot be seen.

It is clear that all characteristics are similar. The sensor that corresponds to a light source detects high intensities at short distances and low values at larger distances. From 0.6 m other sensors start detecting the light source as well. This is because the light coming from the slit does not propagate in a perpendicular beam, but is diverging slightly. It is important to note that corresponding light sensors are calibrated to read the highest intensities between 0 and 1.2 m. The shape of the plots are like they would have been expected from the physics rule that the intensity $I \sim \frac{1}{r^2}$, where r is the distance to the light source.

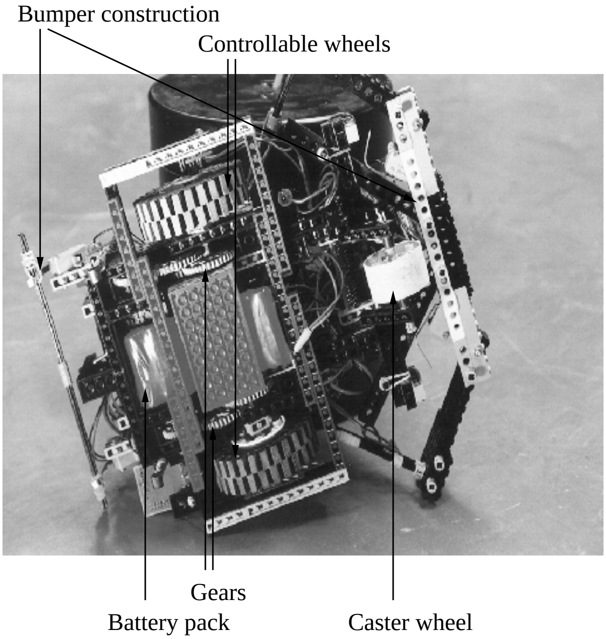
It is noteworthy that each sensor detects noise that comes mainly from ambient light.



(a) Front



(b) Back

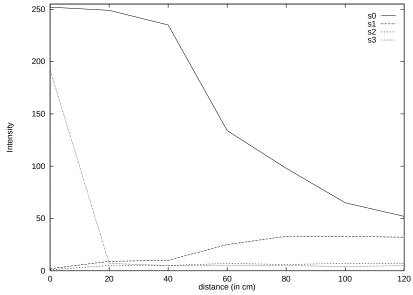


(c) Bottom

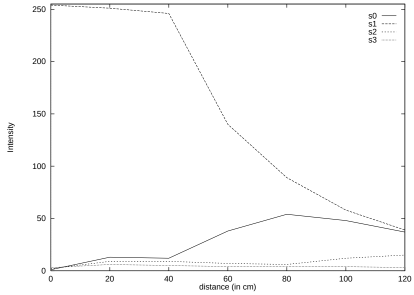
Figure 2.4: Several close ups of one of the robots. Figure (a) shows the front side of the robot. The bumper construction can be seen. The perceptual sensor array consisting of 4 light sensors, the infrared sensors and the infrared emitter are also visible. The radio link module can be seen in (b) as well as a part of the bumper construction on the back. Figure (c) shows the bottom of the robot. We see the wheels, gearing and the battery pack. Also a good view is seen of the bumper constructions.

The robots are also equipped with sensors and actuators that are used for interfacing the robot with the experimenter. It has for instance a serial port for connecting the robot to a PC, a display with 64 LEDS, a pause button, an on/off switch, etc. Since these sensors are not vital for the behaviour of the robots, they are not discussed in more detail here.

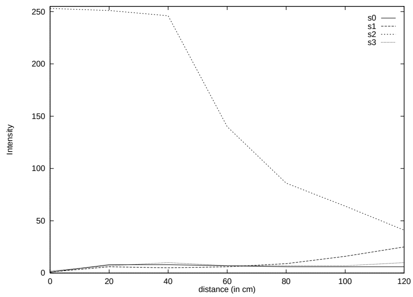
This subsection introduced the sensorimotor equipment that the robot carries in the experiments as discussed throughout this book. The next subsection discusses the sensorimotor board in some more detail.



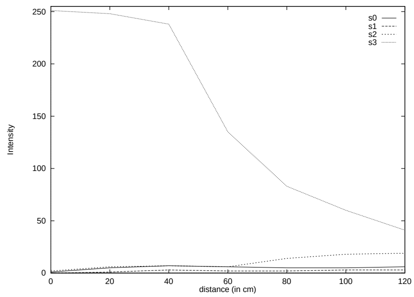
(a) r0, L0



(b) r0, L1



(c) r0, L2



(d) r0, L3

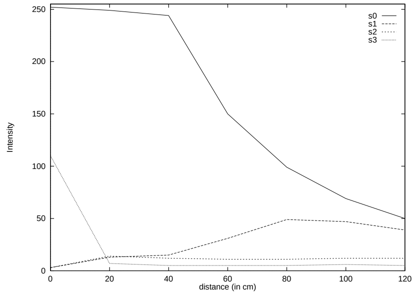
2.2.2 Sensor-motor board II

The computing hardware of the robots is a sensorimotor board, called the SMBII, which is developed at the VUB AI Lab by Dany Vereertbrugghen (1996). It consists of an ADD-ON SMB-2 BOARD and a Vesta Technologies SBC332 micro controller board.

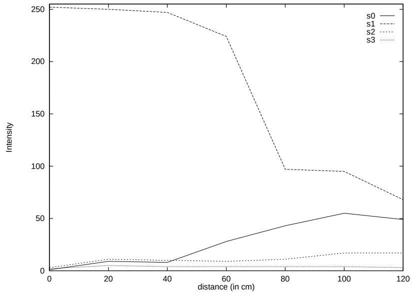
The Vesta board (see Figure 2.6(a)) contains a Motorola MC68332 microcontroller, 128 kB ROM and 1 MB RAM.[2] The board's micro-controller runs at 16.78 MHz at 5 Volt and everything is powered by a pack of rechargeable nickel–cadmium batteries.

The add-on SMB-2 board (Figure 2.6(b)) contains several I/O chips, bus controllers and connectors. The SMBII low-level program is run on the kernel and it can interface a user program written in any language as long a the kernel calls are written in C (Vereertbrugghen 1996). The program that is run on the SMBII for these experiments is written in the Process Description Language PDL.

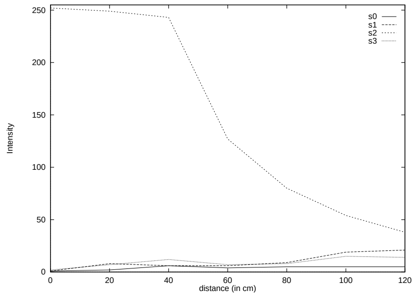
[2] In the original version of the SMBII there were only 256 kB RAM (Vereertbrugghen 1996).



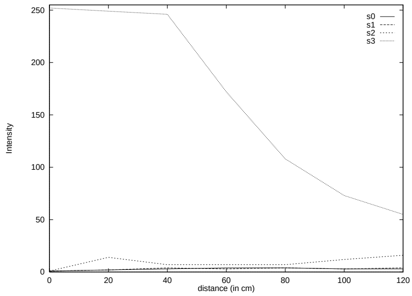
(e) r1, L0



(f) r1, L1



(g) r1, L2



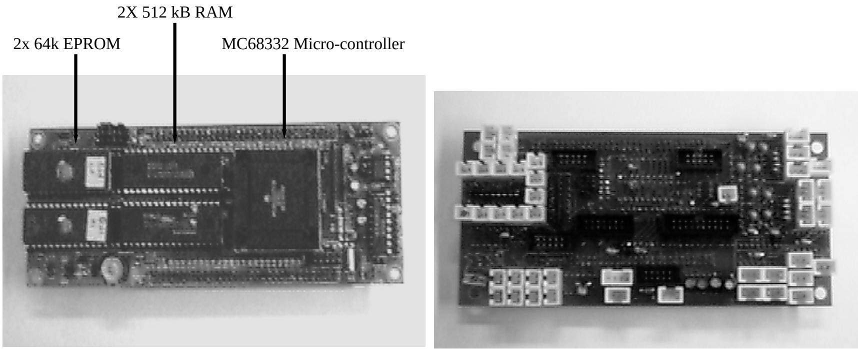
(h) r1, L3

Figure 2.5: The characteristics of the calibrated light sensors as empirically measured for the experimental set-up when exposed to light sources $L0 - L3$. Plots (a) – (d) show the of robot $r0$ and plots (e) – (h) show them for $r1$. The distances are measured from the front of the robots to the boxes. Actual distances from source to sensor are 12 cm further.

2.3 The Process Description Language

The robots are programmed in the so-called Process Description Language (Steels 1992, 1994a,b). PDL is designed as a framework for designing software for autonomous agents according to the behaviour-oriented control.

In PDL one can decompose a behaviour system in a set of dynamical processes. For instance, one can decompose the behaviour of PHOTOTAXIS (i.e. moving towards a light source) into two dynamical processes: (1) moving forward and (2) orienting towards the light. PDL is designed to implement parallel processes that are virtually evaluated simultaneously to output a summated response. So, suppose there are the two parallel processes (1) and (2) that are evaluated simultane-



(a) Vesta board (b) SMB-2 add-on

Figure 2.6: The two components of the SMBII board: (a) the Vesta Technologies SBC332 micro controller board, and (b) the add-on SMB-2 sensorimotor board.

ously. And suppose further that the output of the two processes are summated to give a motor response. Then the emergent behaviour is phototaxis.

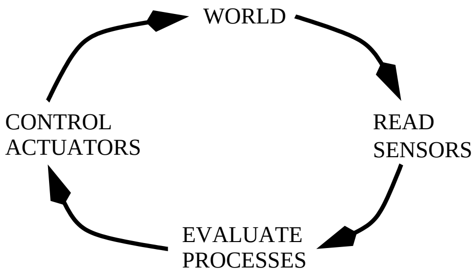


Figure 2.7: The way that PDL interacts with the world via a robot. Every $\frac{1}{40}s$ PDL is going through a cycle as shown in the figure.

PDL cycles the process of READING SENSORS, EVALUATE PROCESSES and CONTROL ACTUATORS (Figure 2.7). During a PDL cycle a robot reads the sensors to detect the current state of a robot in the world. These sensor readings are evaluated by processes that are defined in the software as explained below. The processes output commands to activate the actuators. These actuators in turn change the state of the world. Such a cycle is processed at 40 Hz, so 40 PDL cycles take 1 second. Throughout the book the basic time unit is a PDL cycle ($\frac{1}{40}s$).

The initial implementation of PDL was written in LISP, the currently used version is implemented in ANSI-C. It can compile both the specialised PDL syntax and ANSI-C commands within its architecture. The PDL architecture has as its basic symbolic units so-called `quantities`. A quantity is a `struct` type that has a NAME, a VALUE, an UPPER BOUND, a LOWER BOUND and an INITIAL VALUE. Each quantity can be connected to a serial port, interfacing the program with the physical sensors and actuators. Each type of sensor and actuator is defined within the operating system of the SMBII. The radio module has its own interface, but can be called with a PDL command. The most important parts of a PDL program are the *processes*. Each time a PDL program is compiled, a network of processes is build up. The following example of phototaxis shows how this is done.

The example implements two behaviours: (1) INFRARED ORIENTATION and (2) INFRARED PHOTOTAXIS. In infrared orientation, the goal of the robot is to orient itself in the direction of an infrared source without approaching the source.[3] It is implemented using only one dynamic process called TAXIS. With infrared phototaxis the goal of a robot is to approach the infrared source. It is implemented with an additional process that causes a robot to try to move at a default speed.

After declaration, the quantities have to added to the system as follows:

```
add_quantity(LeftFrontIR,''LeftFrontIR'',255.0f,0.0f,0.0f);
add_quantity(RightFrontIR,''RightFrontIR'',255.0f,0.0f,0.0f);
add_quantity(LeftMotor,''LeftMotor'',100.0f,-100.0f,0.0f);
add_quantity(RightMotor,''RightMotor'',100.0f,-100.0f,0.0f);
```

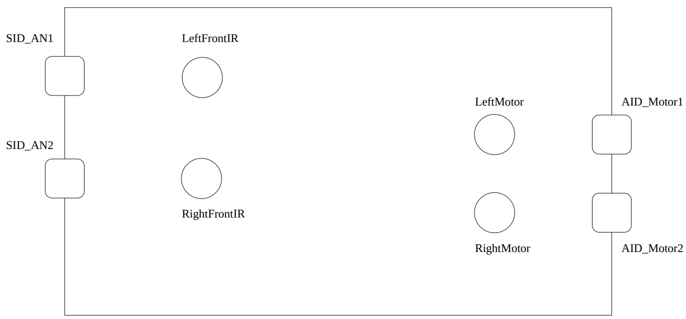


Figure 2.8: The construction of a PDL network. The program has serial ports SID_AN1 and SID_AN2 for analogue sensory input. Ports AID_MOTOR1 and AID_MOTOR2 are serial ports for the motors. The network consists of the quantities `LeftFrontIR`, `RightFrontIR`, `LeftMotor` and `RightMotor`.

[3] In the experiments the robots themselves are infrared sources.

The function `add_q` adds the quantity `LeftFrontIR` to the network, an upper bound of `255.0f` (where `f` stands for "float"), a lower bound of `0.0f` and an initial value of `0.0f`. Likewise the quantities `RightFrontIR, LeftMotor` and `RightMotor` were added, see Figure 2.8. The upper and lower bound of the motors are 100.0 and −100.0, respectively. If, mathematically, an upper or lower bound would be exceeded, PDL sets the quantity-value to its upper or lower bound. The next step is to connect the quantities to the serial ports of the SMBII, which are connected to the sensors and actuators.

```
add_connection(SID_AN1,LeftFrontIR);
add_connection(SID_AN2,RightFrontIR);
add_connection(AID1_Motor,LeftMotor);
add_connection(AID2_Motor,RightMotor);
```

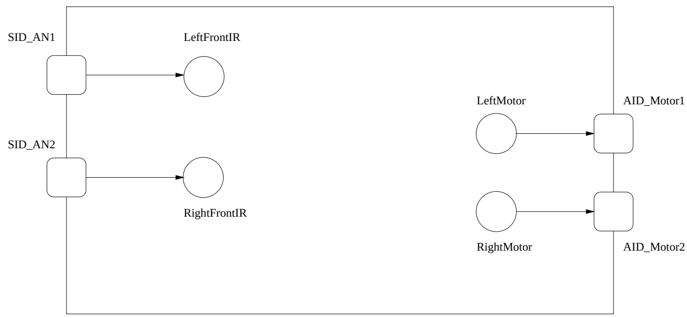


Figure 2.9: This is the PDL network after the quantities are connected with the serial ports.

Now the network looks like in Figure 2.9. The above is part of the initialisation. Another step of the initialisation is to add processes to the network:

```
add_process(Taxis,''Taxis'');
add_process(TowardsDefault,''TowardsDefault'');
```

This leads to the network as shown in Figure 2.10. To couple the sensors with the actuators, the processes have to be defined. The process `Taxis` causes the robot to orient towards an infrared light source.

```
void Taxis()
{
  D=value(RightFrontIR)-value(LeftFrontIR);
  add_value(LeftMotor,C*F(D)*D));
  add_value(RightMotor,-C*F(D)*D));
}
```

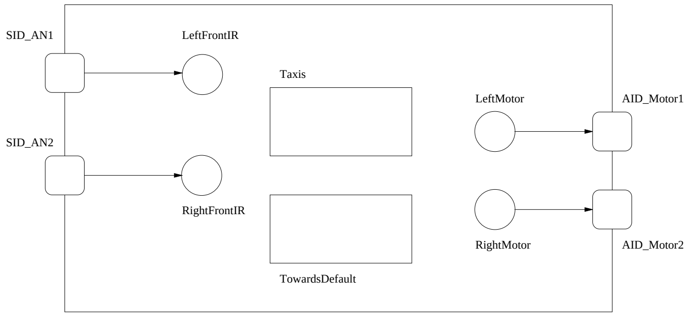


Figure 2.10: This is the PDL network after the processes `Taxis` and `TowardsDefault` are added.

Here, `value(Q)` is a function that returns the value of quantity `Q`, `add_value(Q,V)` adds value `V` to the value of `Q`. The actual update of `Q` is done at the end of each PDL cycle. When more values are added to `Q`, these values are summed before they are added. `C` is a constant and `F(D)` is a scaling factor of difference (`D`) in infrared. $F(x)$ is implemented as an inverse sigmoid function.

$$F(x) = \frac{1}{1 + e^{\alpha \cdot (x - \beta)}}$$

$F(x) \cdot x$ dampens x strongly if x is large, it is less dampened if x is not large. If $F(x)$ is not applied, the robot would exaggerate its wiggling too much.

`Taxis` increases the `LeftMotor` and decreases the `RightMotor` by a value proportionate to `D`. If the infrared source is to the right of the robot, the difference `D` is positive. Hence the value of the `LeftMotor` increases and the value of the `RightMotor` decreases. This in effect causes the robot to turn to the right. When the infrared source is to the left of the robot, the opposite happens. So, the robot will rotate in the direction in which the intensity of infrared is detected the highest. If the robot passes the infrared source, the direction in which the infrared is detected (i.e. the sign of direction changes) and so the robot changes its direction of rotation. This will continue until `D` approaches zero or when the values become so small, that it there is no power left to move the robot.

Although the robot is rotating around its axis in varying directions, it does not move from its place. This is accomplished by introducing the following process:

```
void TowardsDefault()
{
  add_value(LeftMotor,(DefaultSpeed-value(LeftMotor))/Step);
```

```
  add_value(RightMotor,(DefaultSpeed-value(RightMotor))/Step);
}
```

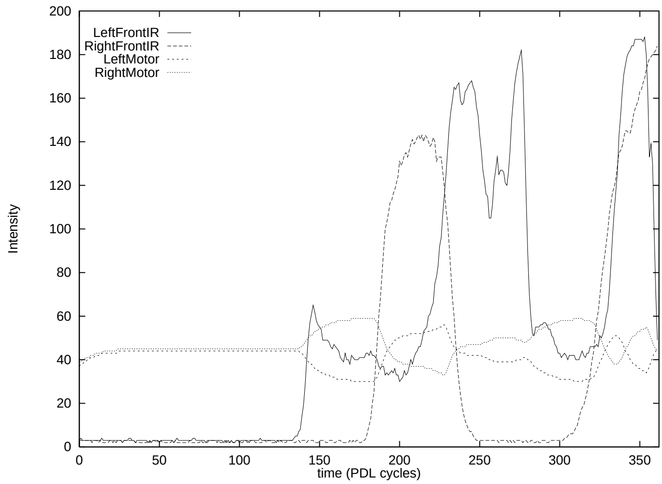


Figure 2.11: This figure shows the evolution of the infrared sensors and motor values in time during phototaxis, i.e. the emergent dynamics of combining the processes `Taxis` and `TowardsDefault`. On the x-axis the time is shown in the basic time unit of the robots, a PDL cycle ($= \frac{1}{40} s$). The y-axis shows the intensity of the infrared sensors and motor signals. The data is taken from a robot that was driving using both processes `Taxis` and `TowardsDefault`. It drove straight forward until at time 140 the robot detected an infrared source after which it adjusted its motor signals to home in on the source.

This process causes the robot to change its speed towards a default speed with certain step size. The step size is introduced to let the robot accelerate smoothly. Note that this way the motor values do not reach the default speed; the values approach it asymptotically. When the processes are defined, the network looks like in Figure 2.12.

Taking the two processes together results in the emergent behaviour that the robot will move wiggling towards the infrared source (see Figure 2.11). Such phototaxis behaviour, although with a slightly different implementation, was introduced for a robot application by Valentino Braitenberg (1984) and has already been discussed extensively in the literature, see e.g. Steels 1994c.

Appendix **??** presents the structure of the implemented PDL program in more

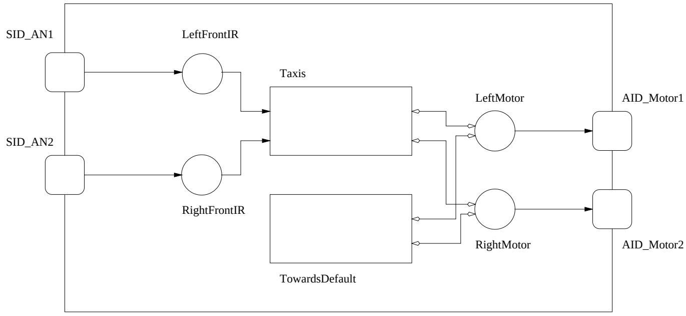


Figure 2.12: Finally the network of quantities and processes for the phototaxis example is complete. The `LeftFrontIR` and `RightFrontIR` are connected to input `Taxis`, which outputs to the motor quantities. The motor quantities are also used to calculate the output, hence this connection is bi-directional. The process `TowardsDefault` does not use any sensors; as in Taxis it only uses values of the quantities `LeftMotor` and `RightMotor` thus giving the bi-directional connection between `TowardsDefault` and the motors.

detail. In the next section the behaviour based architecture is expanded to incorporate planned behaviours as well.

2.4 Cognitive architecture in PDL

To accomplish a complex task like communication, a sequence of actions have to be planned. Reactive behaviours like phototaxis alone do not suffice. To allow the robots to execute planned behaviour a new architecture has been developed. This resulted in what could be called a BEHAVIOUR-BASED COGNITIVE ARCHITECTURE that is primarily based on the behaviour-based control architecture proposed by Luc Steels (1994b). This cognitive architecture could be applied as a general purpose architecture for complex and dynamic tasks like navigation. The architecture executes a script (or plan) through excitation and inhibition of processes that altogether result in some emergent behaviour. The scripts are implemented as finite state automata in which transitions are controlled by state-specific pre- and post-conditions. In each state of the finite state automaton (FSA) a particular set of processes are activated or inhibited. Figure 2.13 shows the basic principle.

In the architecture the sensors *Se* and actuators *A* are coupled through a complex of connections. The agent consists of a set of scripts, which are implemented

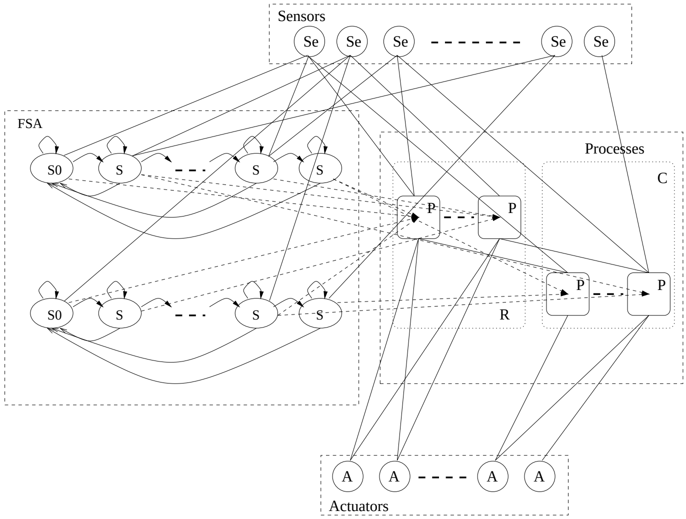


Figure 2.13: A schematic overview of the developed architecture. See the text for details.

as finite state automata. The finite state automata are parallel processes where transitions are regulated by pre- and post-conditions. Usually the pre- and post-conditions are satisfied by some sensory stimuli. A state may also be fed with information coming from some internal process (not shown). Every state *S* has a post-condition that allows the system to enter the default state *S*0 where nothing happens. Each state of the automaton has excitatory and inhibitory connections with dynamic sensorimotor processes *P*. The excitatory connections are drawn as dotted lines, the inhibitory have been left out for clarity of the picture. The processes are divided between reactive (R) and cognitive (C) processes. The reactive processes have more direct processing and can take usually only sensorimotor data as input. The cognitive processes are more complex, and may take also stimuli coming from other internal processes. Note that the finite state automaton could be considered as a cognitive process as well. The configuration of excitatory processes and the dynamics of the robot with its environment cause the robot to perform some emergent behaviour. Hence the system is consistent with the behaviour-based paradigm.

Activation of processes is modelled by invoking motivational factors (cf. Steels 1996d; Jaeger & Christaller 1998). For example if is a state that in which the motivation for doing infrared taxis is present, this state may be a motivational factor `MotIRT` that is set to 1. The process taxis can then look like this:

```
void Taxis()
{
  D=value(RightFrontIR)-value(LeftFrontIR);
  add_value(LeftMotor,MotIRT*C*F(D)*D));
  add_value(RightMotor,-MotIRT*C*F(D)*D));
}
```

A multi-agent system is a parallel process in which two robots cooperate autonomously. In order to synchronise these two parallel processes, the robots use pre-programmed radio communication. The robots playing a language game process dependent, but parallel operating finite state automata. A signal is broadcasted when both robots should transfer to another state simultaneously as the result of the transition of one of the robots.

Because the architecture uses finite state automata, readers may wrongly suggest it is the subsumption architecture proposed by Rodney Brooks (1990). In the subsumption architecture each process is viewed as a finite state automaton on its own with only one state that models a behaviour (Figure 2.14 (a)). The architecture proposed here uses possibly more finite state automata each with a sequence of states that can be entered (Figure 2.14 (b)). These finite state automata are used to control planning. A process in the cognitive architecture can be activated by several states, and a particular state can activate several processes. In addition the processes couple the sensors with the motors, like the behaviour-based architecture proposed by Luc Steels (1994b).

The behaviour-based cognitive architecture has strong similarities with the dual dynamics architecture (Jaeger & Christaller 1998). However, the in the dual dynamics the activation of processes is regulated internally of these processes. There is no explicit finite state automaton that regulates the activation.

The architecture proposed here is also similar to the architecture proposed by Barnes (1996) and Barnes et al. (1997), called the behaviour synthesis architecture (BSA), which synthesises a set of *behaviour patterns* with a certain utility (or strength) for accomplishing a task. A *behaviour script* controls a sequence of *behaviour packets*. Each behaviour packet consists of a set of behaviour patterns, a pre-condition and a post-condition. Comparing the behaviour patterns with the dynamical processes of PDL, the behaviour scripts with the finite state automata and the packets with a single state, then the BSA is very close to the architecture that has been incorporated here. Main differences with the work of Barnes (1996) is the use of utility functions as its synthesis mechanism. Although the architecture here is developed by a human programmer, Barnes et al. (1997) show that planning can be automated using the BSA.

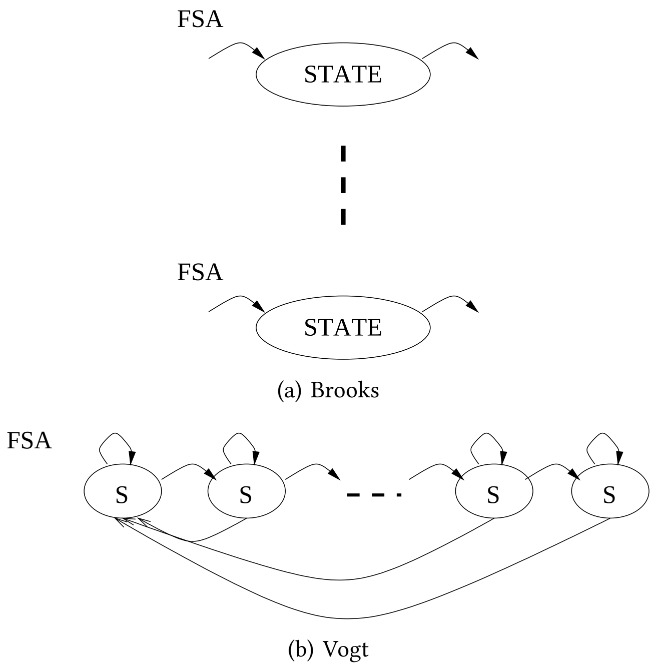


Figure 2.14: The finite state automata as used in the subsumption architecture (a) and in the cognitive architecture (b). In the subsumption architecture the finite state automata usually only has one state that models a particular behaviour. This behaviour can inhibit (or subsume) another behaviour. The cognitive architecture has some finite state automata each modelling a script-like behaviour. Each state excites or inhibits a number of dynamical processes. The finite state automata function independently as a parallel process.

2.5 Summary

In this chapter the basic set-up of the robots, their software and environment were introduced. The experiments use two small LEGO vehicles that are equipped with a set of sensors, actuators, a battery pack and a specialised sensorimotor board SMBII. The robots are programmed in a specialised programming language PDL, which is dedicated to process the dynamics of sensorimotor behaviours in the behaviour-oriented paradigm.

The principles of PDL have been extended to a behaviour-based cognitive architecture. In this new architecture robots can execute planned behaviour as cognitive processes.

The robots as introduced here are the physical bodies with which the agents try to develop their ontologies and lexicons. How they do that is explained in Chapter ??. As shall become clear some processing is done off-board. This is mainly done to experiment more efficiently and to be able to test different approaches on recorded data. In some specific experiments the architecture of the robots has been changed with respect to the description given in this chapter. Relevant changes will be reported when these experiments are discussed.

More detailed information on the PDL program can be found in Appendix ??.

Bibliography

Aitchison, J. 1996. *The seeds of speech: Language origin and evolution.* Cambridge University Press.

Barnes, D. P. 1996. A behavior synthesis architecture for co-operant mobile robots. In J. O. Gray & D. G. Caldwell (eds.), *Advanced robotics and intelligent machines*, 295–314. IEE Control Engineering Series 51.

Barnes, D. P., R. S. Aylett, A. M. Coddington & R. A. Ghanea-Hercock. 1997. A hybrid approach to supervising multiple co-operant autonomous mobile robots. In *Proceedings of the international conference on advanced robotics*, .

Baronchelli, A., M. Felici, E. Caglioti, V. Loreto & L. Steels. 2006. Sharp transition towards shared lexicon in multi-agent systems. *Journal of Statistical Mechanics* P06014.

Barrett, M. 1995. Early lexical development. In P. Fletcher & B. MacWhinney (eds.), *The handbook of child language*, 362–392. Basil Blackwell Ltd.

Belpaeme, T., L. Steels & J. van Looveren. 1998. The construction and acquisition of visual categories. In A. Birk & J. Demiris (eds.), *Learning robots, proceedings of the ewlr-6, lecture notes on artificial intelligence 1545*, Springer.

Berthouze, L. & Y. Kuniyoshi. 1998. Emergence and categorization of coordinated visual behavior through embodied interaction. *Machine Learning* 31/1,2,3. 187 – 200.

Billard, A. 1998. *Drama, a connectionist model for robot learning: Experiments on grounding communication through imitation in autonomous robots*: University of Edingburgh dissertation.

Billard, A. & G. Hayes. 1997. Robot's first steps, robot's first words ... In Sorace & Heycock (eds.), *Proceedings of the gala '97 conference on language acquisition - edingburgh*, Human Communication Research Centre: University of Edinburgh.

Bowerman, M. 1988. The "no negative evidence" problem: How do children avoid constructing an overly general grammer. In J. A. Hawkins (ed.), *Explaining language universals*, chap. 4, 73–101. Basil Blackwell.

Braine, M. D. S. 1971. The acquisition of language in infant and child. In C. Reed (ed.), *The learning of language*, New York: Scribners.

Braitenberg, V. 1984. *Vehicles, experiments in synthetic psychology.* Cambridge MA.: The MIT Press.

Bretano, F. 1874. *Psychology from an empirical standpoint.* Leipzig: Duncker und Humbolt.

Brooks, R.A. 1986. A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation* 2. 14–23.

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.

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

Cangelosi, A. & D. Parisi. 1998. The emergence of "language" in an evolving population of neural networks. *Connection Science* 10. 83–93.

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

Chomsky, N. 1956. Three models for the description of language. *IRE Transactions on Information Theory IT* 2(3). 13–54.

Churchland, Patricia S. 1986. *Neurophilosophy.* Cambridge, MA: The MIT Press.

Clancey, W. J. 1997. *Situated cognition.* Cambridge Univsity Press.

Clark, E.V. 1993. *The lexicon in acquisition.* Cambridge University Press.

Darwin, C. 1968. *The origin of species.* Penguin Books.

Dawkins, R. 1976. *The selfish gene.* Oxford: Oxford University Press.

De Boer, B. 1997. Generating vowels in a population of agents. In C. Husbands & I. Harvey (eds.), *Proceedings of the fourth european conference on artificial life*, 503–510. Cambridge, MA and London: MIT Press.

De Boer, B. 1999. *Self-organisation in vowel systems*: Vrije Universiteit Brussels dissertation.

De Boer, B. & P. Vogt. 1999. Emergence of speech sounds in changing populations. In D. Floreano, J.-D. Nicoud & F. Mondada (eds.), *Advances in artificial life: Proceedings of 5th european conference ecal'99*, Springer.

De Jong, E. D. 1999a. Autonomous concept formation. In *Proceedings of the sixteenth international joint conference on artificial intelligence ijcai'99*, .

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

De Jong, E.D. 1999b. Analyzing the evolution of communication from a dynamical systems perspective. In D. Floreano, J.-D. Nicoud & F. Mondada (eds.), *Proceedings of the fifth european conference on artificial life ecal'99*, Berlin: Springer-Verlag.

De Jong, E.D. 2000. *The development of communication*: VUB AI Lab dissertation. In preparation.

Demetras, M. J., K. Nolan Post & C. E. Snow. 1986. Feedback to first language learners: the role of repetitions and clarification questions. *Journal of Child Language* 13. 275–292.

Dennett, D.C. 1991. *Consciousness explained.* Boston: Little, Brown.

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

Eco, U. 1986. On symbols. In J. Deely, B. Williams & F.E. Kruse (eds.), *Frontiers in semiotics*, 153–180. Bloomington: Indiana University Press.

Edelman, G.M. 1992. *Bright air, brilliant fire.* New York: Basic Books Inc.

Fodor, J.A. 1975. *The language of thought.* Hassocks: Harvester Press.

Gardner, B.T. & R.A. Gardner. 1969. Teaching sign language to a chimpanzee. *Science* 165. 664–672.

Gasser, M. (ed). 1998. The grounding of word meaning: Data and models, papers from the 1998 workshop. Tech. Rep. Technical Report WS-98-06 AAAI Press.

Gilbert, N., M. den Besten, A. Bontovics, B.G.W. Craenen, F. Divina, A.E. Eiben, A. R. Griffioen, Gy. Hévézi, A. Lörincz, B. Paechter, S. Schuster, M.C. Schut, C. Tzolov, P. Vogt & L. Yang. 2006. Emerging artificial societies through learning. *Journal of Artificial Societies and Social Simulation* 9(2).

Greco, A., A. Cangelosi & S. Harnad. 1998. A connectionist model for categorical perception and symbol grounding. In *International conference on artificial neural networks*, .

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

Harnad, S. 1993. Grounding symbols in the analog world with neural nets. *Think* 2. 12 – 78.

Harnad, S., H. Steklis & J. Lancaster (eds.). 1976. *Origins and evolution of language and speech*, vol. 280. (Annals of the New York Academy of Sciences) The New York Academy of Sciences.

Jaeger, H. & T. Christaller. 1998. Dual dynamics: Designing behavior systems for autonomous robots. *Artificial Life and Robotics* 2. 108–112.

Kaplan, F. 2000. *L'émergence d'un lexique dans une population d'agent autonomes*: Université Pierre-et-Marie-Curie (Paris 6) dissertation. In preparation.

Kirby, S. & J. Hurford. 1997. Learning, culture and evolution in the origin of linguistic constraints. In C. Husbands & I. Harvey (eds.), *Proceedings of the fourth european conference on artificial life*, Cambridge, MA and London: MIT Press.

Kirby, S. & J. R. Hurford. 2002. The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi & D. Parisi (eds.), *Simulating the evolution of language*, 121–148. London: Springer.

Kröse, Ben, Roland Bunschoten, Nikos Vlassis & Yoichi Motomura. 1999. Appearance based robot localization. In G. Kraetzschmar (ed.), *Ijcai-99 workshop on adaptive spatial representations of dynamic environments*, 53 – 58.

Lakoff, G. 1987. *Women, fire and dangerous things*. The University of Chicago Press.

MacLennan, B. 1991. Synthetic ethology: an approach to the study of communication. In C.G. Langton, C. Taylor & J.D. Farmer (eds.), *Artificial life ii, vol. x of sfi studies in the sciences of complexiy*, Redwood City, CA: Addison-Wesley Pub. Co.

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

Nagel, T. 1974. What is it like to be a bat? *The Philosophical Review* 83. 435–450.

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

Newell, A. 1990. *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Nöth, W. 1990. *Handbook of semiotics*. Bloomington: Indiana University Press.

Oates, T. 1999. Identifying distinctive subsequences in multivariate time series by clustering. In *Proceedings of the fifth international conference on knowledge discovery and data mining*, .

Oates, T., Z. Eyler-Walker & P.R. Cohen. 1999. Using syntax to learn semantics: An experiment in language acquisition with a mobile robot. Tech. Rep. Technical Report 99-35 University of Amherst Amherst.

Oliphant, M. 1997. *Formal approaches to innate and learned communication: Laying the foundation for language*: University of California, San Diego dissertation.

Oliphant, M. 1998. Rethinking the language bottleneck: Why don't animals learn to communicate? In Chris Knight & J. R. Hurford (eds.), *The evolution of language (selected papers from the 2nd international conference on the evolution of language, london, april 6-9 1998)*, .

Oliphant, M. 2000. The learning barrier: Moving from innate to learned systems of communication. To appear in *Adaptive Behavior*.

Oudeyer, P.-Y. 1999. Experiment in emergent phonetics. Rapport de stage de deuxième année. Magistère informatique et modélisation, ENS Lyon.

Peirce, C.S. 1931–1958. *Collected papers*, vol. I-VIII. Cambridge, MA: Harvard University Press.

Pfeifer, R. & C. Scheier. 1999. *Understanding intelligence.* MIT Press.

Piaget, J. 1996. *La naissance de l'intelligence chez l'enfant.* Neufchatel: Delachaux et Netlé.

Pinker, S. 1984. *Learnability and cognition.* Cambridge, MA: MIT Press.

Premack, D. 1971. Langauge in chimpanzee? *Science* 172. 808–822.

Pylyshyn, Z.W. (ed.). 1987. *The robot's dilemma.* New Jersey: Ablex Publishing Corporation.

Rosenstein, M. & P. R. Cohen. 1998a. Concepts from time series. In *Proceedings of the fifteenth national conference on artificial intelligence*, Menlo Park Ca.: AAAI Press.

Rosenstein, M. & P. R. Cohen. 1998b. Symbol grounding with delay coordinates. In *Working notes of the aaai-98 workshop on: The grounding of word meaning*, Menlo Park Ca.: AAAI Press.

de Saussure, F. 1974. *Course in general linguistics.* New York: Fontana.

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

Searle, J.R. 1984. *Minds, brains and science.* Cambridge, MA: Harvard University Press.

Siskind, J. M. 1996. A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition* 61. 39–91.

Smith, K., A.D.M. Smith, R. Blythe & P. Vogt. 2006. Cross-situational learning: a mathematical approach. In P. Vogt, Y. Sugita, E. Tuci & C. Nehaniv (eds.), *Symbol grounding and beyond*, Springer.

Steels, L. 1992. The pdl reference manual. AI Lab Memo 92-05.

Steels, L. 1994a. Artificial life roots of artificial intelligence. *Arificial Life Journal* 1(1). 89–125.

Steels, L. 1994b. Building agents with autonomous behavior systems. In L. Steels & R.A. Brooks (eds.), *The "artificial life" route to "artificial intelligence". Building situated embodied agents*, New Haven: Lawrence Erlbaum Associates.

Steels, L. 1994c. A case study in the behaviour-oriented design of autonomous agents. In *Proceedings of the third simulation of adaptive behavior conference*, Cambridge, MA: The MIT Press.

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

Steels, L. 1996b. Emergent adaptive lexicons. In P. Maes (ed.), *From animals to animats 4: Proceedings of the fourth international conference on simulating adaptive behavior*, Cambridge, MA: The MIT Press.

Steels, L. 1996c. Perceptually grounded meaning creation. In M. Tokoro (ed.), *Proceedings of the international conference on multi-agent systems*, Menlo Park Ca.: AAAI Press.

Steels, L. 1996d. A self-organising spatial vocabulary. *Artificial Life Journal* 3(2).

Steels, L. 1996e. The spontaneous self-organization of an adaptive language. In S. Muggleton (ed.), *Machine intelligence 15*, Oxford: Oxford University Press.

Steels, L. 1997a. Language learning and language contact. In W. Daelemans, A. Van den Bosch & A. Weijters (eds.), *Workshop notes of the ecml/mlnet familiarization workshop on empirical learning of natural language processing tasks*, 11–24. Prague.

Steels, L. 1997b. Synthesising the origins of language and meaning using coevolution, self-organisation and level formation. In J. Hurford, C. Knight & M. Studdert-Kennedy (eds.), *Evolution of human language*, Edingburgh: Edingburgh University Press.

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

Steels, L. 1999. *The talking heads experiments. volume 1. words and meanings.* Special pre-edition for LABORATORIUM, Antwerpen.

Steels, L. 2000. The emergence of grammar in communicating autonomous robotic agents. In W. Horn (ed.), *Proceedings of ecai-2000*, Amsterdam: IOS Press.

Steels, L. & J. De Beule. 2006. Unify and merge in fluid construction grammar. In P. Vogt, Y. Sugita, E. Tuci & C. Nehaniv (eds.), *Symbol grounding and beyond: Proceedings of the third international workshop on the emergence and evolution of linguistic communication* (Lecture Notes in Computer Science), 197–223. Springer.

Steels, L. & R. Brooks. 1995. *The "artificial life" route to "artificial intelligence". building situated embodied agents.* New Haven: Lawrence Erlbaum Ass.

Steels, L. & F. Kaplan. 1998. Stochasticity as a source of innovation in language games. In *Proceedings of alive vi*, .

Steels, L. & F. Kaplan. 2000. Aibo's first words. the social learning of language and meaning. *Evolution of Communication* 4(1).

Steels, L., F. Kaplan, A. McIntyre & J. Van Looveren. 2002. Crucial factors in the origins of word-meaning. In A. Wray (ed.), *The transition to language*, 252–271. Oxford, UK: Oxford University Press.

Steels, L. & A. McIntyre. 1999. Spatially distributed naming games. In *Advances in complex systems*, .

Steels, L. & P. Vogt. 1997. Grounding adaptive language games in robotic agents. In C. Husbands & I. Harvey (eds.), *Proceedings of the fourth european conference on artificial life*, Cambridge, MA and London: MIT Press.

Steels, Luc (ed.). 2012. *Experiments in cultural language evolution*, vol. 3 (Advances in interaction studies). John Benjamins.

Tani, J. & S. Nolfi. 1998. Learning to perceive the world as articulated: An approach for hierarchical learning in sensory-motor systems. In R. Pfeifer, B. Blumberg, J-A. Meyer & S. Wilson (eds.), *From animals to animats 5, proceedings of the fifth internation conference on simulation of adaptive behavior*, Cambridge, MA: MIT Press.

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

Tomasello, M., S. Mannle & A. Kruger. 1986. The linguistic environment of one to two year old twins. *Developmental Psychology* 22. 169–176.

Van Looveren, J. 1999. Multiple word naming games. In E. Postma & M. Gyssens (eds.), *Proceedings of the eleventh belgium-netherlands conference on artificial intelligence*, University of Maastricht.

Vereertbrugghen, D. 1996. *Design and implementation of a second generation sensor-motor control unit for mobile robots*. Vrije Universiteit Brussel MA thesis.

Vogt, P. 1997. *A perceptual grounded self-organising lexicon in robotic agents*. Cognitive Science and Engineering, University of Groningen MA thesis.

Vogt, P. 1998a. The evolution of a lexicon and meaning in robotic agents through self-organization. In H. La Poutré & J. van den Herik (eds.), *Proceedings of the netherlands/belgium conference on artificial intelligence*, Amsterdam: CWI Amsterdam.

Vogt, P. 1998b. Perceptual grounding in robots. In A. Birk & J. Demiris (eds.), *Learning robots, proceedings of the ewlr-6, lecture notes on artificial intelligence 1545*, Springer.

Vogt, P. 1999. Grounding a lexicon in a coordination task on mobile robots. In E. Postma & M. Gyssens (eds.), *Proceedings of eleventh belgium-netherlands conference on artificial intelligence*, Univ. of Maastricht.

Vogt, P. 2000a. Bootstrapping grounded symbols by minimal autonomous robots. *Evolution of Communication* 4(1). 89–118.

Vogt, P. 2000b. Grounding language about actions: Mobile robots playing follow me games. In Meyer, Bertholz, Floreano, Roitblat & Wilson (eds.), *Sab2000 proceedings supplement book*, Honolulu: International Society for Adaptive Behavior.

Vogt, P. 2002. The physical symbol grounding problem. *Cognitive Systems Research* 3(3). 429–457.

Vogt, P. 2003a. Anchoring of semiotic symbols. *Robotics and Autonomous Systems* 43(2). 109–120.

Vogt, P. 2003b. THSim v3.2: The Talking Heads simulation tool. In W. Banzhaf, T. Christaller, P. Dittrich, J. T. Kim & J. Ziegler (eds.), *Advances in artificial life - proceedings of the 7th european conference on artificial life (ecal)*, Springer Verlag Berlin, Heidelberg.

Vogt, P. 2005a. The emergence of compositional structures in perceptually grounded language games. *Artificial Intelligence* 167(1–2). 206–242.

Vogt, P. 2005b. On the acquisition and evolution of compositional languages: Sparse input and the productive creativity of children. *Adaptive Behavior* 13(4). 325–346.

Vogt, P. 2006. Language evolution and robotics: Issues in symbol grounding and language acquisition. In A. Loula, R. Gudwin & J. Queiroz (eds.), *Artificial cognition systems*, 176–209. Hershey, PA: Idea Group Publishing.

Vogt, P. 2007. Variation, competition and selection in the self-organisation of compositionality. In B. Wallace, A. Ross, J. B. Davies & T. Anderson (eds.), *The mind, the body and the world: Psychology after cognitivism?*, 233–256. Exeter: Imprint Academic.

Vogt, P. 2012. Exploring the robustness of cross-situational learning under Zipfian distributions. *Cognitive Science* 36. 726–739.

Vogt, P. & H. Coumans. 2003. Investigating social interaction strategies for bootstrapping lexicon development. *Journal for Artificial Societies and Social Simulation* 6(1). Http://jasss.soc.surrey.ac.uk.

Vogt, P. & B. de Boer. 2010. Language evolution: Computer models for empirical data. *Adaptive Behavior* 18(1). 5–11.

Vogt, P. & F. Divina. 2007. Social symbol grounding and language evolution. *Interaction Studies* 8(1). 31–52.

Vogt, P. & E. Haasdijk. 2010. Modelling social learning of language and skills. *Artificial Life* 16.

Vogt, P. & E. Lieven. 2010. Verifying theories of language acquisition using computer models of language evolution. *Adaptive Behavior* 18. 21–35.

Vogt, P. & J. D. Mastin. 2013. Anchoring social symbol grounding in children's interactions. *Künstliche Intelligenz* 27. 145–151.

Werner, G.M. & M.G. Dyer. 1991. Evolution and communication in artificial organisms. In C.G. Langton, C. Taylor & J.D. Farmer (eds.), *Artificial life ii, vol. x of sfi studies in the sciences of complexiy*, Redwood City, CA: Addison-Wesley Pub. Co.

Winograd, T. 1972. *Understanding natural language.* Academic Press.

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

Yanco, H. & L. Stein. 1993. An adaptive communication protocol for cooperating mobile robots. In J-A. Meyer, H.L. Roitblat & S. Wilson (eds.), *From animals to animats 2. proceedings of the second international conference on simulation of adaptive behavior*, 478–485. Cambridge, MA: The MIT Press.

How mobile robots can self-organise a vocabulary

One of the hardest problems in science is the symbol grounding problem, a question that has intrigued philosophers and linguists for more than a century. With the rise of artificial intelligence, the question has become very actual, especially within the field of robotics. The problem is that an agent, be it a robot or a human, perceives the world in analogue signals. Yet humans have the ability to categorise the world in symbols that they, for instance, may use for language.

This book presents a series of experiments in which two robots try to solve the symbol grounding problem. The experiments are based on the language game paradigm, and involve real mobile robots that are able to develop a grounded lexicon about the objects that they can detect in their world. Crucially, neither the lexicon nor the ontology of the robots has been preprogrammed, so the experiments demonstrate how a population of embodied language users can develop their own vocabularies from scratch.

ISBN 978-3-944675-43-5

9 783944 675435