Chapter 1
Grounding Language through Evolutionary Language Games

Luc Steels[1,2]

This paper is the author's draft and has now been officially published as:
Luc Steels (2012). Grounding Language through Evolutionary Language Games. In Luc Steels and Manfred Hild (Eds.), *Language Grounding in Robot*, 1-22. New York: Springer.

Abstract This chapter introduces a new experimental paradigm for studying issues in the grounding of language and robots, and the integration of all aspects of intelligence into a single system. The paradigm is based on designing and implementing artificial agents so that they are able to play language games about situations they perceive and act upon in the real world. The agents are not pre-programmed with an existing language but with the necessary cognitive functions to self-organize communication systems from scratch, to learn them from human language users if there are sufficiently frequent interactions, and to participate in the on-going cultural evolution of language.

Key words: robotics, computational linguistics, whole systems AI, integrated AI, evolutionary language games, semiotic cycle

1.1 Introduction

This book addresses three challenges for getting us closer to intelligent robots.

The First Challenge: Integrated AI

The depth and breadth of knowledge accumulated for all the different aspects of intelligence has now grown to such an extent that nobody can still be an expert in the whole field. On the other hand, everybody agrees that robust, open-ended flexible intelligent robots, and particularly language communication with robots, can only be achieved if all the different subfields of AI, from sensorimotor intelligence to high

[1]ICREA Institute for Evolutionary Biology (UPF-CSIC) Barcelona, e-mail: steels@arti.vub.ac.be
[2]Sony Computer Science Laboratory Paris

level knowledge representation, reasoning and verbal interaction, are integrated into a single system. In other words, when we use a *whole systems approach*.

There has been promising integration work in the past, particularly in the early seventies such as the project Shakey at SRI (Nilsson, 1984), but more recently the quest for artificial systems that could plan and physically move around in the world, understand language and produce themselves goal-directed communication has slowed down. This book wants to put integration again on top of the AI research agenda for the following reasons:

1. The *symbol grounding problem* has been plaguing AI for a long time (Steels, 2008). The field has since the very beginning always had two research thrusts. The first one, focusing on segmentation, feature extraction, pattern recognition, and dynamic control, is primarily based on continuous mathematics. The second research line, focusing on symbolic representations and complex operations over these representations, is primarily based on discrete mathematics, algorithmic computer science and logic. The grounding problem is about how these two worlds can get connected. How can we go from the continuous real world with sensory and motor streams to the world of symbols and language and back? A whole systems approach is our best hope for tackling this problem because only that way can we tightly integrate components from sensorimotor intelligence with symbolic knowledge representation.
2. When one focuses on a single aspect of intelligence, for example computer vision or inference, there is a risk to make components more complex than they need to be. For example, a computer vision system may work hard to derive an accurate world model which may not actually be needed by the agent for the task he is currently engaged in. A parser may plough through heavy combinatorial search to come up with the best matching syntactic analysis whereas the meaning of the utterance could already be grasped from the words, context and on-going dialog. A planner may try to plan in fine-grained detail motor commands whereas it is much easier to leave this to low level dynamical processes that can exploit the physics of the body and the world. Taking a whole systems approach can therefore lead to *simpler solutions* which are easier to develop and employ.
3. Integrating all aspects of intelligence in a single system also helps to deal with *robustness*, one of the key challenges in intelligent robotics. For example, 100 % accurate speech recognition is not possible because of noise and speaker errors in articulation. But if other components are available, they can make up for this either by retrieving a word from the lexicon that matches partly or by ignoring the missing word and trying to grasp the meaning from the remainder of the utterance. Or, in dialog the speaker may describe an object which is partly occluded for the hearer but thanks to information from language the hearer can nevertheless make a reasonable guess about the type, location and visual characteristics of the object involved.
4. But the most important reason for a whole systems approach lies in the *learning opportunities* it creates for autonomous learning and development. Current machine learning methods are very powerful but they almost invariable require that a human trainer carefully creates annotated data and well defined restricted

learning goals. When different aspects of intelligence are integrated in a single system, one component can scaffold the learning conditions for another one and so we can achieve open-ended autonomous learning. For example, the acquisition of words can be greatly improved with syntactic bootstrapping, based on syntactic expectations created by partial parsing, or semantic bootstrapping, based on semantic expectations from partial understanding, context and further dialog. Feedback from language can help concept acquisition under the assumption that objects which are named differently are categorized differently by other agents.

The Second Challenge: Language Grounding on Robots

Human language-like communication with robots remains today a very distant goal. A few decades ago the problem was almost entirely on the side of robots. There were not enough physical robots to work with and the scarce robots that were available were unreliable, difficult to control and had only weak sensing capabilities. Also the computing power and electronics available for sensing and motor control had strict limitations. This situation has changed significantly over the past few years. There are now thousands of powerful robots in the world and their capacities in terms of embodiment, sensorimotor potential and computing power are quite sufficient for high level tasks. The intense activity around the Robocup and the new developments towards standardized components for robotics, such as ROS, are illustrative of this trend and it bodes well for future research. On the other hand, research on natural language processing appears not ready to exploit these new robotic capabilities. After promising work with systems like Shrdlu (Winograd, 1972) and Shakey in the early seventies, the ARPA speech understanding projects in the eighties (Klatt, 1990), and the Verbmobil project in the nineties (Walhster, 2000), the quest for artificial systems that could understand language and produce themselves goal-directed communication slowed down and research in computational linguistics became dominated by statistical language processing.

There is no doubt that the statistical approach has been very successful and is of practical use. Statistical language processing relies on a large corpus of example sentences (the larger the better) and on general purpose machine learning algorithms. It basically attempts to develop language models that predict the probability of a word occuring in a sentence given the previous words. This approach stands in contrast to the one explored in earlier deep natural language processing research which used sophisticated grammars based on linguistic theory and procedural semantics for the precise interpretation of meaning in terms of world models derived from sensing and actuating. Parsers try to extract rich grammatical structures of sentences before interpreting them and producers used sophisticated planning techniques to determine what to say and then map meaning into words and grammatical constructions.

The main reasons why statistical language processing became more popular are as follows:

1. Human languages are unlike programming languages in the sense that sentences are rarely fully grammatical. Often only partial fragments are communicated and errors in meaning, grammar use, word choice, or pronunciation are very common due to the speed with which utterances need to be produced. Consequently, parsers that rely on sentences being grammatical easily break down on real input. Statistical language processing handles this problem by being rather shallow in terms of the syntactic structures that are extracted, sometimes even relying only on sequential structure instead of hierarchy (Frank and Bod, 2011). Often these shallow structures are enough for tasks that are needed by search engines.
2. Grammars of human languages are extraordinarily complicated. It therefore became clear quite early in language processing research that it would be extremely hard to design grammars and lexicons by hand. Some form of automatic language learning is essential, and the most effective way to do so at the moment is to use statistical machine learning techniques.

But what if the goal is to use language for interacting with complex devices such as robots? Shallow parsing is not sufficient because the rich grammatical structures underlying sentences are there to help listeners grasp meaning. If we ignore this, we deprive ourselves of an important source of information. Lack of semantics or shallow semantics is too risky because it may lead to actions by the robot which are inappropriate or outright dangerous. Language production must rely on careful planning of meaning and this meaning needs to be the basis of sentence formulation as opposed to retrieving from memory sentence fragments that have tended to occur in similar circumstances. Most importantly it is also crucial that meaning gets grounded in the context through the sensorimotor apparatus of the robot, and unless we have corpora that contain vast amounts of data on grounded interactions, it is not possible to apply statistical machine learning techniques.

The Third Challenge: Artificial Cultural Evolution

Language is a system that is in constant flux. Whenever humans interact there is the possibility that a new word gets invented or the meaning of an existing word gets stretched, a grammatical construction may be used in an 'odd' way but this may further propagate and become the norm, words gain or shift to new syntactic categories, and so on. This suggests that if we want to see grounded communication between humans and robots, these robots need to be flexible enough to participate in this ever evolving cultural system, and this creates an intriguing convergence between research on grounded language in robots and studies in the cultural evolution of language, a topic which has recently come to the foreground in many disciplines interested in language, from evolutionary biology and anthropology to linguistics.

There has been a large amount of interdisciplinary activity in language evolution research lately (see for example the bi-annual conferences starting from Hurford et al, 1998), but there is still no widely accepted *explanatory* theory of the cultural evolution of language of the same stature as current theories of biological evolution. Such a theory should on the one hand propose some general principles by which

languages can become more complex and it should on the other hand make concrete proposals for the cognitive functions and interaction patterns that are needed to see the emergence of specific linguistic forms and the conceptualizations they express, for example a tense-aspect system, argument structure realization, a basic color term vocabulary, a system of quantifiers, an internal agreement system, etc. (Steels, 2011).

Three concrete research paradigms are currently being used for working out and testing theories of cultural language evolution. The first paradigm takes primarily a *linguistic* point of view. It starts from concrete data of language change as found in the historical record or in situations where special circumstances have lead a community of people to develop a new language (Mufwene, 2001) and attempts to find the kind of cognitive operations and strategies that underly the observed grammaticalization processes (Heine, 1997). The second paradigm follows a *psychological* approach. It proposes to perform 'semiotic experiments' with human subjects in order to find out what kind of strategies they employ for coming up with a new communication system (Galantucci and Garrod, 2010). These experiments typically put humans in challenging situations where they have to interact without being able to use their existing language. Remarkably, they are able to build up new communication systems rather quickly, even though some people are much better than others (Galantucci, 2005).

The third paradigm, which interests us here, is based on *modeling* because that will yield a mechanistic theory of cultural language evolution that we can then apply to robots. A particular proposal for the cognitive functions, ecological conditions and interactions patterns that are needed for language is operationalized and then used to simulate the emergence of language systems in populations of artificial agents. This approach started in the early nineties (see an early review in Steels, 1998) and has flourished considerably during the past decade (Lyon et al, 2007; Minett and Wang, 2005; Nolfi and Miroli, 2010). The language systems that emerge in these computational experiments are of course never equal to English or Hindi, given the historical contingencies that play a role in normal cultural language evolution, however, by using strategies reconstructed from human languages or by scaffolding the experiment with a vocabulary or partial grammar from an existing human language, the artificial languages are closer to a human source language, which makes the experiment more relevant and the evolution easier to follow.

Even if one chooses the synthetic route, there are still many different ways to model cultural language evolution. In this book we will explore a theory of cultural language evolution based on linguistic selection and self-organization (Steels, 2012b). This theory argues that situated communication between embodied individuals plays a major role in shaping, selecting and self-organizing language systems. We therefore need to employ a modeling approach which has communication (and not only vertical transmission) at its core and we will therefore frame communication in terms of *language games*, following up on proposals originally made by Wittgenstein (1953). The study of evolutionary language games started from timid beginnings in the early nineties but right now dozens of experiments have been performed for many different aspects of language, ranging from perceptually grounded

vocabularies to grammar (Steels, 2012a). The rest of this chapter discusses the notion of a language game in more detail. The next parts of the book discuss the different components needed to operationalize language games. And the book ends with some examples of integrated experiments.

1.2 Language Games

A language game is embedded in a cooperative activity in which communication is useful. It attempts to model situated dialog in contrast to the isolated sentences that are commonly used today in formal linguistics. Consequently, a language game requires a population of individuals, a context, and a communicative purpose, so that pragmatics is part of the modeling effort from the start. Wittgenstein gives the example of the builder and his assistant. The builder requires stones of a certain kind to be handed to him and hence they need a language for talking about types of stones. A language game involves joint attention to some objects and activities in the immediate context and a routinized turn taking interaction that may involve symbolic communication as well as physical actions or gestural communications. A language game takes place based on a specific embodiment that grounds the participants in the world and within a particular environment and ecological setting. These factors co-determine what kind of communicative goals participants may have and what kind of concepts they might be able to use. For example, if the environment contains only black and white objects or if speakers and hearers are all color blind, a hue-based color language cannot (and should not) emerge. If the world is such that objects do not move, a language for talking about events and their temporal structure is irrelevant.

1.2.1 Examples

Here is the scenario of a typical language game called the *Naming Game*, first introduced by Steels (1995). The Naming Game is a *Game Of Reference*, the speaker attempts to draw the attention of the hearer to an object in the world by naming a characteristic feature of the object. If the object is a specific recognizable individual, then a proper name can be used. It is also possible to name colors, shapes, sizes, as long as they are distinctive.

The game is played by a population `P` of agents and involves a world `W` consisting of objects. Each object is characterized by a point in an n-dimensional feature space. For example, the color of an object is a point in the three-dimensional color feature space with the dimensions red-green, yellow-blue, and lightness. Two members are randomly selected from the population to take on the roles of speaker and hearer respectively. A context `C` is established which contains a subset of the objects in the world `W`. Then the following interaction takes place:

1. The speaker selects one object out of the context, further called the topic `T`.
2. The speaker finds the distinctive category for the object and names this category.
3. The hearer looks up which object is associated with this category in his memory and examines the context to find out whether there is an object which has this distinctive characteristic.
4. The hearer then signals to the speaker which object was intended according to him, for example by pointing.
5. The speaker checks whether the hearer selected the same object as the one he had originally chosen.
   a. If they are the same, the game is a success, and the speaker signals this outcome to the hearer.
   b. If they are different, the game is a failure. The speaker signals this outcome and then points to the topic he had originally chosen.

A 'solution' to the game is a particular language strategy that agents can use to build up a shared set of distinctive categories and names for these categories such that they are successful in the game. The agents do not know these categories nor their names in advance. The language strategy contains diagnostics and repairs for concept formation and concept acquisition and routines for concept alignment, as well as diagnostics and repairs for vocabulary formation and vocabulary acquisition and routines for vocabulary alignment.

There are always many language strategies possible for a language game depending on the specific cognitive functions that are used for playing the game, for learning an existing language system or forming one and particularly for alignment. Each of these strategies has different performance characteristics which can be systematically investigated for the same experimental parameters in order to find the 'linguistic niche' of a strategy. For example, for the Naming Game, we can change the number of objects in the context, the relevant categorial dimensions, how close objects are within the feature space used to form categories, the size of the population, whether the world is dynamic or static, whether the population is dynamic or static, and so on.

Another class of language games are *Action Games*. The speaker tries to get the hearer to do a particular action, such as turn around, raise the left arm, pick up an object or go to a particular location in the room. Action games are particularly useful for studying how names for actions can emerge in a population. One type of Action Game are *Posture Games* where the speaker does not describe the action but the bodily posture that he expects the hearer to adopt, such as "arms raised" or "sitting" (Steels and Spranger, 2012).

The Posture Game is again played by a population `P` of agents which have a physical body which they can control to execute actions in the world and a sensory system to get feedback about their own actions (proprioception) and to observe actions by others (through vision). Two members are randomly selected from the population to take on the roles of speaker and hearer respectively.

1. The speaker chooses a posture from his inventory of postures.

2. The speaker retrieves the name for this posture in his vocabulary and transmits that to the hearer.
3. The hearer retrieves the posture by looking up the name in his own vocabulary and evokes the motor behavior that could achieve this posture.
4. The speaker observes the posture adopted by the hearer and checks whether it fits with the prototypical visual body-image of the posture he had originally chosen.
   a. If this is not the case, the speaker signals failure. The speaker activates his own motor behavior for achieving this posture in order to repair the communication, so that there is an opportunity for the hearer to learn the speaker's name for this posture.
   b. Otherwise the speaker signals success.

Again, this game definition is just a setting. The solution takes the form of concrete proposals for language strategies by which speaker and hearer can invent, learn, and coordinate names for postures, as well as learn the visual image schemata of a posture, the motor control programs to achieve the posture, and the associations between the two. Language games almost always raise many fundamental issues in cognitive science. For example, to be able to play the posture game, the players need to have a mirror system so that they can recognize actions of others in terms of their own actions (Rizzolatti and Arbib, 1998). But there are also recurrent problems that come up in almost every game. For example, both the Naming Game and the Action Game require that the population establishes lexical conventions, even though the game script, the conceptual system, and the strategies of the agents are different.

1.2.2 The Semiotic Cycle

Playing a fully grounded language game requires that speakers and hearers go through the *semiotic cycle* shown in Figure 1.1. The relevant processes take place against the background of turn-taking and attention sharing behaviors and scripts monitoring and achieving the dialog.

The processes relevant for the speaker are:

1. *Grounding:* The first set of processes carried out by both the speaker and the hearer must maintain a connection between the internal factual memory and the states and actions in the world that dialog partners want to talk about. They include segmentation, feature extraction, object recognition, event classification, object tracking, object manipulation, etc.
2. *Conceptualization:* The second set of processes must select what needs to be said and then conceptualize the world in a way that it can be translated into natural language expressions which satisfy the communicative goal that the speaker wants to achieve (Talmy, 2000). For example, if we say "the car is in front of the tree", we have conceptualized the tree as having a front which is directed towards us, and the car as being in between ourselves and this front.

3. *Production:* (also known as verbalization or formulation; Levelt, 1989): This set of processes takes a semantic structure and turns it through a series of mappings into a surface form, taking into account the lexical, grammatical, morphological and phonological conventions of the language as captured by various constructions.
4. *Speech Articulation:* This set of processes renders a sentence into the fast movements of the articulatory system required to produce actual speech and gestures.

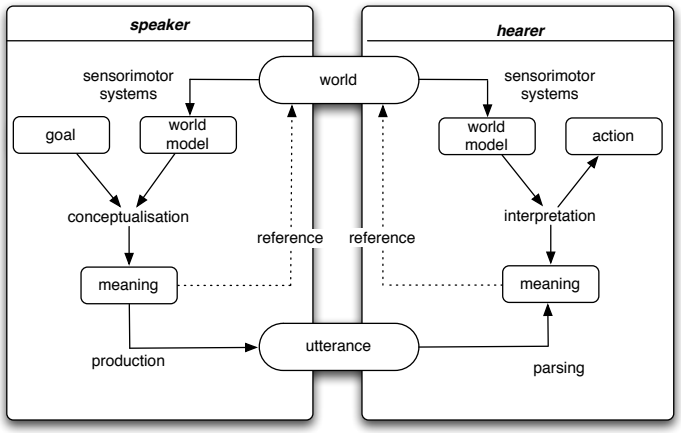

Fig. 1.1 The semiotic cycle summarizes the main processes that the speaker (left) and the hearer (right) go through. Sentence parsing and production is only one of the activities within this cycle.

The processes relevant for the hearer are:

1. *Speech Recognition:* The speech signal needs to be processed through a battery of signal processing and pattern recognition processes to get a reasonable set of hypotheses about the speech elements that might be present.
2. *Parsing:* The hearer uses these data to reconstruct as well as possible the meaning of the utterance that is transmitted by the speaker. Again, this process is highly elaborate due to the complexity of natural language and the presence of ambiguities.
3. *Interpretation:* The hearer must then confront the meaning resulting from the parsing process with his or her own factual memory of the world and understanding of the dialog context in order to find a correct interpretation of the utterance that fits with his or her own expectations and observations. For example, the hearer must retrieve the object in the scene that the speaker wanted him or her to pay attention to.
4. *Grounding:* The hearer must therefore also maintain a connection through perception and action between his or her internal factual memory and the states of the world, possibly including the mental states of the speaker.

On top of this cycle are the meta-level processes that perform diagnostics and repairs and the alignment routines which are executed based on the outcome of a game (Steels, 2012b).

1.3 Language Game Experiments

It is almost impossible to understand the consequences of a proposed language strategy, particularly for complex strategies with rich conceptualizations and complex grammar, which is the main reason why computer simulations and robotic experiments of language games are necessary. A language game experiment is intended to explore some aspect of language in depth, both from the viewpoint of how grounded communication is possible and how it could have evolved. For example, a concrete experiment might focus on 'fuzzy' quantifiers, such as "some", "a few", "almost all", etc. (Pauw and Hilferty, 2012).

Setting up an experiment requires defining (i) an environment and an embodiment of the agents, (ii) a communicative goal within this environment, which is assumed to be part of a larger cooperative interaction that is usually not modeled, (iii) a game script, that not only deals with routine use but also with feedback and repair, and (iv) a population structure. The challenge is then to come up with an operational language strategy so that agents endowed with this strategy cannot only routinely use the aspect of language under study but also learn it from others and invent and coordinate a language system from scratch given this strategy. Some experiments go one step further and provide agents only with the necessary components to evolve new language strategies and try them out in a selectionist semiotic dynamics, but the goal of the experiment is the same, namely to arrive at a shared communication system that is adequate for the embodiment, environment, and cooperative goal of the language game.

If carried out properly, language game experiments are able to establish in an objective fashion whether a proposed strategy works, what its performance characteristics are, how different alternative strategies compare and might compete with each other in a selectionist dynamics and how new strategies could form by recruitment or by exaptation from an existing strategy.

Many different types of language games can be imagined and realistic human dialogs always involve many different games at the same time. For the purposes of methodical scientific investigation, it is however useful to focus on a single language game so that the linguistic phenomena of interest can be studied in isolation. The landscape of possible language games is vast and only very small patches have been explored so far in concrete experiments, but this has already yielded a body of technical expertise, tools, and fundamental insights that carry across different language games.

1.3.1 Environment and Embodiment

Artificial agents in language evolution experiments can be in software, operating in a virtual world, or they can take the form of physically instantiated robotic agents which move about and act in the real world. When the agents are operating purely in a virtual environment, we talk about *non-grounded language games*, otherwise we talk about *grounded language games*. Using robots is highly relevant for language evolution research because language undoubtly originated for talking about the real world as experienced by sensors and embedded in cooperative physical actions in the world. Abstract concepts are usually based on the metaphorical extension of embodied concepts and language (Lakoff, 1987), which suggests that we should attempt to explain the origins of perceptually grounded language first. Moreover, language is a non-modular system. None of the components needed for language (world perception, speech, grammatical processing, conceptualization) is reliable on its own. To achieve robust production and comprehension of language requires therefore a *whole systems approach* in which one component can make up for the weakness of another one and different components need to co-evolve to explain how the whole system could emerge.

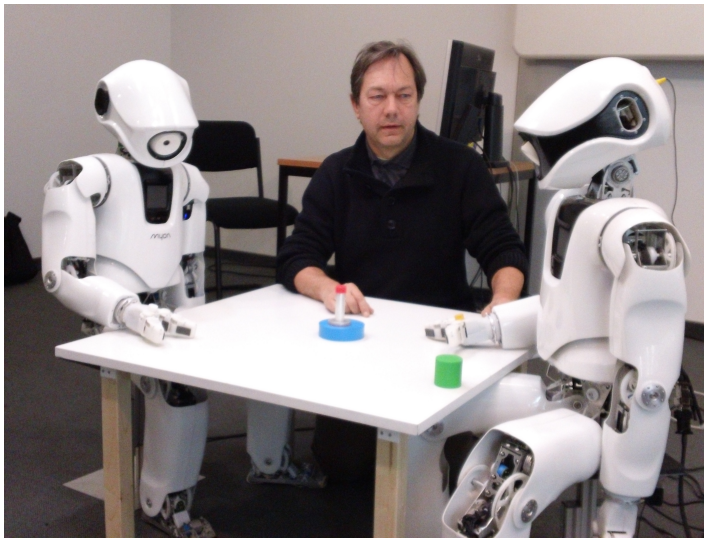

Fig. 1.2 Experimental setup for language games played with embodied agents. There are at least two autonomous robots (in this case MYON robots) and possibly a human experimenter (in this case the author of this chapter).

The earliest embodied language game experiments already took place in the late nineties. They used very basic 'cybernetic' robots, built out of Lego blocks, simple sensors for infrared and touch, and basic motor control (Steels and Vogt, 1997), or pan-tilt cameras oriented towards a wall with geometric images so that large

series of 'Talking Heads' experiments could be done (Steels et al, 2002). As robotics technologies matured, robots became more complex, first using a typical circular base, then 4-legged robots such as the AIBO (Steels and Kaplan, 2001; Steels and Loetzsch, 2008) and more recently humanoid robots (as shown in Figure 1.2) with much more powerful hardware for sensorimotor intelligence and computation. The experiments reported later all use a humanoid platform.

To have any scientific value, we always need to do several experimental runs to get statistically significant results. Unfortunately, robots are still a very scarce resource today and it takes quite a while (usually on the order of several minutes) to perform a single grounded language game.

We overcome the first limitation through a *'teleporting' facility* (McIntyre et al, 1999) that was pioneered in the Talking Heads experiment (Steels and Kaplan, 2002). The internal states of the agents (their perceptual system, vocabularies, grammars, concepts, memory of past interactions, etc.) are stored as data on a central server. When a game has to start, two agents are selected from the population to play the role of speaker and hearer and their state is downloaded into the robots so that they become fully autonomous. No other interaction is possible between these robots except through the world. After the game has ended, the changed software state of the two agents is uploaded again to the server and statistics are collected.

The teleporting facility makes it possible to have a large population, even if there are only a few physical robots available. The Internet can be used to send around agent states from one physical location to another and to have several games at the same time. The Talking Heads experiment thus managed to handle thousands of embodied agents which performed hundreds of thousands of language games, in different locations in Belgium, France, the Netherlands, the United Kingdom, and Japan.

We can overcome the second limitation by systematically *storing* the complete sensorimotor states of agents as they engage in situated grounded interactions, which is useful for analysis anyway, and to use these data later as input for other games, possibly by other agents. This makes it not only possible to do a much larger number of experiments. It also becomes possible to test alternative solutions on the same sensorimotor data. The only aspect of a game which is then no longer 'real' are the actions taken as a consequence of a communication, for example pointing gestures, but these actions can be deduced and simulated. For example, the direction of pointing of the hearer can be projected into the visual scene perceived by the speaker and the speaker can formulate and communicate feedback about success and failure directly and even correct the hearer again with a new simulated pointing gesture.

Given this approach, it becomes furthermore possible to do experiments where the sensorimotor data of the agents is entirely based on a simulated virtual world. For example, experiments in color language games can use a database of perceptual color data or pictures from real world scenes. However, the more we use simulated worlds the more we move away from reality and therefore risk to make wrong assumptions about what perceptual systems could possibly provide, and what actions

in the world (such as pointing gestures) could effectively be performed and recognized.

There are a number of dimensions for possible environments in language games that have an important impact on whether a particular language strategy will work or not:

1. *Static vs. dynamic*: There is obviously a huge difference whether the objects in the world do or do not move, and whether the agents move or not. The differences concern partly the sensorimotor processing that agents will have to be capable of to recognize moving objects and event structure, and which actions they can perform. But also in terms of conceptualization and language there are huge differences. Many aspects of language, such as temporal structure expressed through tense or aspect, are irrelevant in static worlds.
2. *Open vs. closed*: An *open world* is one where new objects may appear at any time or when the characteristics and activities of objects may change. In a *closed world* the set of possible objects and their characteristics are fixed at the beginning of the experiment. When the world is open, many machine learning methods that rely on supervised learning or on scanning of a complete example set no longer work. Instead, all components of a language strategy have to be incremental.
3. *Degree of stochasticity*: When conducting experiments with physically instantiated agents in real environments there are many sources of error, noise, or just stochasticity. This occurs both for the speech signal and speech articulation, for the behavior of the objects in the environment, and for the non-verbal feedback that is provided. For example, pointing gestures by robotic agents unavoidably include error either in producing the gesture or in recognizing it, and this may have a significant impact on which strategy is used or on the behavior of a strategy (Steels and Kaplan, 1998b).

1.3.2 Communicative Goal

The communicative goal determines whether a communication succeeds or fails. The goal is itself something beyond communication, i.e. something that the speaker wants to achieve, such as draw attention to an object in the environment, get an object, or obtain information. Communication is therefore embedded in cooperative interactions even if they are not modeled explicitly in most language game experiments.

Typologies of communicative goals have been proposed in the literature on speech acts. For example, Austin (1975) has made a distinction between illocutionary acts, which express or imply desires, locutionary acts, which describe or inquire about the situation itself, and perlocutionary acts, which are like commands where the speaker wants the hearer to do something. Searle (1969) has made a distinction between different types of illocutionary speech acts, including representatives, where the speaker commits to the truth of a proposition, directives, where the

speaker commands an action, commissives, where the speaker makes a promise of a future action, etc.

In language game experiments, the following communicative goals have so far been explored:

1. *Game of reference*: This is a language game where the speaker draws the attention of the hearer to an object in the environment, which is usually present in the immediate context. From a semantic point of view, games of reference imply a *Discrimination Game* because the speaker has to find a unique property or set of properties that distinguishes the object from others in the present context. Feedback on communicative success can take place by a pointing gesture of the hearer, and a repair can again take the form of a pointing gesture by the speaker. Many variants of the Naming Game exist depending on the complexity of the language or the uncertainty involved. Here are two important cases:
   a. *Naming Game*: The term Naming Game is employed for settings where there is no ambiguity about the semantic domain that is used to discriminate the topic with respect to the other objects in the context and therefore when a single word can be used to name the distinctive category that identifies the referent. The Naming Game is particularly useful for studying the semiotic dynamics underlying conventionalization (Baronchelli et al, 2006), and the co-evolution between the formation of perceptually grounded categories (such as colors or spatial relations) and an emerging vocabulary (see for example Bleys, 2012).
   b. *Guessing Game*: The complexity of a Game of Reference increases substantially when there is more than one possible distinctive category for the topic, because this introduces much more uncertainty for the hearer about the meaning of an unknown word. This is Quine's "Gavagai" problem (Quine, 1960). The biggest challenge in Guessing Games is to combat the combinatorial explosion of possible meanings (Wellens and Loetzsch, 2012).
   c. *Multi-Word Guessing Game*: As soon as there are multiple meaning dimensions and more than one meaning is required to identify the referent, there is the possibility to use more than one word, and this raises the next level of difficulty: How to distribute meanings over different words?
2. *Action game*: This is a language game where the speaker wants the hearer to perform a particular action. From a semantic point of view, an Action Game again implies a Discrimination Game as well because the speaker has to find unique distinctive properties of the action that he wants the hearer to perform. On the other hand, feedback on communicative success can no longer be based on pointing gestures but must be based on recognizing the intended effect of the action, and this implies the speaker and hearer develop a mirror system linking visual images of actions or their outcome with motor control programs to achieve them. The speaker can repair a dialog by performing the action himself.
   a. *Posture game*: In a Posture Game, the speaker describes a particular posture, e.g. “lie”, and expects the hearer to perform the necessary motor control ac-

tions to achieve that posture (Steels and Spranger, 2012). Posture games are of great interest because many languages use postures by extension for grammatical meanings, such as aspect, or transfer postures from the action domain to the spatial domain which is common in Germanic languages like Dutch (Spranger and Loetzsch, 2009).

b. *Movement game*: A Movement Game describes the action that the speaker wants the hearer to perform, such as "stand up". Often the objects involved in the action and their roles need to be described as well, as in "pick up the big red block", which then brings in the need for games of reference to identify the objects, and expression of argument structure. One of the main challenges of actions involving objects in language game experiments is that this requires greater motor control capacities from the robot.

3. *Description Game*: In a Description Game, the speaker describes a particular situation. This task is in itself too open-ended to define success criteria and provide adequate feedback and repair. Variants of this game define further constraints and hence different types of communicative success.

   a. *Factual Description Game*: This game is successful if the hearer agrees that the description produced by the hearer is valid for the situation that speaker and hearer experienced or are experiencing. It is for example used in van Trijp (2012) to study how case markers can emerge.
   b. *Differential Description Game*: This game involves two different scenes, for example shown as two separate videoclips, and the hearer has to select which of these two fits best with the description produced by the speaker. This game implies again a Discrimination Game to find enough unique characteristics to uniquely identify which scene should be chosen. A Differential Description Game has been used for example in Gerasymova et al (2009) to study the emergence of aspect and it is frequently used in psycholinguistic 'preferential looking' experiments, particularly with very young children.

The complexity of the language in each of these games can vary considerably from single words to multi words and from morphological markers or particles attached to words to item-based constructions and then fully hierarchical grammatical descriptions. It is certainly possible to identify many more communicative goals and many additional types of games. The list above enumerates only the language games that have been studied in depth so far. One must keep in mind that setting up and conducting concrete experiments takes years of work, often by a team of technically highly skilled researchers. Even for the games listed here, there are still many issues that are left unexplored.

1.3.3 Population Structure

A language game experiment not only involves an embodiment and an environment but also a population of agents, which can be physically embodied or not. This population may have a number of characteristics that play an important role in the semiotic dynamics observed in the experiment. Moreover not all language strategies are effective for all population structures.

1. *Structured vs. Unstructured Populations*: In the case of an *unstructured population*, all agents have an equal chance to interact with any other agent. For a *structured population*, agents are either divided up into different subsets or there is some network structure imposed on the population which influences the probability with which two agents interact (Baronchelli et al, 2006). Structured populations are of course the rule in human languages, because language users have different geographic or social characteristics and some speakers have much greater influence than others because they reach large groups of hearers.

2. *Active vs. Passive*: Agents can differ with respect to how much they are actively participating in the invention of the language or just passively adopting what others have invented. Indeed, it is well known that some human language users have a more creative approach to language use than others, and first or second language learners usually adopt the attitude that they must accept the norm rather than invent their own. The attitude towards alignment also differs. Some language users tend to strongly align, even within the course of a single dialog, whereas others do not align at all and tend to stick with their own inventions. These different attitudes are clearly observable within experimental semiotics experiments with human subjects (Galantucci, 2005).

3. *Coordination vs. Iterated Learning*: In the case of a fully *coordinated population*, all agents participating in a language game take turns and have equal rights in shaping the language. This means that they can always act both as speaker or hearer and as speaker they can invent new words, coerce words into new functions, apply grammatical constructions to new situations, introduce new categories, etc. Coordination models rely critically on alignment to achieve convergence in the population (Garrod and Anderson, 1987). In *Iterated Learning* models (Brighton et al, 2005) there is a strict division between a subpopulation of tutors and a subpopulation of learners. Learners only act as hearers and have no influence over the language of the tutors. Iterated learning models are useful because they isolate the potential role of the learner in language evolution, but this restriction also leaves out other factors that are crucial in the selectionist theory explored here, in particular the role of communicative success.

4. *Dynamic vs. Static Populations*: A *static population* has a fixed set of members of a given size which does not change in the course of an experiment. A *dynamic population* shows a turn over in its constitution, with some members entering the population without any knowledge of the existing language systems or strategies and other members leaving, taking with them the knowledge that they acquired. Dynamic populations can be used both for Coordination and Iterated Learning models. In Iterated Learning models, the learners become tutors for the next cohorte of

learners, creating a chain. In coordination models, new agents entering the population participate with full rights but they will have almost no influence on the total language. If the population in- and out-flux is very high then existing language structures may collapse and new paradigmatic choices of language systems or even new strategies may have to emerge (Steels and Kaplan, 1998a).

1.3.4 Scaffolding

Implementing a complete semiotic cycle is enormously challenging, particularly with grounded autonomous robots, hence it is customary to *scaffold* certain aspects of the cycle, depending on the research question one wants to ask:

1. *Scaffolding speech*: Although there have also been experiments in the evolution of speech sounds (de Boer, 1999; Oudeyer, 2005), usually the speech actions of a language game are scaffolded, in the sense that agents communicate using symbolic transmission of the utterance. The symbols are constructed by the agents themselves by the random assembly of syllables. This approach makes it possible to better focus on issues of lexicon, grammar or concept formation.

2. *Scaffolding world models*: It is possible to scaffold all aspects related to perception and physical behavior by providing the speaker and the hearer directly with a world model that is already in a conceptual form. World models are often defined using an ontology with which possible instances can be generated. When the world model is scaffolded we speak about *non-grounded language games*, otherwise about *grounded language games*.

3. *Scaffolding conceptualization*: It is possible to scaffold the conceptualization process, which implies that the speaker is directly given a meaning to express. In the simplest case (known as *direct meaning transfer*), the hearer is given the same meaning and the game is a success if the hearer has been able to reconstruct the meaning that the speaker had to produce. Direct meaning transfer is not a very desirable method because it ignores the fact that communicative success is the goal of a language game, and success may be reached even if the meanings of speaker and hearer are not the same. Direct meaning transfer also ignores the possibility that the conceptual system and the linguistic system co-evolve, which is one of the central tenets of the selectionist theory explored here. So none of the language experiments discussed later in this book uses direct meaning transfer, although the technique is useful in a first preparatory phase of an experiment.

4. *Scaffolding the language system*: It is often useful to test a particular experimental set up by providing all agents in the population with a particular language system and establish a baseline performance. When the language system is a reconstruction from an existing human language, then it becomes easier to follow the grammaticalization process. Usually those aspects of language which are not being studied but nevertheless essential to get a verbal interaction operational are scaffolded. For example, the study of determiners can only take place when there are nominal phrases, but the semantics and grammar of such phrases can be scaffolded.

5. *Scaffolding the language strategy*: Finally it is useful to scaffold the language strategy the agents are using, instead of having the strategy evolve through recruitment or exaptation. When strategies are based on reconstructions from human language more realism and hence more relevance is obtained.

1.3.5 Measures

It is standard practice to simulate games sequentially and use only two agents per game, even though it is entirely possible (in fact each proposed strategy should allow it) that several agents participate in a single game or that many language games go on at the same time. It is also standard practice to monitor communicative success as the games unfold over time and plot the running average. Communicative success is not determined by an outside observer but always by the agents themselves, and particularly by the speaker.

Usually several experiments are carried out with the same world conditions and the running average for different experiments is plotted with standard deviation. Measurements are often scaled with respect to the number of games played per agent so that results are represented as scaled with respect to population size. Figure 1.3 shows an example (taken from Steels and Loetzsch, 2012). It shows (left y-axis) the communicative success of a particular language strategy for the Naming Game in a population of 10 agents.

Other measures are often taken as games unfold, such as the average number of categories in the ontologies of the agents, the central points of prototypes, the emerging set of Aktionsart distinctions or the contours of spatial categories, the number of semantic roles and syntactic cases, the set of markers playing a role in agreement, the number of abstract argument structure constructions, and so on. It is also very useful to track the evolution with respect to selectionist criteria, such as expressive adequacy, cognitive effort, learnability, and social conformity, or to monitor how the preference for paradigmatic language system choices or strategies is changing over time.

1.4 Conclusion

The past decade a new methodology has emerged to develop scientific models of cultural language evolution and these models are beginning to provide important insights into how we can develop grounded language interactions with real robots. The methodology is based on designing and implementing artificial agents so that they are able to play language games about situations they perceive and act upon in the real world. The agents are not pre-programmed with an existing language but only with the necessary cognitive functions to self-organize a communication system without human intervention. This chapter discussed what kind of language

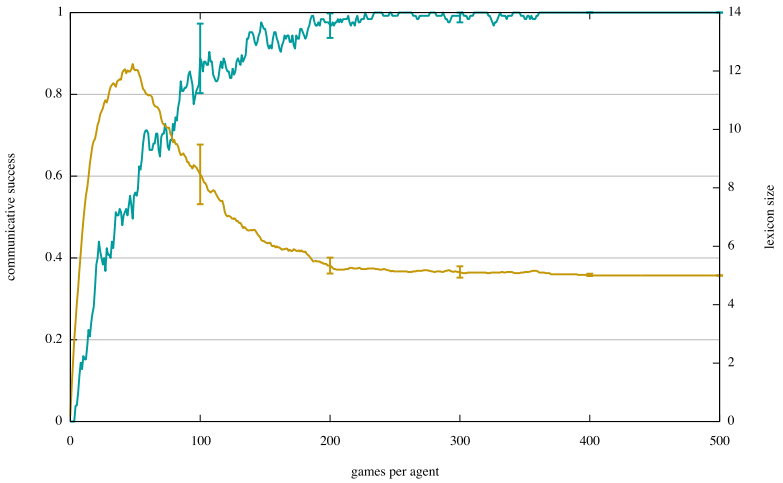

Fig. 1.3 Semiotic Dynamics of the Non-Grounded Naming Game with 10 agents self-organizing a vocabulary to name 5 unique objects using proper names. The number of games per agent is shown on the x-axis. Both the running average of communicative success (left y-axis) and of the average vocabulary size (right y-axis) are shown. The agents reach 100 % communicative success and lexical variation gets damped so that an optimal vocabulary of 5 names, one for each individual object, emerges.

games have been researched, the different aspects of an experiment, and the kinds of methodological steps that are being taken to make language game experiments as relevant as possible to research into the origins and evolution of human languages as well as human-robot communication.

Evolutionary language game experiments often take the extreme stance to eliminate human influence altogether, in order to focus entirely on how new meanings, new words, and new grammatical constructions may arise. Of course, in human societies there is always already a very rich language and human children or second language learners acquire this language before they become creative themselves and extend the language to deal with new communicative problems that they encounter. So it seems initially that research on evolutionary language games deviates from the goal of getting human-robot interaction. But this is not the case.

First of all, in evolutionary language games the agents must also be able to learn new concepts and language from other agents through situated embodied interactions, otherwise inventions would not propagate. Consequently, these very skills can be applied as well to learn from humans instead of other robots. We must keep in mind however that humans are a very different species from robots. Their sensorimotor capabilities are very different and their intelligent systems are vastly more complex than what we can currently conceive and built. On the other hand, the per-

ceptual and conceptual structures of two agents/humans do not necessarily have to be equal to have successful communication.

Second, we can make the experiments more relevant to human languages by starting from linguistic phenomena that are actually observed in those languages. For example, we might be interested in spatial language, and then start by a reconstruction of the language strategies used in an existing human language. Several examples of this approach are discussed in Steels (2012a) and in the case studies in the final part of this book.

Acknowledgements

Research discussed in this chapter was conducted at the Sony Computer Science Laboratory in Paris and the Artificial Intelligence Laboratory of the Free University of Brussels (VUB). We are indebted to Mario Tokoro, president of the Sony Computer Science Laboratories, for his continuing long term support of this research direction. Project funding was augmented by a variety of sources, including the FP6 EU project ECAgents and the FP7 EU project ALEAR. The experiments reported here require a large highly competent and dedicated team and the author is indebted to the many graduate students and collaborators who participated in creating the technological platforms and carrying out the experiments reported here. The author is also indebted to the Wissenschaftskolleg in Berlin for the opportunity to lay the foundations for the present work.

References

Austin J (1975) How To Do Things with Words. OUP, Oxford

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

Bleys J (2012) Language strategies for color. In: Steels L (ed) Experiments in Cultural Language Evolution, John Benjamins, Amsterdam

de Boer B (1999) Self-organisation in vowel systems. PhD thesis, Vrije Universiteit Brussel, Brussels

Brighton H, Kirby S, Smith K (2005) Cultural selection for learnability: Three principles underlying the view that language adapts to be learnable. In: Tallerman M (ed) Language Origins: Perspectives on Evolution, OUP, Oxford, p Chapter 13

Frank S, Bod R (2011) Insensitivity of the human sentence-processing system to hierarchical structure. Psychological Science 22(6):829–834

Galantucci B (2005) An experimental study of the emergence of human communication systems. Cognitive Science 29(5):737–767

Galantucci B, Garrod S (2010) Experimental semiotics: A new approach for studying the emergence and the evolution of human communication. Interaction Studies 11(1):1–13

Garrod S, Anderson A (1987) Saying what you mean in dialogue: A study in conceptual and semantic coordination. Cognition 27:181–218

Gerasymova K, Steels L, van Trijp R (2009) Aspectual morphology of Russian verbs in Fluid Construction Grammar. In: Taatgen N, van Rijn H (eds) Proceedings of the 31th Annual Conference of the Cognitive Science Society, Cognitive Science Society

Heine B (1997) The Cognitive Foundations of Grammar. OUP, Oxford

Hurford J, Studdert-Kennedy M, Knight C (eds) (1998) Approaches to the Evolution of Language: Social and Cognitive Bases. Edinburgh University Press, Edinburgh

Klatt D (1990) Review of the arpa speech understanding project. In: Klatt D (ed) Readings in Speech Recognition, Morgan Kaufmann Publishers Inc., San Francisco, Ca

Lakoff G (1987) Women, Fire, and Dangerous Things: What Categories Reveal about the Mind. The University of Chicago Press, Chicago

Levelt W (1989) Speaking. MIT Press, Cambridge MA

Lyon C, Nehaniv C, Cangelosi A (eds) (2007) Emergence of Language and Communication. Lecture Notes in Computer Science, Springer Verlag, Berlin

McIntyre A, Steels L, Kaplan F (1999) Net-mobile embodied agents. In: Proceedings of Sony Research Forum 1999, Tokyo

Minett JW, Wang WSY (2005) Language Acquisition, Change and Emergence: Essays in Evolutionary Linguistics. City University of Hong Kong Press, Hong Kong

Mufwene S (2001) Competition and selection in language evolution. Selection 3(1)

Nilsson N (1984) Shakey the robot. In: SRI Technical note 323, SRI International, Menlo Park, CA

Nolfi S, Miroli M (eds) (2010) Evolution of Communication and Language in Embodied Agents. Springer Verlag, Berlin

Oudeyer PY (2005) The self-organization of speech sounds. Journal of Theoretical Biology 233(3):435–449

Pauw S, Hilferty J (2012) The emergence of quantification. In: Steels L (ed) Experiments in Language Evolution, John Benjamins, Amsterdam

Quine W (1960) Word and Object. MIT Press, Cambridge Ma

Rizzolatti G, Arbib MA (1998) Language within our grasp. Trends in Neurosciences 21(5)

Searle J (1969) Speech Acts. Cambridge University Press, Cambridge Ma

Spranger M, Loetzsch M (2009) The semantics of sit, stand, and lie embodied in robots. In: Cognitive Science 2009

Steels L (1995) A self-organizing spatial vocabulary. Artificial Life 2(3):319–332

Steels L (1998) Synthesising the origins of language and meaning using coevolution, self-organisation and level formation. In: Hurford J, Studdert-Kennedy M, Knight C (eds) Approaches to the Evolution of Language: Social and Cognitive Bases, Edinburgh University Press, Edinburgh, pp 384–404

Steels L (2008) The symbol grounding problem has been solved. so what's next? In: Glenberg A, Graesser A, de Vega M (eds) Symbols, Embodiment and Meaning, OUP, Oxford, pp 506–557
Steels L (2011) Modeling the cultural evolution of language. Physics of Life Reviews doi:10.1016/j.plrev.2011.10.014
Steels L (ed) (2012a) Experiments in Cultural Language Evolution. John Benjamins Pub., Amsterdam
Steels L (2012b) Self-organization and selection in language evolution. In: Steels L (ed) Experiments in Cultural Language Evolution, John Benjamins, Amsterdam
Steels L, Kaplan F (1998a) Spontaneous lexicon change. In: Proceedings COLING-ACL 1998, Morgan Kaufmann, San Francisco, CA, pp 1243–1250
Steels L, Kaplan F (1998b) Stochasticity as a source of innovation in language games. In: Adami C, Belew RK, Kitano H, Taylor CE (eds) Proceedings of the Sixth International Conference on Artificial Life, MIT Press
Steels L, Kaplan F (2001) Aibo's first words: The social learning of language and meaning. Evolution of Communication 4(1):3–32
Steels L, Kaplan F (2002) Bootstrapping grounded word semantics. In: Briscoe T (ed) Linguistic Evolution through Language Acquisition: Formal and Computational Models, Cambridge University Press, Cambridge, pp 53–73
Steels L, Loetzsch M (2008) Perspective alignment in spatial language. In: Coventry K, Tenbrink T, Bateman J (eds) Spatial Language and Dialogue, OUP, Oxford
Steels L, Loetzsch M (2012) The grounded naming game. In: Steels L (ed) Experiments in Cultural Language Evolution, John Benjamins, Amsterdam
Steels L, Spranger M (2012) Emergent mirror systems for body language. In: Steels L (ed) Experiments in Language Evolution, John Benjamins, Amsterdam
Steels L, Vogt P (1997) Grounding adaptive language games in robotic agents. In: Husbands P, Harvey I (eds) Proceedings of the 4th European Conference on Artificial Life, The MIT Press, Brighton, U.K., pp 473–484
Steels L, Kaplan F, McIntyre A, Van Looveren J (2002) Crucial factors in the origins of word-meaning. In: Wray A (ed) The Transition to Language, OUP, Oxford, UK
Talmy L (2000) Toward a Cognitive Semantics, Concept Structuring Systems, vol 1. MIT Press, Cambridge, Mass
van Trijp R (2012) The emergence of case marking systems for marking event structure. In: Steels L (ed) Experiments in Cultural Language Evolution, John Benjamins, Amsterdam
Walhster W (ed) (2000) Verbmobil: Foundations of Speech-to-Speech Translation. Berlin, Springer-Verlag
Wellens P, Loetzsch M (2012) An adaptive flexible strategy for lexicon formation. In: Steels L (ed) Experiments in Cultural Language Evolution, John Benjamins, Amsterdam
Winograd T (1972) A procedural model of language understanding. Academic Press, New York
Wittgenstein L (1953) Philosophical Investigations. Macmillan, New York