Social Learning and Verbal Communication With Humanoid Robots

Luc Steels

Sony Computer Science Laboratory - Paris
VUB Artificial Intelligence Laboratory — Brussels
steels@arti.vub.ac.be

Abstract

The paper discusses issues in the design of open-ended and grounded communication with humanoid robots. A number of design principles and design maxims are proposed. The key idea of the paper is that social learning can play a crucial role in boot-strapping a humanoid robot into lnguistic culture.

1. Introduction

Recent advances in robotics, as demonstrated for example by the SONY AIBO [12] and the SONY SDR humanoid robot, are creating fascinating opportunities to push forward research on natural language communication with autonomous robots. Moreover there are clear advances in pattern recognition: better algorithms in many sensory domains, particularly in computer vision and speech

processing, novel statistical learning techniques like ICA, and more powerful computers or dedicated hardware so that solutions which were too slow in the past can now work in real time on real datastreams [2]. All these techniques help to bridge the gap between the subsymbolic world of sensing and effecting and the representations needed for language and higher order cognition.

On the other hand, language communication for humanoid robots is much less advanced. The goal of this paper is to analyse the requirements, understand key properties of natural language communication and why they make it difficult to build natural language systems for humanoids, and then propose a general framework for pursuing further progress. This discussion is based on prior technical work in implementing language games for autonomous robots [24]. On the one hand, we have done a large-scale experiment (the Talking Heads experiment) in which thousands of agents interacted through robotic bodies with the world and constructed a shared set of concepts and language to communicate about them [25]. We have also done experiments with the Sony AIBO robot for the acquisition of words through social learning [26]. Although fully open-ended grounded communication with humanoids through natural language will not be possible any time soon, the paper expresses optimism about the long term outlook.

2. Requirements

The use of natural language for open communication with humanoid robots is not a luxury but a necessity. Three requirements must be met.

First the communication must be grounded, meaning that the utterances must be about the shared situation in the real world in which robot and human find themselves. For example, it should be possible to say ’go over there’, while pointing to some corner of the room, and the robot should interpret the gesture and the request for action in the given situation and execute it. This makes the challenge very different from most current applications of natural language processing, such as front-ends of databases, where all semantics is purely symbol processing. Research on communication with virtual agents [1] is more relevant but in this application domain the world models are completely and accurately known to the agent. In the case of a robot which needs to gather information about the world through a sensory apparatus, world models can only be partially known and will always remain to a large extent uncertain.

A second requirement is that the communication is open-ended. A humanoid robot will have a repertoire of behaviors that expands based on novel combination or learning. The situations that will be encountered can not all be foreseen. Consequently, the communication system cannot be like a traditional computer interface where all possible commands are fixed in advance. Of course, we could build such a command interface to get communication started, but unprepared humans will quickly find this to be very unsatisfactory, particularly in the case of entertainment robots. Moreover detailed research on grounded natural language dialogs has shown that the conventions of a natural language are not fixed. In real world conversations, participants negotiate the way they express meanings as part of the conversational interaction [7]. It follows that learning and ’language invention’ must be essential parts of any language communication system for a humanoid robot.

Proceedings of the IEEE-RAS International Conference on Humanoid Robots

Third, the communication must be speech-based. Those interacting with a humanoid expect undoubtly that they can simply speak to it. There will be no facility for typing in text nor for clicking with a mouse. At the moment no hundred procent reliable speech recognition system exists and it will probably never exist given the large amounts of noise and variation in speech data and the open-ended character of natural languages. Nevertheless, some restricted speech systems now exist for applications like airline reservation and humanoid dialogs can build further on these achievements [27].

Note that each of these requirements generates uncertainty: uncertainty about the world state being communicated due to the need to ground language in sensori-motor behavior, uncertainty about the language conventions to be adopted, due to the open-ended nature of human language, and uncertainty about the speech input due to the nature of real speech. This may seem to make the task hopeless, but the advantage of a humanoid robot is that the total can be more than the sum of its parts. The situation and strong predictions from the world model plus gestures and other types of non-verbal communication can help to cope with uncertain speech input or gaps in knowledge about the language. Language communication can help to deal with world grounding because it provides additional sources of knowledge, complementary to constraints coming from the situation itself. Both the world and the conventions of the language can help to deal with the unreliability of the speech input by producing top down expectations.

In any case, despite these various sources of uncertainty, children in the time span of only a few years and with relatively poor cognitive capacities, at least compared to adults, manage to acquire enough natural language capacity to engage in open-ended dialogs. New words are learned very rapidly and new grammatical constructions are effortlessly absorbed. What is the magic bullet that can explain this remarkable achievement? If we could understand it, we might be able to bootstrap humanoid robots into human culture using a similar mechanism.

3. Social Learning

There have been extensive debates in the cognitive science literature to explain the rapid acquisition of language. One approach relies strongly on genetic predeterminism [20]. The main thesis is that language ability is so complex and it is learned so quickly with only weak stimuli and little direct feedback that there must be very rich innately given brain structures. It has therefore been suggested that most categories and conceptualisations underlying language are already known to the learner [13].

Language acquisition has been compared to setting parameters in a universal grammatical framework [6]. From the viewpoint of humanoid robotics there is nothing against this hypothesis. On the contrary, if we could find out what needs to be programmed in to make rapid acquisition possible, noone would object. The problem is to find out what exactly the universally innate structures should be. Languages not only exhibit universal tendencies but also a lot of idiosynchratic features and observations of real dialogs show constant shifts and renegotiations in the lexicon and even the grammar. The differences in the conceptualisations underlying different languages occasionally run very deep, even for domains like color terms [10] or spatial descriptions [3], which are usually believed to be universal. Perhaps this is the reason why nobody has been able to work out a sufficiently detailed proposal for universal conceptualisations and grammar that could be the basis of computer implementation.

An alternative approach is to de-emphasise innate structure in favor of individualistic learning. It is typified by a lot of neural network approaches to language acquisition. The learner is assumed to receive a stream of inputs (speech sounds, situations where sound and meaning co-occur, examples of temporal constructions, etc.) and to develop an internal coding that eventually enables communication [11]. These proposals have the advantage of being very concrete and testable, but have not yet lead to the rapid learning observed in human children. Learning usually takes an enormous amount of time and the data needs to be prepared carefully. Neural networks do not appear to be the 'magic bullet' that can explain the enormous speed of language learning. I believe that this magic bullet is social learning. This hypothesis has also been advanced by a number of cultural psychologists, in particular [27].

Here are some concrete examples which illustrate the concept of social learning. They all involve at least two players in a shared situation, one acting principally as a mediator and the other one as a learner.

Example 1: A giraffe, a dog, and an animal never seen before are put on a table by the individual playing the role of the mediator. The animal has white fur, one leg, three eyes, and some other bizarre features.

M(ediator) Can you give me the wrob? [said while pointing and looking at the novel animal]
L(earner): This thing? [said while handing over the animal to the mediator]
M: Yes, thank you.

Clearly, the mediator has set up the context and indicated the goal through non-verbal cues. The learner has succeeded to guess the intention of the mediator, even though he never heard the word "wrob" before. Note that there is no explicit linguistic feedback here. Much has been made of the apparent absence of feedback in language but it is usually ignored that the context gives the pragmatic feedback from which linguistic feedback can be deduced. Thus the learner can guess here that "wrob" must be the name of one of the objects in this particular situation. If he knows already words for giraffe and dog, that would in itself be evidence that wrob must refer to the third type of animal, but even without this knowledge the pointing and gaze direction of the speaker provide sufficient constraints. The learner is not only learning that "wrob" is the name of an object, or perhaps of a class of objects, but is also learning about the class itself. Wrob is a first example that can be stored in memory and it could be used later by analogy. When more examples of wrobs are seen, more views can be stored and the wrob-class becomes progressively richer and more refined.

Example 2: There are two cups on the table close to each other, a blueish-green one with white dots and an orange one with black stripes.

T(eacher): Can you give me the zoopy cup?
L(earner) [gives blueish green one]
T: No, the zoopy one.
L: [hands over orange one]
T: Yeah, thanks.

The mediator again sets up the context although non-verbal cues are weaker than in example 1. The learner figures out that "zoopy" must refer to a characteristic of the cup because the mediator assumes that the meaning of "zoopy" is enough to identify which cup is meant. Zoopy is without doubt an adjective. The learner can deduce that "zoopy" has to do with the visual appearance of the cup. Of course it could also refer to its position, the orange cup being closer to the learner or more to its left, but this is less likely given the greater salience here of color differences. As in the previous example the learner not only acquires the meaning of a new word but also a new category ’orange with black stripes’.

From these examples we can gather a number of features of social learning, which have more to do with the setting than with a specific learning mechanism.

1. The first feature is that the learning event involves at least two individuals in a shared environment. They could be called the learner and the mediator. The mediator could be a parent and the learner a child, but children (or adults) can and do teach each other just as well. In fact sometimes learning takes place because two individuals engage in a dialog without anyone of them already knowing in advance the solution.

2. The goal of the interaction is not really teaching, and therefore calling the mediator a teacher is misleading. The goal is rather something practical in the world, for example to identify an object or an action, like when someone says: "Give me that pen", and the other person identifies the pen, picks it up, and hands it over. Learning takes place when the listener did not know the word "pen" but grasped it from the present context and now remembers it for future use.

3. The mediator has various roles: She sets constraints on the situation to make it more manageable (a process often called scaffolding), gives encouragement on the way, provides direct feedback, and acts upon the consequences of the learner’s actions. For example, if the learner picks up a piece of paper instead of the pen, the mediator might say: "No, not the paper, the pen", and point to the pen. This information is not only crucial to succeed in the task but supplies the learner with all the information necessary to acquire new knowledge.

4. The learner actively tries to guess the intentions of the mediator. This is probably the most important feature of social learning. The intentions are of two sorts. First of all, the learner must guess what the goal is that the mediator wants to see realised (like ’pick up the pen on the table’). But, the learner must also guess the way that the mediator construes the world. This is particularly true for language. Language communication, as mentioned earlier, always involves two things: a way of categorising and conceptualising the world and words and constructions that express these conceptualisations. For example, if someone says "give me that pen", she has categorised the object to be given in a particular way, and conceptualised the desired behavior as an action with an agent, a recipient, and an object to be given. Many things are left out such as how it should be given or what trajectory should be followed. Other languages might conceptualise the same desired behavior in a very different way.

5. The social interaction between learner and mediator may completely succeed, in which case both of them obtain evidence that the learner was able to guess the intentions of the mediator based on the concrete situation and available knowledge sources (including knowledge of the language). The interaction might also fail, in which case additional exchanges between learner and mediator can try to put it on the right track again. Most importantly, failure and their repairs are an opportunity for the

learner to acquire new meanings of words and grammatical constructions and to learn new ways in which the world can be construed.

6. The knowledge individuals need to engage in these social interactions is never definite and static but always changing and being refined as new shades of meaning become known, new views of objects are learned, etc. Every interaction is therefore a new opportunity to learn or test out existing knowledge. There is no substitute for life-long learning.

Social learning is clearly different from individualistic learning, in which the learner is assumed to be confronted with a stream of data that provides examples and possibly counterexamples. Such a type of learning is well modeled with existing neural networks or symbolic learning algorithms which perform induction to arrive at the most compact representations of the examples seen. I am not claiming that this kind of learning does not occur. It is probably of great relevance for different types of sensori-motor intelligence, for example for learning the most efficient color coding of natural scenes or learning associations between sounds and images. However I do believe that for the acquisition of language and other forms of high level cognition, isolated individualistic learning plays a minor role.

Social learning (at least as narrowly construed here) goes beyond imitation, although imitation clearly establishes the first foundations of social learning from the very first moments of neonate life [18]. The acquisition of sounds and body movements are typical examples where imitation plays a major role. They follow on the footsteps of 'babbling' which consists of the individualistic exploration of the sound space or the bodily movement space. In the case of imitation, the main purpose is to reenact movements whose effect was heard or seen as produced by another person, and so it is a matter of learning to categorise the movements of the other in the auditory or visual domain and retrigger them based on your own movements. The crucial additional feature in social learning, as defined here, is the need to guess the mental state of the other person with respect to an external situation, specifically the goals and conceptualisations of the other person in that situation.
The guess is based on what you would do yourself in the same situation. It requires reflections of the sort: What makes sense to me in these circumstances, and so probably makes sense to the other? What is a potential goal here? What possible meanings fit with this situation?

Based on this analysis, we have made the social learning of language the target of our own research in dialogs with humanoid robots, particularly because we have been able to demonstrate this kind of learning already on less complex robots [25,24]. Social learning for a humanoid robot does not stand on its own but must be integrated in the fully interactive grounded behavior of the robot. More specifically we have to face three major difficulties before social learning itself can be embedded properly.

The first difficulty is that a vast array of behavioral and cognitive competences are required, ranging from picking out people in the image and detecting the direction in which they are looking [22], interpreting emotional state based on prosody [19] or facial expressions, detecting pointing gestures and identifying to which object they are directed, computational auditory scene analysis [5] setting up an appropriate turn-taking dynamics [4], etc. We need relatively robust solutions for these various components and we need above all powerful ways to integrate them into a coherent total behavior. Fortunately, recent work in AI and pattern recognition has made significant advances for all these components.

The second difficulty is that we must understand better the nature of human social relations, the relation to novel artefacts, and how these relations get established and confirmed through language. This research can build further on recent efforts towards the study of sociality in robots [9, 8].

The third difficulty is to set up adequate social interactions and more specifically dialogs. Dialogs are flexible and multi-faceted. They do not follow a strict pattern but the coherence arises in an emergent fashion through turn-taking. We need design principles for setting up such social interactions and a methodology to turn them into concrete dialogs, building further on earlier work in dialog management [14,28] but adding considerable flexibility. The rest of the paper elaborates some design ideas for building open-ended grounded dialogs in which social learning can be embedded.

4. Designing dialogs for social learning

4.1. Games

A deeper investigation of social interactions, particularly by very young children (2 years), shows that they can be structured in terms of games [17]. A game is a series of interactions between at least two participants and involving objects that play some role in the game. Participants say certain things at certain times, do certain things in a predictable sequence, and repeat the same sort of actions to the same sort of objects. It is only after the interaction has become routine that variations can be brought in so that the learner can acquire new knowledge.

Here is a simple example, a game undoubtly familiar to many parents:
M(ediator): What does the cow say? [shows cow] Mooooh.
L(earner): [just observes]
M: What does the dog say? [shows dog] Woof.
L: [observes] ...
M: What does the cow say? [shows cow again and then waits]
L: Mooh
M: Yeah!

The learner in this case is learning to imitate the sound of various animals, the names of the animals, and their possible appearance. The game is typical, in the sense that (1) it involves many sensory modalities and abilities (sound, image, language), (2) it contains a routinized interaction which is well entrenched after a while so that it is clear what is expected, (3) the learner plays along and guesses what the mediator wants and the mediator sets up context, constrains the difficulties, and gives feedback on success or failure.

Conversations between adults become less structured but can nevertheless still be studied in terms of games as research on conversational game theory has shown [21]. The main primitive unit proposed in this type of analysis are moves, such as instruct, explain, initiate-game, close-game, confirm, clarify, etc. [16]. They are combined together in transition networks that implement turn-taking behavior. Conversational game theory is mostly used for analysis of human dialogs but can be used in dialog design.

An important type of game is known as symbolic play, which starts to occur in children around the same time as language takes off. In symbolic play objects and people are given imaginary roles and a narrative is constructed and developed. A pen may become an airplane, a person may become a doctor, a long imaginary trip is played out, etc. Symbolic play appears to be an exercise ground for working out human relationships, for taking on different perspectives, and for mastering complex realities. The power and qualities of symbolic play are an important indicator of the cognitive and social development of children and therapeutic methods have been developed to help children with developmental disorders [23]. (Small) humanoid robots could potentially play a significant role in fostering symbolic play.

Social games, particularly as designed for young children are intended to be fun, which can be achieved by introducing surprise, or just the opposite, sticking to repetition so that the learner feels secure enough to go just a little step further. Fun can be enhanced by introducing humor, for example by giving the humanoid a recognisable character that deviates in unexpected ways from the norm. The fun aspect induces positive motivational states which are known to be facilitators of learning.

4.2. Mirroring

A crucial feature of social learning is the assumption that the other is like yourself. This helps to predict what the other person is going to do or what she may know and suggests what you should do yourself. One consequence of this principle is that every game and every step in the game should have its mirror. The agents (human and robot) must have scripts for both situations: one where they take the initiative and one where they follow the game line set out by the other. Here are some examples of the application of this design principle.

Agents involved in social interaction need to know as much as possible about each other's state. This has given rise to extensive rituals in many societies, including standard phrases for greetings, the exchange of business cards, etc. One aspect of state is the physical location of the other partner. Locating a human requires processes detecting and tracking faces, processes identifying the location of sound sources, the matching of faces with sound sources, and the maintenance of a model (both in absolute and relative coordinates) of the position of the partner. For each of these sources of location information, mirror behaviors can be performed by a humanoid to make it easier for a human to detect its physical location: moving around or making gestures to become noticed, look at the human so that the face is easier to recognise, make a sound, etc. Another important item is the social relation and emotional state of the agent. This can be gleaned from prosody, facial expressions, and bodily movements. The humanoid can execute mirror behaviors that communicate its own emotional state.

One of the remarkable features of human learning is the ability to quickly take turns from observer and/or participant to player, which means that roles get reversed and the player now carries out the actions seen first through observation.

4.3. Interaction dynamics

The next general principle concerns the interaction dynamics. This needs to be such that a smooth turn-taking takes place. Turn-taking cannot be clocked but should emerge from the myriad of interactions embedded in the world, as already illustrated in Kismet [4]. There are a number of ways to establish and maintain synchronised interaction and each has mirror behaviors:

+ Ongoing interest must be shown. This can be done by nodding, interjections, etc. The robot must

monitor whether interest is still strong by monitoring these cues but mirror them by producing similar cues.
+ Dialogs have rhythms which need to be respected. So the robot should try to detect the rhythm by measuring times between turn-taking, speed with which the other interrupts, or not, etc. Once the rhythm is known, it can help to time one's own interventions.

+ Emotional states must be matched. Given an emotional state of the human, the robot must exhibit an appropriate 'reply'. For example, given anger, the robot could either also exhibit anger, or just go in the opposite direction to calm the situation.

+ Spatial dynamics. Moving away during a dialog is seen as the end of an interaction, but coming too close might be seen as an invasion of private space. So the robot must monitor spatial distances and help regulate.

+ Gesture repetition. It is known from the anthropological literature that people who are intensely engaged in social interaction (for example in courtship) will also imitate each other's gestures and body postures. A humanoid must attempt to recognise certain gestures to induce a similar feeling of close coordination.

+ Attention focusing. The agent should establish in all possible ways whether the partner is paying attention and try to detect on which object the partner is focused. Attention is potentially shifting at any moment from one object to another or from one person to another. So the agent must be constantly on the look out and must itself indicate as clearly as possible to what it is paying attention. There is no simple unique way to do this. All possible sources of knowledge and behaviors must be exploited.

5. Embedded learning

5.1. Learn whenever you can

Most social interactions contain novel elements, otherwise they would not take place. The agent must therefore be constantly ready to extract from the interaction information that could be relevant for the future, or induce missing pieces.

5.2. Immediately test new knowledge

Learning opportunities arise when there is a breakdown in the interaction but feedback from the play partner, the situation, and the routinised interaction of the game are sufficient to fill in gaps and hence conduct a repair. Repair strategies are particularly important for the acquisition of language, as discussed briefly later. This leads to the next design maxim: All new knowledge being acquired should be tested right away when there is still sufficient context to the interaction. The nature of the test depends of course on the kind of knowledge that was acquired. For example, the robot should repeat a new word to make sure that the pronunciation was acquired correctly and then wait until there is confirmation. If a new categorisation of the world has been learned (e.g. as the meaning of 'zoopy') it should be tested when a new situation arises (such as two orange cups but one with stripes and the other one without them). A confirmation may not necessarily come, which by default can be interpreted as positive evidence. If the partner thinks the wrong knowledge was acquired she is likely to react. Note that the mirror principle applies here as much as anywhere else. The robot should be on the look out for ways in which the partner tests new knowledge and be prepared to provide confirmations or corrections.

5.3. An example

Even though initially a game may follow a rigid pattern (and very young children often insist that things are exactly the same) quickly many variations develop to play the same game. So it is better to think of a game in terms of a number of desirable subgoals which are posted when a particular game is entered and which can be achieved in more than one way, rather than in terms of a fixed set of interactions in a specified order. Research on conversation analysis may provide inspiration to identify what useful building blocks are.

Here are some steps in a possible greeting game between a person (Luc) and a robot (Robbie), which illustrate the maxims introduced in this section. The greeting game fits within the general goal of gathering as much information as possible about the partner, specifically the identify of the partner.

Luc: Hi, [get attention. proceed if robot pays attention]
Luc: what is your name? [indicate nature of the game]
Robbie: Robbie. [robot detects game and plays along]
Luc: Robbie, [person tests word]
Luc: that is a nice name. establish positive social relation]
Luc: I am Luc [mirroring. robot acquires new name. New name is associated with certain sound characteristics or facial features uniquely identifying the person.]
Robbie: Luc? [test new knowledge]
Luc: Yes [confirmation]

Here is another way to play the same game. The surface form looks very different but similar subgoals are involved:

Luc: Hello , [get attention. proceed if robot pays attention]
Luc: are you Francesca? [test knowledge]
Robbie: No, I am Robbie [correct knowledge]
Luc: Oh, Hi Robbie. [confirm knowledge]
Robbie: I am Luc. [mirroring]

These examples illustrate that for every conversational goal, the agent must be able to produce many variants and recognise many variants. To increase robustness and make the dialog richer and less predictable for humans.

6. Conclusions

The paper analysed requirements and processes needed for open-ended grounded dialogs with autonomous humanoid robots and proposed a number of ideas, principles, and maxims for designing such dialogs. The main hypothesis is that social learning is the key towards the bootstrapping of language abilities and the acquisition of their underlying conceptualisations. Social learning means that both partners in the interaction attempt to guess the state of the other one based on a model of themselves. Social learning must be embedded in flexible dialogs. We have found that it is very effective to structure them along the lines of games.

7. Acknowledgement

I am indebted to many people from the Sony Computer Science Laboratory in Paris, the Digital Creatures Lab in Tokyo, and the VUB Artificial Intelligence Laboratory in Brussels. In particular, Frederic Kaplan of Sony CSL who has done various experiments with language games on the Sony AIBO, Angus McIntyre, who has been implementing a computational foundation for flexible dialog management, and Toshi Doi and Masahiro Fujita and their team, creators of the AIBO and the SDR humanoid.

8. References

[1] Badler, N., M. Palmer, R. Bindiganavale (1999) Animation control for real-time virtual humans. Communications of the ACM 42(8), August 1999, pp. 64-73.
[2] Beymer, D. and K. Konolige (1999) Real-time Tracking of Multiple People Using Continuous Detection. 7th IEEE Conference on Computer Vision, Greece.
[3] Bowerman, M. and S. Levinson (2001) Language acquisition and conceptual development. Cambridge University Press, Cambridge.
[4] Breazeal, C. (1998) A Motivational System for Regulating Human-Robot Interaction", Proceedings of AAAI98, Madison, WI.
[5] Bregman, Albert S., Auditory Scene Analysis: The Perceptual Organization of sound. Cambridge, Massachusetts: The MIT Press, 1990
[6] Chomsky, N. (1980) Rules and representations. Blackwell Publishers, Oxford.
[7] Clark, H. H., and Brennan, S. A. (1991). Grounding in communication. In: Resnick, L.B., J.M. Levine, and S.D. Teasley (Eds.). Perspectives on socially shared cognition. Washington: APA Books.
[8] Conte, R. (2000) Intelligent Social Learning. In: Proceedings of the AISB'00 Symposium on Starting from Society - the Application of Social Analogies to Computational Systems, Birmingham, UK:AISB, 1-13.
[9] Dautenhahn, K. (2000) Human Cognition and Social Agent Technology John Benjamins Publishing Company, Amsterdam.
[10] Davidoff, J., I. Davies, J. Roberson (1999) Color categories in a stone-age tribe. Nature, vol 398. 230-231.
[11] Dominey, P. and F. Ramus (2000) Neural network processing of natural language I : Sensitivity to serial, temporal and abstract structure of language in the infant. Language and Cognitive Processes.
[12] Fujita, M and H. Kitano (1998). Development of an autonomous quadruped robot for robot entertainment. Autonomous Robotics, 5:1--14.
[13] Fodor, J. (1983) The modularity of mind. The MIT Press, Cambridge Ma.
[14] Grosz, B. and C. Sidner (1987) Attention, Intentions and the Structure of Discourse. Computational Linguistics 12(3):175-204.
[15] Konolige, K. and K. L. Myers, E.H. Ruspini and A. Saffiotti (1999) The Saphira Architecture: A Design for Autonomy. Journal of Experimental and Theoretical Artificial Intelligence, 9(1): 215-235, 1997.
[16] Lewin, I. (2000) A formal model of Conversational Game Theory. In: Gotalog: 4th Workshop on Semantics and Pragmatics of Dialogue, June 2000
[17] Nadel, J. (1986) Imitation et communication entre jeunes enfants. Presses Universitaire France, Paris.
[18] Nadel, J. and G. Butterworth (1999) Imitation in infancy. Cambridge University Press, Cambridge.
[19] Oudeyer, P-Y (1998) this volume.
[20] Pinker, S. (1994) The Language Instinct. Morrow, New York.
[21] Powel, R. (1979) The organization of purposeful dialogues. Linguistics 17: 107-152.

[22] Scassellati, B. (1998) Eye Finding via Face Detection for a Foveated, Active Vision system. In Proceedings of AAAI-98. AAAI Press Books.
[23] Slade, A. and D. P. Wolf (1994) Children at Play. Clinical and Developmental Approaches to Meaning and Representation.
Oxford University Press, Oxford.
[24] Steels, L. (2001) Language Games for Autonomous Robots. IEEE Intelligent Systems. in press.
[25] Steels, L., F. Kaplan, A McIntyre and J. Van Looveren (2001) Crucial factors in the origins of word-meaning. In Wray, A., editor, The Transition to Language, Oxford University Press. Oxford, UK, 2002.
[26] Steels, L. and F. Kaplan (2001) AIBO's first words. The social learning of language
and meaning. Evolution of Communication 4(1).
[27] Zue, V. (1997) Conversational Interfaces: Advances and Challenges. Proceedings Eurospeech Conference, Rhodes pp. 9-18.