An Interactivist Perspective of Autonomous Robotic Sign Users

Erich Prem

Austrian Research Institute for Artificial Intelligence
Schottengasse 3
erich.prem@oefai.at

Abstract. In this paper we present an interactivist perspective on design principles for autonomous sign users such as autonomous robots. Past discussions have centered around symbol grounding or anchoring where concepts are considered prototypes in the perceptual space of a robot or animal. Here we advocate a more complex view starting from the purpose of using signs or pursuing *sign acts*. We argue that concepts actually serve a number of purposes and revisit techniques for representing and memorizing concepts. Our own proposal focuses on the intentional and anticipatory aspects of sign acts and on anticipatory representations. This research is aimed at designing architectures for growing up autonomous sign users.

Introduction

Research in embodied Artificial Intelligence and robotics has recently addressed the problem of how to generate language-like behaviour in autonomous agents. Researchers in these fields deal with the question as to how it is possible for a robotic agent to acquire the meaning of words through interaction with the environment and with other speaking agents (artefacts or humans). At least three main directions can be identified in this kind of research: (i) *symbol anchoring* refers to the more engineering aspects of how to establish and maintain over time the connection of a symbol to an entity in the robot's world, cf. (Coradeschi & Saffiotti 01). (ii) *symbol grounding* deals with the more philosophical question as to how it is possible for an agent to acquire symbols that possess intrinsic meaning, cf. (Harnad 90, 93).

Both approaches share an interest in how to design agents that acquire through experience the capability to correctly use signs. But while it is sufficient for *anchoring* to develop such a technique in a way that works from an engineering point of view, *grounding* is more interested in what the conditions for sign users are such that the meaning of the sign is original in the system and not just "parasitic" on a clever programmer's mind.

Another approach to the study of robotic sign users is pursued by a group of researchers that are interested in developmental or evolutionary studies of linguistic behaviours, e.g. (Steels & Vogt 97, Steels 01). In approaches like these, agents learn to exchange signals that refer to objects or states of affairs in the agent's environment through trial and error and based on communication with other agents in the same

environment. The resulting systems serve as models or metaphors of the development or evolution of communication.

Sign acts of autonomous sign users

The focus of this paper is on agents that make use of signs and on the underlying structures that enable these agents to use signs adequately, i.e. for the purpose of reliable indication. Signs in this paper are entities with referential properties ranging from signals, object tokens, stigmata to words. Sign users are agents who either use signs passively or who create signs for a range of purposes. Here we do not limit ourselves to spoken human language. On the contrary, our aim is to discuss the question as to what the fundamentals are of using signs that we find in all autonomous sign users. Also, sign usage in this paper is not strictly limited to conveying information about the agent's environment. Most importantly, this paper advocates a view in which signs are not necessarily used for descriptive purposes alone. We are also interested in using signs for greeting, receiving attention, navigation, etc. Thus, we do not use the term "symbol grounding" in what follows.

Examples for autonomous sign users performing sign acts include humans who wave at people to get their attention, people who use the turning signal of their car to indicate the direction in which they are going or persons who navigate through a city using signs and arrows. Sign users in this paper could also be animals as in the following two examples:

Bees are known to encode distance and angle to a food source in their dance. The angle is encoded with respect to the sun's azimuth in the waggle portion of the dance referring to gravity as zero. The dance itself is usually performed on a vertical surface within the hive and taken up by other bees as an indicator for food sources.

Vervet monkeys are known to produce a diverse set of warning calls as soon as they detect a predator. The referential repertoire of vervet monkeys covers a relatively wide range from snakes, raptors, cats to primates. Depending on the type of alarm call, fellow vervet monkeys react with an appropriate behaviour, i.e. either escape towards the ground or into the trees.

In the following, we take a closer look at how the usage of (linguistic) signs was studied to date in most of the autonomous agents literature.

Concept Spaces

Nearly all accounts of techniques to generate autonomous sign users start with separating signs and sign acts from the underlying representation that is said to possess referential properties. These underlying representations are often related to the notion of concepts. The basic idea is that signs label concepts which become represented in the mind of the sign using agent based on experience. The philosophy behind this idea is the classical semiotic triangle in which a sign (symbol) designates a referent (or object) through its relation to the concept (or idea, meaning, etc.). The

technical problem in symbol anchoring and grounding then is to create a corresponding representation that guarantees the validity of the desired connection of sign and referent.

In such a view, sign acts really serve to create the idea of the referent in the listener's mind. In using a sign that refers to an object in agent A's environment, it activates a representation in agents B's mind that will ensure that the same object becomes connected to the sign used. Before we propose a different approach to the semantics of sign acts, we review previous accounts of concept spaces for signs users.

Representations for concepts

Following Harnad (87) categorical perception lies at the basis of such an approach to language and symbol usage. Categorical perception in turn is based on discrimination and identification. Discrimination requires a subject to tell apart stimuli presented in pairs, indicating whether they are the same or different. Identification requires the subject to categorize stimuli using labels. Categorical perception occurs when there is a quantitative discontinuity in discrimination at the category boundary.

Harnad (90) suggests that three kinds of representations are necessary in a system capable of grounding symbols:
- Iconic representations: sensor projections of the perceived entities
- Categorical representations described before, i.e. learned and innate feature detectors that pick out the invariant features of object and event categories from their sensory projections
- Symbolic representations: symbolic strings describing category membership.

Harnad argues much in favour of a connectionist network as the suitable tool for learning invariant features (Harnad 93).

In Davidsson's ((94) and (95) for autonomous agents) general concept representation framework a concept is represented by a composite description consisting of several components. Different kinds of representations should be available for different purposes, e.g. for perceptual categorization or for high-level reasoning. In this view, the description consists of the following entities:
- Designator: the name or symbol used to refer to the category
- Epistemological representation: used to recognize the instances of the category
- Inferential representation: a collection of "encyclopaedic" knowledge about the category and its members used to infer non-perceptual information and to make predictions.

An example would be the symbol "chair" with the epistemological representation of sensory properties etc. and the encyclopaedic representation describing the chair as something to sit on etc. Although Davidsson does not necessarily imply a certain type of representation, the proximity to Minsky's (75) concept of frames is obvious.

Results in Cognitive Science point to the fact that there are no general necessary and sufficient conditions for category membership. Instead, prototype-based concept descriptions appear more appropriate. These representations can be based on one or more of the following techniques:

- Memorization of specific instances (exemplar approach)
- Construction of a probabilistic representation (probabilistic approach) in which not the prototype but the similarity to the prototype is stored.
- Discriminatory approach in which not the categories themselves, but boundaries between categories are learned. Here the underlying assumption often is that all categories are present in the exemplar set. Multi-layer perceptrons and decision trees often implement this approach. As a consequence, these techniques do not easily allow for the generation of a specific prototype exemplar.
- Generative or characteristic models: here the aim is to discriminate the instances of a category from all other possible instances. These models concentrate on the similarities between the members of a category, boundaries are an implicit by-product of this approach.

It is worthwhile to point out that all the accounts above orient themselves at a view of representations that serves to carve out descriptive properties of the objects such that a reliable connection between sign and referent or object and concept can be established. Harnad' view has drawn heavy criticism from many different sides in the past. The ongoing focus on his proposed approach of symbol grounding is all the more surprising. In the following, we would like to propose a shift towards more intentional and anticipatory aspects of concept space representations and sign usage when it comes to questions of how to design autonomous agents or robots.

Representational Interactivism

In an effort to critically examine Harnad's view of symbol grounding and, more generally, the notion of representation in Artificial Intelligence, Bickhard & Terveen (95) suggested that in order for representations to be about something it is necessary that representations can be wrong and thus that they are tightly connected to interaction outcomes. Actually, what really becomes represented in a representation is the anticipated interaction outcome internal to an agent. It was argued in Prem (98) that if concepts are acquired through an adaptive agent's interaction with the world it can be shown that the resulting representations necessarily contain a model of the world and the agent's interaction with it. More precisely, the resulting representations necessarily are anticipations of interaction outcomes.
It is important to note that such a view of the nature of representation means a divergence of merely descriptive approaches to the problem of concept formation. In this view, the features of any "object" in the agent's environment are entirely based on its functional properties for the agent.

In previous work, most notably in (Prem 00), we suggested to use this view also as the basis of approaching the nature of sign usage in autonomous agents. The essence of being semantically autonomous is that signs used by an agent should possess meaning with respect to the agent's own goals or activities. Of course, meaning here is related to interaction and interaction outcome, *not* to sensory properties of the sign's referent. More precisely, symbols or signs should be meaningful due to the properties they possess for the agent in interactions with the world. In many cases,

sign-based interactions may be the use of symbols for communicative purposes, but the aforementioned purpose of orientation or receiving attention is another plausible semantically autonomous sign act in this view.

Again, sign acts are not usually treated in this way in autonomous robot research. "Symbol anchoring" does not focus on the inherent relativitiy of the concepts developed by the agent. It actually looks more like a way to automate the process of developing sensor-based feature detectors for classes of objects. Symbol anchoring does not address the question as to why the represented object is important to the robotic agent. Symbol anchoring also does not deal with how signs are used by the robotic agent and in this way distinguishes itself from symbol grounding. But even symbol grounding research is usually not focused on interaction outcomes of symbol usage. However, in what follows we propose the fundamentals of an approach to grow concept spaces for autonomous sign users that are entirely built on the instrumental character of sign acts.

Sign acts and autonomous sign users

Action circuits

We propose to approach the problem of creating concept spaces for autonomous sign users based on the ethological construct of an action circuit. Following von Uexküll (28), the action circuit is a model of the tight coupling that all living systems exhibit in their interactions with the word (see Fig.1).
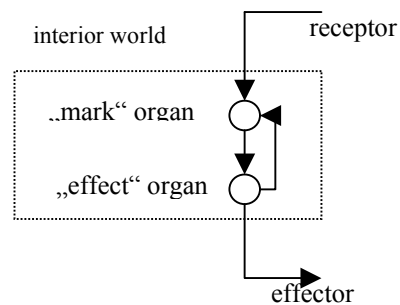

Fig. 1. The action circuit, first described by von Uexküll. The properties of the mark organ and of the system-environment interaction determine the interior world of the animal. The dynamics itself is driven by the interaction purpose, i.e. the internal outcome of the interaction.

Von Uexküll used the notion of the "mark" organ (sensory system) and the "effect" organ with which an animal creates its way of interacting with the world. Through these purpose-driven interactions, the animal creates an internal world view that

becomes the world according to the animal and its discovery and precise description usually is a difficult task for the biological scientist.

The idea was taken up in robotics mostly in Brooks' (91) famous behaviour-based architecture that consists of layered modules driving a robot. These modules receive inputs that are highly tuned to the specific behaviours and their output can either directly influence the robot's behaviour or indirectly suppress other behaviours. As a consequence, behaviour-based robots interact with the world at high and previously unseen interaction dynamics.

From what we have said before, it should be straightforward to realise that reference in autonomous sign users is based on the tool-character of signs. However, this reference is not something that happens merely because signs are used to refer. In contrast, signs are used for circumstances and purposes in which (we use a terminology close to Heidegger (27) and Dreyfus (90) here) an agent already dwells. In these circumstances, the agent can already cope with the situations it encounters. Signs are merely another tool to achieve the desired interaction outcome. Signs are, however, peculiar in pointing out parts of the contextual whole of references in which an agent might already find itself, i.e. signs can point out the "wherein" of living. The reader may wish to compare to Bickhard's discussion of context in what he calls the level of apperception in (Bickhard 98).

We thus propose to regard active sign acts, i.e. acts of creating or putting signs as anticipations of successful interactions of indication. Passive sign usage, i.e. following or taking up signs encountered in the environment on the other hand should be regarded as giving orientation to action circuits.

To make this proposal more easily understandable, consider the following examples: Car drivers use turning signals to indicate the direction in which they plan to go. Sign usage in this example is clearly based on interaction outcome. Using the turning signal is learned as a consequence of our experience that it makes driving easier, less problematic, and less dangerous. An example for an autonomous robot would be to cry for "Help!" which serves to get out of places where the robot is stuck. The meaning of "Help!" is the anticipation of getting out of the current problematic situation.

Using the turning signal then is the anticipation of the successful outcome of indicating one's direction to others, i.e. the appropriate behaviour of other participants in traffic. These participants in turn also use the signal in orienting themselves towards the sign. Here, we realize that the action circuit of "driving home" etc. receives its orientation through the sign that is actively used by the driver in front. The following table exhibits a number of potential sign acts that we can find in living beings or could make sense for autonomous robots.

Sign Act	Behaviour
Greeting	The agent reacts to a greeting or salutation or to another agent with a specific greeting behaviour.
Set mark	The agent marks an interesting location or object in the

	environment so as to retrieve it later more easily. This would be a form of auto-stigmeric communication.
Warn	Produce an alarm signal to make group members aware of danger or make them run away.
Flee	React to a warning signal by running away.
Follow arrow	The agent moves in the direction to which an arrow points. This sign orients an action circuit such as *move* or *find target*.
Find place	The agent navigates to a designated place. Examples include "here", "there", "home", etc.

Table 1. Examples of sign using behaviors (sign acts).

Growing representations to anticipate interaction outcome

If an autonomous robotic system succeeds in using signs actively in the way proposed here, the resulting system would need a representational subsystem capable of representing the interaction outcome of indication actions. In effect this means that whenever the robot encounters a situation in which sign usage could be helpful, this representation should become active and thus provide an anticipation of the outcome of actively indicating something to others.

It was argued previously (Prem 00) that adaptive autonomous systems generate representations which lead to a view of the world as the possibility to act, i.e. the world according to such an agent appears as full of things to interact with in order to pursue one's goals. Accordingly, when we add sign acts to this picture, a social world with someone to signal to will appear equipped with potentials to signify.

For passive sign users the situation is somewhat simpler in that only a system of behaviours is needed that allows for action circuits (or behavioural elements) to become orientated at signs encountered in the environment.

From anticipations of indication outcome to descriptive signs

It is clear that the above proposal is not very precise because it allows for an extremely wide range of potential sign acts. In particular, one interesting question in this context is how do we get from such a view to merely descriptive language, to grounded or anchored symbols that seem to be so useful in everyday life or even for robotics? The answer is that once we develop a mechanism that enables the generation of anticipated indication outcomes, we should also be able to develop a system for using descriptive signs.

Let us for the moment assume that the first "descriptive" nouns (something with a meaning like "mummy" or "daddy") in baby talk are produced so as to generate parental reward, *not* to denote the corresponding objects. It would then be necessary to devise a method for re-using a representation anticipating this kind of reward for

other rewards or purposes. As an example, it would be interesting as a next step to also receive reward from others for correctly pointing to mum and dad or for crying for help etc. The essence of all these potential anticipations of internal indication outcomes, this is the idea here, should then converge towards a mainly referential representation that happens to denote mum and dad in a number of contexts.

Staying with baby talk for a moment, note that there is a close relation of "Give me X!" and "This is X!", because the former sentence happens to produce the desired outcome, if "X" refers to X. Note, however, that "X" can still mean many different things in other contexts. Similarly with vervet monkeys, "leopard!" will model the leopard, but always in order to make one's friends jump up the trees. As a consequence, it is not necessary to enrich "connotations" of the word "leopard" after its descriptive properties are learned. They will, quite to the contrary of such a view, ensure in the first place that the sign is properly used.

Conclusion

In this paper we have presented fundamental elements for growing concept spaces for autonomous sign users. Out of dissatisfaction with current views of language in autonomous robot research, it is proposed here to first study general properties of sign usage and to avoid a premature commitment to linguistic signs only.

Secondly, we propose to change from the use of merely descriptive (feature-based) concept spaces to interactivist representations as already suggested by Bickhard and Terveen. As a consequence, we regard representations as anticipations of interaction outcome. It is then only a logical next step to also view many sign acts from the point of view of indication outcomes, since signs are tools used for indication purposes. (There may be other purposes and aspects of signs that we do not touch upon here.)

We have argued that representational spaces for signs will then allow for the easy anticipation of these indication outcomes such that the environment of an agent will appear as an opportunity to indicate in specific situations.

Finally it is argued here that the re-use of the developed anticipatory representations for different indicatory purposes in language behaviours can lead to signs that exhibit referential properties in the more traditional sense.

Acknowledgements

References

M. H. Bickhard, L. Terveen, Foundational issues in AI and Cognitive Science, Elsevier Science Publishers, 1995.

M. H. Bickhard, Levels of Representationality. Journal of Experimental and Theoretical Artificial Intelligence, 10(2), 179-215, 1998

R.A. Brooks, Intelligence without representation. Foundations of Artificial Intelligence, Artificial Intelligence, 47 (1-3), 1991.

S. Coradeschi, A. Saffioti, Anchoring symbols to sensor data in single and multiple robot systems. Papers from the 2001 Fall Symposium. Technical Report FS-01-01, AAAI Press, North Falmouth, MA, 2001.

P. Davidsson, A Framework for organization and representation of concept knowledge in autonomous agents. In R. Trappl (ed.), Cybernetics and Systems '94, Vol.II, World Scientific Publishing, Singapore/London, pp.1427-1434, 1994.

P. Davidsson, On the concept of concept in the context of autonomous agents, Proc. of the 2nd World Conference on the Fundamentals of Artificial Intelligence (WOCFAI-95), pp. 85-96, 1995.

S. Harnad, Categorical Perception: The Groundwork of Cognition. Cambridge University Press, Cambridge, UK, 1987.

S. Harnad, The symbol grounding problem, Physica D, 42, pp. 335-346, 1990.

S. Harnad, Symbol grounding is an empirical problem, Proc.of the 15th Annual Conference of the Cognitive Science Society, Boulder, CO, Lawrence Erlbaum Associates, pp. 169-174, 1993.

M. Minsky, A framework for representing knowledge. In: P.H. Winston (ed.), The Psychology of Computer Vision, McGraw-Hill, London, 1975.

E. Prem, Semiosis in embodied autonomous systems, Proc. of the IEEE International Symposium on Intelligent Control, IEEE, Piscataway, NJ, pp.724-729, 1998.

E. Prem, Changes of representational AI concepts induced by embodied autonomy. In: Communication and Cognition – AI, the journal for the integrated study of AI, Cognitive Science and Applied Epistemology, Vol. 17 (3-4), pp. 189-208, 2000.

L. Steels, Language games for autonomous robots. In: Semisentient Robots, IEEE Intelligent Systems, 16 (5), pp. 16-22, 2001.

L. Steels, P. Vogt, Grounding adaptive language games in robotic agents. In: P.Husbands & I. Harvey (eds.), Fourth European Conference on Artificial Life (ECAL97), MIT Press/Bradford Books, Cambridge, MA, pp. 474-482, 1977.

J. von Uexküll, Theoretische Biologie (Theoretical Biology), Frankfurt/Main, Suhrkamp, 1928.