

Robosemiotics and embodied enactive cognition¹

Tom Ziemke
Dept. of Computer Science
University of Skövde
P.O. Box 408
S-541 28 Skövde
SWEDEN
tom@ida.his.se

© This paper is not for reproduction without permission of the author(s).

ABSTRACT

Much research in cognitive science and AI has recently been devoted to the study of adaptive autonomous agents, such as robots and artificial life systems. Such agents are typically said to ‘learn, ‘develop’ and ‘evolve’ in close interaction with their environments. It could be argued that these self-organizing properties solve the problem of representation grounding in AI research, and thus also place such ‘artificial organisms’ in a position of semiotic interest. This paper discusses the use of autonomous agents as models of sign processes and embodied enactive cognition. Furthermore, it addresses the question whether or to what extent such agents are themselves autonomous and capable of semiosis in that sense that living organisms are.

1 INTRODUCTION

Much research in cognitive science, and in particular artificial intelligence (AI) and artificial life, has since the mid-1980s been devoted to the study of so-called *autonomous agents*. These are typically robotic systems situated in some environment and interacting with it using sensors and motors. Such systems are often self-organizing in the sense that they artificially learn, develop and evolve in interaction with their environments, typically using computational learning techniques, such as artificial neural networks or evolutionary algorithms. Due to the biological inspiration and motivation underlying much of this research (cf. Sharkey & Ziemke 1998), autonomous agents are often referred to as ‘artificial organisms’, ‘artificial life’, ‘animats’ (short for ‘artificial animals’), ‘creatures’

¹ Most of the material in this paper has previously appeared elsewhere in very similar form (Ziemke 2001, 2002; Ziemke & Sharkey, 2001; Ziemke & Thieme; in press).

or ‘biorobots’. These terms do not necessarily all mean exactly the same; some of them refer to physical robots only, whereas others include simple software simulations. But the terms all express the view that the mechanisms referred to are substantially different from conventional artifacts, and that to some degree they are ‘life-like’ in that they share some of the properties of living organisms. Throughout this paper this class of systems will be referred to as ‘artificial organisms’ or ‘autonomous agents/robots’ interchangeably.

This paper addresses the semiotic status and relevance of such artificial organisms. One question is whether and to what extent they are autonomous and capable of semiosis. This is not straightforward since semiosis is often considered to necessarily involve living organisms. Morris (1946), for example, defines semiosis as “a sign-process, that is, a process in which something is a sign to some organism”. Similarly, Jakob von Uexküll¹ considered signs to be “of prime importance in all aspects of life processes” (T. von Uexküll 1992), and made a clear distinction between organisms, which as *autonomous subjects* respond to signs according to their own *specific energy*, and inorganic mechanisms, which lack that energy, and thus remain *heteronomous* (to be discussed in more detail later).

Mechanisms can, of course, be involved in sign processes, in particular computers and computer programs. Sebeok, for example, wrote (in personal communication cited by T. von Uexküll 1982) that “the criterial feature of living entities, and of machines programmed by humans, is semiosis”. The latter are, however, typically considered to lack ‘first hand semantics’, i.e., “intrinsic meaning” (Harnad 1990) or “contents for the machine” (Rylatt et al. 1998), and to derive their semantics from the fact that they are programmed, observed and/or interpreted by humans. Andersen et al. (1997) have argued in detail that computers/programs, when it comes to semiosis, fall somewhere in between humans and conventional mechanisms, but that they ultimately derive their semiotic ‘capacities’ from the interpretation of their designers and users. The major difference, they argued, was that living systems are autopoietic, i.e., self-creating and -maintaining, whereas machines are not (this issue will be discussed in further detail later). Hence, their “tentative conclusion” was that

... the difference between human and machine semiosis may not reside in the particular nature of any of them. Rather, it may consist in the condition that machine semiosis presupposes human semiosis and the genesis of the former can be explained by the latter. (Andersen et al. 1997: 569)

Cognitive science and AI research has traditionally, since its beginning in the 1950s, been dominated by the so-called *computer metaphor for mind*, i.e., the view that the human mind works very much like a computer program. This has led decades of traditional AI research to fall into the *internalist trap* (Sharkey & Jackson 1994) of focusing solely on

disembodied computer programs and internal *representations* supposed to ‘mirror’ a pre-given external reality (cf. Varela et al. 1991), while forgetting about the need for grounding and embedding these in the world they were actually supposed to represent. Hence, for cognitive scientists the use of embodied, situated agents offers an alternative, bottom-up approach to the study of intelligent behavior in general, and internal representation and sign usage in particular.

Artificial organisms, unlike computer programs equipped with robotic capacities of sensing and moving, *do* interact with their environments, and they appear to do so independently of interpretation through external users or observers. Moreover, such systems are often *self-organizing* in the sense that they ‘learn’, ‘develop’ and ‘evolve’ in interaction with their environments, using computational intelligence techniques and often attempting to mimic biological processes. Some examples of this type of self-organization in adaptive robots will be illustrated in Section 2. The sign processes and functional circles by which artificial organisms interact with their environments are therefore typically self-organized, i.e., the result of adaptation in interaction with an environment, rather than programmed or built-in by a designer, and thus often not even interpretable to humans (cf. Prem 1995). Hence, unlike computer programs, their genesis typically can *not* be explained with reference to human design and interpretation alone. Thus it has been argued that autonomous agents are, at least in theory, capable of possessing ‘first hand semantics’ (e.g., Harnad 1990; Franklin 1997; Bickhard 1998). Their semiotic and epistemological interest, it might be held, arises because unlike conventional machines, their use of signs and representations is self-organized, and thus, as for living systems, largely private and typically only meaningful to themselves. Many researchers therefore no longer draw a strict line between animals and autonomous robots. Prem (1998), for example, refers to both categories as ‘embodied autonomous systems’, and does not at all distinguish between living and non-living in his discussion of semiosis in such systems. We will return to this question in Section 3, after providing some examples of our own experimental work with robots in Section 2.

2 EXPERIMENTAL WORK: ROBOTS AND RECURRENT NEURAL NETS

Much research in *adaptive robotics* is concerned with making robots’ control mechanisms, mapping sensory signals to motor commands, adaptive using artificial evolutionary or learning techniques. In particular artificial neural networks (ANNs) are commonly used as ‘artificial nervous systems’ connecting a robot’s receptors to its effectors. The robots used in this type of research are often mobile robots (see Figure 1 for a typical example), typically receiving sensory input from, e.g., infrared proximity sensors or simple cameras, and controlling the motion of their wheels by motor outputs.

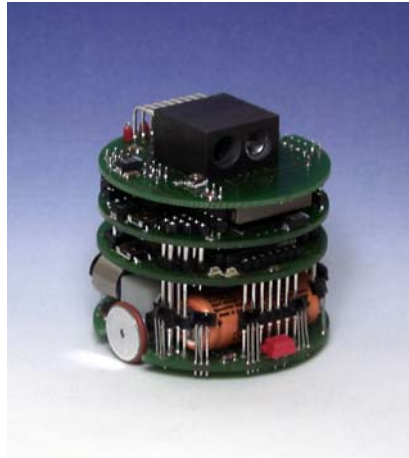


Figure 1: The Khepera, a wheeled miniature mobile robot commonly used in adaptive robotics research (manufactured by K-Team SA; for details see Mondada et al. 1993). The model shown here is equipped with eight short-range infrared sensors around the body and a simple long-range camera on top.

One interesting aspect of this type of research is that it, at least to some degree, offers a way around the problem of *symbol* or *representation grounding* in AI research (cf. e.g. Searle 1980; Harnad 1990; Sharkey & Jackson 1994; Bickhard 1998; Ziemke 1999). This is due to the fact that the controller network can actually, by means of the robot body (sensors and effectors), interact with the physical objects in its environment, independent of an observer's interpretation or mediation. Hence, it could be argued that its internal signs/representations, now formed in physical interaction with the world they 'represent' or reflect, can be considered (physically) grounded.

Our own research focuses on the use of recurrent neural networks, which will be discussed further in Section 2.1, before presenting some concrete examples in Section 2.2.

2.1 RECURRENT ANNS

As long as we are using a feed-forward network, i.e., a network in which activation is only passed in one direction, namely from input to output units, the mapping from input to output will always be the same (given that the network has already learned and does not modify its connection weights anymore). Hence, the controlled robot will be a "trivial machine" (cf. T. von Uexküll 1997), i.e., independent of past or input history, same inputs will always be mapped to same outputs. In semiotic terms, this corresponds to a semiosis of information where the input corresponds to the sign, the input-output mapping to the interpretant (or causal rule) and the output to the signified (T. von Uexküll 1997a).

However, if we add internal feedback through recurrent connections to the network it becomes a “non-trivial” machine. That means, the mapping from input to output will vary with the network’s internal state, and thus the machine, depending on its past, can effectively be a ‘different’ machine in each time step. An analogy in semiotic terms could be a semiosis of symptomization (cf. T. von Uexküll 1997) where the interpretant varies and the system’s input-output behavior can inform an observer about the current interpretant. For the robot itself this means that it no longer merely reacts to ‘external’ stimuli, but it ‘interprets’ incoming stimuli/signs depending on its own internal state.

Recurrent neural networks play an important role in the study and modeling of cognitive representation and their construction. This is due to the fact that they account for both the (long-term) representation of learning experience in connection weights as well as the (short-term) representation of the controlled agent’s current context or immediate past in the form of internal feedback. Moreover, Peschl (1997) has pointed out that RNNs, like real nervous systems, are “structure determined” (cf. also Maturana and Varela 1980), which means that their reaction to environmental stimuli always depends on the system’s current state (or structure), and thus is never determined by the input alone. Peschl referred to this as the “*autonomy* of a representational system”. He further argued that in such recurrent networks traditional concepts of knowledge representation (as a ‘mirror’ of external reality) are not applicable due to the fact that there is “no stable representational relationship of reference”. Hence, the “goal of representation” in such systems, he argued, could not be to achieve an accurate mapping of an external environment to internal referential representations. Instead, recurrent neural systems should be viewed as “physical dynamical devices embodying the (transformation) knowledge for sensorimotor [input-output] integration and generating adequate behavior enabling the organism’s survival”. Peschl’s view of knowledge as mechanisms of adequate sensorimotor transformation becomes particularly clear in his characterization of knowledge as “representation without representations”:

The internal structures do not map the environmental structures; they are rather responsible for generating functionally fitting behaviour which is *triggered* and *modulated* by the environment and *determined* by the internal structure (... of the synaptic weights). It is the result of *adaptive* phylo- and ontogenetic processes which have changed the architecture over generations and/or via learning in an individual organism in such a way that its physical structure embodies the dynamics for maintaining a state of equilibrium/homeostasis. (Peschl 1997)

Aware of the limitations of disembodied neural networks, Peschl further suggested a *system relative* concept of representation as “determined not only by the environment”, but also highly dependent on “the organization, structure, and constraints of the representation system as well as the sensory/motor systems which are embedded in a particular body structure”.

As discussed in detail elsewhere (Ziemke 2001), this view of the role of knowledge/representation in general, and the use of recurrent neural networks in adaptive robots in particular, is largely compatible with an enactive view of cognition, as put forward by Varela et al. (1991), and the closely related view of *radical constructivism* which von Glasersfeld (1995) summarized as follows:

- Knowledge is not passively received either through the senses or by way of communication;
- knowledge is actively built up by the cognizing subject.
- The function of cognition is adaptive, in the biological sense of the term, tending towards fit or viability;
- cognition serves the subject's organization of the experiential world, not the discovery of an objective ontological reality. (von Glasersfeld 1995: 51)

2.2 EXAMPLES OF EXPERIMENTAL WORK

This section will just very briefly describe two examples of experimental work in our lab. For details the reader is referred to the original papers/reports. The first example is illustrated in Figure 2. In these experiments (Thieme 2002; Thieme & Ziemke 2002; Ziemke & Thieme in press), simulated robots of the type illustrated in Figure 1, controlled by different types of recurrent and non-recurrent networks, were trained, using evolutionary algorithms, to handle mazes of different complexity. In the example illustrated in Figure 2 the robot always starts in the lower left part of the environment and has to reach the goal area (indicated by a white circle). The direction(s) in which it should turn in the T-junctions are indicated by the two lights that it passes in the first corridor on its way towards the first junction. In this particular case the side on which the first light appears indicates the correct direction in the first junction whereas the meaning of the second light is reversed, i.e., it appears on the side towards which the robot should *not* turn in the second junction.

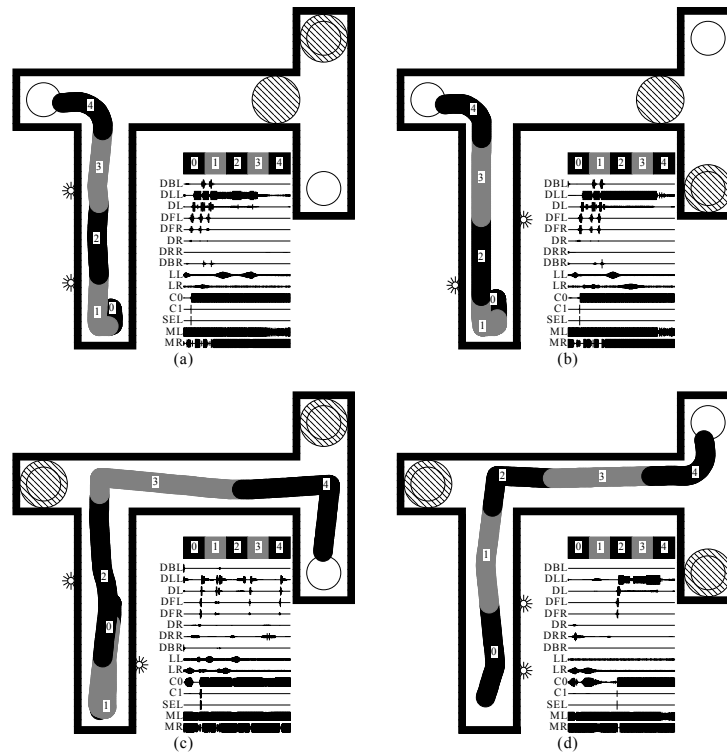


Figure 2: A simulated robot (and its trajectories) in four instantiations of a multiple T-maze task. Furthermore the activations of sensors (D), motors (M) and internal units in the controller network are illustrated over time. Adapted from Thieme (2002) and Ziemke & Thieme (in press).

Of course, the robot is not given any of the above information, but has to ‘find out’ itself in the course of the evolutionary training process how to make sense of its sensorimotor interaction with environment, i.e., what meaning to attribute to different stimuli in which context. Having full access to everything that happens in the controller network and in the environment allows us to analyze in detail exactly how the robot learns to do this (for details see Thieme 2002; Ziemke & Thieme in press).

The second example is illustrated in Figure 3. These experiments (Buason 2002) were concerned with the competitive co-evolution of simulated predator and prey robots (of the general type illustrated in Figure 1). That means, the predator’s task was to catch the prey, whereas the prey’s task was to avoid being caught. In a series of experiments of increasing complexity not only the neural controllers, but also different aspects of the robot morphology were evolved, such as direction, range and angle of the camera. Certain constraints were imposed on the robots, e.g. for more powerful vision they had to ‘pay’ with reduced speed.

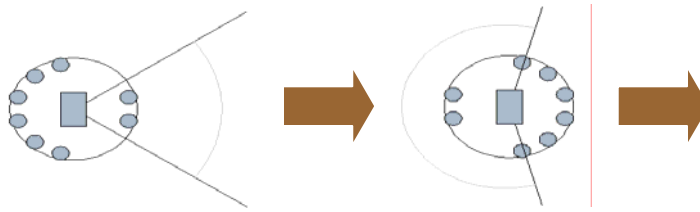


Figure 3: Predator robot (left) and prey robot (right), their visual fields (indicated by black lines and gray curves) and the directions in which they move. The small gray circles indicate the positions of the short-range infrared sensors, six on one side and two on the other.

In the example illustrated in Figure 3 the predator robot has evolved a relatively narrow visual field (indicated by the light gray curve), and consequently higher speed, whereas the prey has evolved a relatively wide visual field. It should be noted that both of them have evolved to have the camera facing the opponent, i.e., the prey looks ‘backwards’, i.e., towards the side with fewer infrared sensors, while moving ‘forward’. Moreover, both of them have the camera pointing in the direction where they have fewer short-range infrared sensors, i.e., the predator actually moves ‘backwards’ and looks for the prey in the same direction. As in the first example, having full access to the internal neural mechanisms and the robots’ behavior in the environment, allows us to analyze in detail the ongoing sign processes and their co-evolution over generations (for details see Buason 2002).

3 DISCUSSION

The question to what degree robots are autonomous and thus capable of semiosis has recently occupied a number of researchers in cognitive science and semiotics (e.g. Emmeche 2001; Nöth 2001; Sharkey & Ziemke 1998; Ziemke & Sharkey 2001; Ziemke 2001, 2002; Zlatev 2001). The question what exactly the semiotic status of such a robot would be apparently has no simple answer. The reason is that the distinction between organisms and mechanisms has become somewhat blurred, as also indicated by terms such as ‘artificial organisms’ or ‘artificial life’. Many would argue that in the case of robots which self-organize and develop in long-term interaction with their environment, relatively independent of their human designers, as illustrated with a couple of simple examples in the previous section, it is simply not the case that the genesis of robosemiosis can be (fully) explained with reference to human semiosis.

The ‘problem’ that makes it difficult, at least at a first glance, to make a sharp distinction between living organisms and today’s adaptive robots and artificial organisms, is that the latter nowadays have a number of the qualities/properties of the former. Ziemke and Sharkey (2001), for example, discussed in detail that three properties which Jakob von Uexküll (1928, 1982) considered unique for organisms (adaptation/growth, use of signs, centrifugal construction) can to some degree also be found in today’s robots. Similarly, Nöth (2001: 695-696) identified “four reasons why robots interact in the same way with their environment as organisms do” which “support the argument that not only organisms, but also robots have an Umwelt in [von] Uexküll’s sense”: (a) both robots and organisms have an Umwelt (or in fact Merkwelt) in the sense that, limited by available senses/sensors, they can only sense part of their physical environment; (b) both process environmental stimuli selectively; (c) both can have “internal representations of their Umwelt”; (d) both are equipped with perceptual organs/modules and effector organs/modules.

There are, however, also a number of differences remaining. Robot bodies are, for example, in many ways extremely different from living bodies, and thus unlikely candidates for supporting the same kind of phenomenal mind or Umwelt. In particular, robot bodies (hardware) and control systems (software) are not at all integrated the way living bodies are. Robot bodies do, for example, not grow. Furthermore, Ziemke & Sharkey (2001) argued in detail that robots lack *endosemiosis* and therefore also lack what T. von Uexküll *et al.* (1993) referred to as the *neural counterbody*, formed and updated in human brains as a result of the continual information flow of proprioceptive signs from the muscles, joints and other parts of our limbs, and thus giving rise to the experience of the living body as the center of subjective reality.

Furthermore, we have argued in detail elsewhere (Sharkey & Ziemke 1998, 2001; Ziemke & Sharkey 2001; Ziemke 1999, 2001), that the key to understanding mind and semiosis might be to understand the *autonomous* and *autopoietic*, i.e., self-creating and –maintaining, nature of living systems (Maturana & Varela 1980). Autopoietic systems are far more integrated than today’s robots in the sense that their ontogenesis does in fact start from a single cell from which they grow in a centrifugal fashion (von Uexküll 1982; cf. Ziemke & Sharkey 2001; Ziemke 2001). Moreover, they have a natural intentionality in the sense that they are autonomous unities naturally concerned with assimilation/dissimilation of material from/into their environment for the purpose of self-maintenance and survival, whereas our predator- and prey-robots really only are struggling for survival in the eye of the observer. Similarly, Nöth (2001: 696-697) concluded his discussion of whether or not robots have an *Umwelt* as follows:

Needless to say, a machine, in spite of a certain autonomy in its agency, can never be said to have its ultimate goal within itself. The objectives of a machine have always been established from outside, namely by the engineer who designed it and the user who switches it on and off. Thus, the robot's ultimate framework of reference, its final causality, is elsewhere, and thus the resulting semiotic process is allreferential.

ACKNOWLEDGEMENTS

I would like to thank João Queiroz for inviting me to present this paper, and Mikael Thieme and Gunnar Buason for carrying out the experiments mentioned in Section 2. This work was supported by a grant (1507/97) from the Knowledge Foundation, Stockholm.

REFERENCES

- Andersen, Peter B., Hasle, Per, and Brandt, Per A. 1997. Machine semiosis. In Posner, R., Robering, K., and Sebeok, T. A., editors, *Semiotik / Semiotics - Ein Handbuch zu den zeichentheoretischen Grundlagen von Natur und Kultur / A Handbook on the Sign-Theoretic Foundations of Nature and Culture*, pages 548-571. Walter de Gruyter, Berlin / New York.
- Bickhard, Mark. 1998. Robots and representations. In *From animals to animats 5 - Proceedings of the Fifth International Conference on Simulation of Adaptive Behavior*, pages 58-63, Cambridge, MA. MIT Press.
- Buason, Gunnar. 2002. *Competitive Co-Evolution of Sensory-Motor Systems*. Masters dissertation HS-IDA-MD-02-004, Department of Computer Science, University of Skövde, Sweden.
- Franklin, Stan 1997. Autonomous agents as embodied AI. *Cybernetics and Systems*, 28(6):499-520.
- Harnad, Stevan 1990. The symbol grounding problem. *Physica D*, 42:335-346.
- Maturana, Humberto R. and Varela, Francisco J. 1980. *Autopoiesis and Cognition - The Realization of the Living*. D. Reidel Publishing, Dordrecht, Holland.
- Mondada, Francesco, Franzi, E., and Ienne, P. 1993. Mobile robot miniaturisation: A tool for investigating in control algorithms. In *Third International Symposium on Experimental Robotics*, Kyoto, Japan.
- Morris, Charles W. 1946. *Signs, Language, and Behavior*. Prentice Hall, Englewood Cliffs.
- Nöth, Winfried 2001. Semiosis and the Umwelt of a robot. *Semiotica*, 134(1-4): 695-699.
- Peschl, Markus 1997. The representational relation between environmental structures and neural systems: Autonomy and environmental dependency in neural knowledge representation. *Nonlinear Dynamics, Psychology and Life Sciences*, 1(3).
- Prem, Erich. 1995. Understanding complex systems: What can the speaking lion tell us? In Steels, L., editor, *The Biology and Technology of Autonomous Agents*. Springer, Berlin, Germany. NATO ASI Series F, Vol. 144.
- Prem, Erich 1998. Semiosis in embodied autonomous systems. In *Proceedings of the IEEE International Symposium on Intelligent Control*, pages 724-729, Piscataway, NJ. IEEE.
- Rylatt, Mark, Czarnecki, Chris A., and Routen, Tom W. 1998. Beyond physical grounding and naive time: Investigations into short-term memory. In *From animals to animats 5 - Proceedings of the Fifth International Conference on Simulation of Adaptive Behavior*, pages 22-31, Cambridge, MA. MIT Press.

- Searle, John 1980. Minds, brains and programs. *Behavioral and Brain Sciences*, 3:417-457.
- Sharkey, Noel E. and Jackson, Stuart A. 1994. Three horns of the representational trilemma. In Honavar, Vasant and Uhr, L., editors, *Symbol Processing and Connectionist Models for Artificial Intelligence and Cognitive Modeling: Steps towards Integration*. Academic Press.
- Sharkey, Noel E. and Ziemke, Tom 1998. A consideration of the biological and psychological foundations of autonomous robotics. *Connection Science*, 10(3- 4):361-391.
- Sharkey, Noel E. and Ziemke, Tom 2001. Mechanistic vs. Phenomenal Embodiment: Can Robot Embodiment Lead to Strong AI?. *Cognitive Systems Research*, 2(4):251-262
- Thieme, Mikael 2002. Intelligence without hesitation. Masters dissertation HS-IDA-MD-02-001, Department of Computer Science, University of Skövde, Sweden.
- Varela, Francisco J.; Thompson, Evan and Rosch, Eleanor 1991. *The Embodied Mind Cognitive Science and Human Experience*. MIT Press, Cambridge, MA.
- von Glasersfeld, Ernst 1995. *Radical Constructivism – A Way of Knowing and Learning*. London: Falmer Press.
- von Uexküll, Jakob 1928. *Theoretische Biologie*. Springer Verlag, Berlin.
- von Uexküll, Jakob 1982. The Theory of Meaning. *Semiotica*, 42(1):25-82.
- von Uexküll, Thure 1982. Introduction: Meaning and science in Jakob von Uexküll's concept of biology. *Semiotica*, 42(1):1-24.
- von Uexküll, Thure 1992. Introduction: The sign theory of Jakob von Uexküll. *Semiotica*, 89(4):279-315.
- von Uexküll, Thure 1997. Biosemiose. In Posner, R., Robering, K., and Sebeok, T. A., editors, *Semiotik / Semiotics - Ein Handbuch zu den zeichentheoretischen Grundlagen von Natur und Kultur / A Handbook on the Sign-Theoretic Foundations of Nature and Culture*, pages 447-457. Walter de Gruyter, Berlin / New York.
- von Uexküll, Thure; Geigges, Werner, and Herrmann, Jörg M. 1993. Endosemiosis. *Semiotica*, 96(1/2):5-51.
- Ziemke, Tom 1999. Rethinking Grounding. In Riegler, A.; Peschl, M. and von Stein, A., editors, *Understanding Representation in the Cognitive Sciences*. Plenum Press, New York.
- Ziemke, Tom 2001. The Construction of 'Reality' in the Robot. *Foundations of Science*, 6(1):163-233.
- Ziemke, Tom 2002. On the Epigenesis of Meaning in Robots and Organisms. *Sign Systems Studies*, 30(1), 101-111.

Ziemke, Tom and Sharkey, Noel E., editors 1998. Biorobotics. Special issue of *Connection Science*, 10(3-4).

Ziemke, Tom and Sharkey, Noel E. 2001 A stroll through the worlds of robots and animals: Applying Jakob von Uexküll's theory of meaning to adaptive robots and artificial life. *Semiotica*, 134(1-4):653-694

Ziemke, Tom and Thieme, Mikael in press. Neuromodulation of reactive sensorimotor mappings as a short-term memory mechanism in delayed response tasks. *Adaptive Behavior*, to appear.

Zlatev, Jordan 2001. The epigenesis of meaning in human beings, and possibly in robots. *Minds and Machines*, 11, 155-195.