Noname manuscript No.
(will be inserted by the editor)

A Short review of Symbol Grounding in Robotic and Intelligent Systems

Silvia Coradeschi · Amy Loutfi · Britta Wrede

Received: date / Accepted: date

Abstract This paper gives an overview of the research papers published in Symbol Grounding in the period from the beginning of the 21st century up 2012. The focus is in the use of symbol grounding for robotics and intelligent system. The review covers a number of subtopics, that include, physical symbol grounding, social symbol grounding, symbol grounding for vision systems, anchoring in robotic systems, and learning symbol grounding in software systems and robotics. This review is published in conjunction with a special issue on Symbol Grounding in the Künstliche Intelligenz Journal.

Keywords Symbol Grounding · Anchoring · Cognitive Robotics · Social Symbol Gronding

1 Introduction

The main dream of Artificial Intelligence has been to create autonomous and intelligent systems that can reason and act in the real world. For such a dream to become true an essential ingredient is to establish and maintain a connection between what the system reasons about and what it can sense in the real world. This can be considered as an aspect of the Symbol Grounding Problem. The Symbol Grounding Problem (SGP) has been defined by Harnad in [29] as the problem of how to ground the meanings of symbol tokens in anything different than other (meaningless) symbols. Since its definition, symbol grounding has been an area of interest both in the fields of psychology as well as artificial intelligence. Its practical application has also been studied in robotics and intelligent systems, with particular emphasis on the problem of grounding symbols to the data acquired by physically embedded sensors. It is this practical application which is the focus of this paper. The review covers the recent literature in the subject and in particular the period from 2000 to 2012 and is organized into two subtopics which relate to the current approaches to SGP in robotics and intelligent systems: Physical Symbol Grounding and Social Symbol Grounding. The "Physical Symbol Grounding" as been defined by Vogt in [74] as the grounding of symbols to real world objects by a physical agent interacting in the real world; while its social component, "Social Symbol Grounding", refers to the collective negotiation for the selection of shared symbols (words) and their grounded meanings in (potentially large) populations of agents as defined by Cangelosi in [10].

These are both significant and hard problems. As explained in [74] Physical Symbol Grounding requires constructing a consistent relation between percepts that may vary under different conditions, and which often have a high dimensionality. Categorising the dimensionalities may yield different categories, which however should be related to one concept often with the help of invariant feature detectors. According to Vogt [77] the social symbol grounding problem may even be a harder problem to solve, because to learn what a word-form refers to can result in Quine's referential indeterminacy problem: the unknown word can -theoretically- refer to an infinite number of objects. Vogt investigated in [77] a number of heuristics from child language acquisition literature that help to reduce this indeterminacy: joint attention, principle of contrast and corrective feedback. In [76] mutual exclusivity, and a few potential dialogues

Örebro University, AASS
Bielefeld University
E-mail: silvia.coradechi@oru.se E-mail: amy.loutfi@oru.se
E-mail: bwrede@techfak.uni-bielefeld.de

that help to reduce referential indeterminacy have been also implemented.

It is worth to note that a recent review of Symbol Grounding has been published in 2005 by Taddeo [69] which specifically addresses SGP as a general problem from a philosophical perspective. A volume edited by Belpaeme [2] presents current views on symbol grounding both from a philosophical and robotics perspective.This review is complementary as it focuses on work of more practical relevance to robotics and intelligent systems. Finally Cangelosi in [12] discusses the current progress and solutions to the symbol grounding problem and specifically identifies which aspects of the problem have been addressed and issues and scientific challenges that still require investigation.

It is the authors belief that the focus of this review is especially timely as the steps towards the solution of the SGP will be key to creating the next generation of robotic systems that are capable of high level reasoning.

The review is structured in a number of subtopics. In the Physical Symbol Grounding section learning of categories on the basis of sensor data and grounding of actions are considered. In addition the concept of Anchoring of symbols to sensor data is defined and the work in this topic is summarized. The review ends with a summary of works in Social Symbol Grounding and works on Symbol Grounding applied to the semantic web.

2 Physical Symbol Grounding

When dealing with the Physical Symbol Grounding, one of the basic challenges examined in the literature is to ground symbols to perceptual representations (sensor data), where the symbols denote categorical concepts such as color, shape and spatial features. Typically, the sensor data come from vision sensors but other modalities have also been used. The methods explored are often inspired by connectionist models and a wide range of learning algorithms have been applied. Unsupervised methods have been investigated by Vavrecka in [73] where a biologically inspired model for grounding spatial terms is presented. Color, shape and spatial relations of two objects in 2D space are grounded. Images with two objects are presented to an artificial retina and five-word sentences describing them (e.g. "Red box above green circle") are inputed. The implementation is done using Self-Organizing Map and Neural Gas algorithms. The Neural Gas algorithm is found to lead to better performance especially in case of scenes with higher complexity. In [36] Kittler considers a visual bootstrapping approach for the unsupervised symbol grounding. The method is based on a recursive clustering of a perceptual category domain controlled by goal acquisition from the visual environment.

A supervised method is used in a framework for modeling language in neural networks and adaptive agent simulations by Cangelosi [9]. In this work symbols are directly grounded into the agents' own categorical representations and have syntactic relationships with other symbols. The grounding of basic words, acquired via direct sensorimotor experience, is transferred to higher-order words via linguistic descriptions.

Emphasizing the dynamic nature of language, Pastra suggests that Symbol Grounding is a bi-directional process (double-grounding) [55, 56]; its use in artificial intelligence agents allows one to tie symbols of different levels of abstraction to their sensorimotor instantiations (catering thus for disambiguation) and at the same time, to untie sensorimotor representations from their physical specificities correlating them to symbolic structures of different levels of abstraction (catering thus for intentionality indication). In other words, going bottom up (from sensorimotor representations to symbols) the agent acquires a hierarchical composition of human behaviour, while going top-down (from symbols to sensorimotor representations) the agent gets intentionality-laden interpretations of those structures.

Such two-way grounding has been captured in an automatically built knowledge base, the PRAXICON, which comprises a semantic network of embodied concepts and pragmatic relations [57, 58]. The concepts have multiple representations (linguistic, visual, motoric) and their rich relational network builds upon findings from neuroscience that have led to an action-centric structure of the network [59]. This is a semantic memory-like module with its own reasoning mechanism for allowing an agent to generalise over learned schemas and behaviours and deal with unexpected situations creatively. Both the reasoner and the language processing tools for the automatic population of such memory module advocate the embodied cognition perspective, coupling symbols to their references, dealing with abstract concepts and their indirect grounding to sensorimotor experiences, as well as with figurative language phenomena, such as metonymy and metaphor [61, 62].

The tools and knowledge bases developed within the double-grounding perspective have been employed in a number of robotic applications with the iCub humanoid, including (a) the robot doer in response to verbal requests for performing everyday activities and (b) the robot active observer in visual scenes where the actions of a human are being observed and verbalised by the robot [1] [60].

[1] POETICON++ and POETICON projects (2008-2015) at http://www.poeticon.eu and http://www.csri.gr/Poeticon

A few works consider the combination of both visual and auditory data, where the combination gives a better result than using one modality alone. In [78] a multimodal learning system is presented by Yu that can ground spoken names of objects in their physical referents and learn to recognize those objects simultaneously from vocal and vision input. The system collects image sequences and speech input while users perform natural tasks and grounds spoken names of objects in visual perception, also learning to categorize visual objects using teaching signals encoded in co-occurring speech. Also Nakamura uses in [49, 50] vision and speech for multimodal categorization and words grounding by robots. The robot uses its physical embodiment to grasp and observe an object from various view points, as well as to listen to the sound during the observing period. The method used is Latent Dirichlet allocation (LDA)-based framework and experimental results with 40 objects (eight categories) show an improvement with respect to just visual categorization and show the possibility of a conversation between a user and the robot based on the grounded words. In [51] a system involving vision and audio data is presented by Needham that is capable of autonomously learning concepts (utterances and object properties) from perceptual observations of dynamic scenes. This work goes beyond categorical learning and learns also protocols from the perceptual observations. The motivation is the development of a synthetic agent that can observe a scene containing interactions between unknown objects and agents, and learn models of these sufficient to act in accordance with the implicit protocols present in the scene. The system is tested by learning the protocols of simple table-top games where perceptual classes and rules of behaviors from real world audio-visual data is learnt in an autonomous manner.

Additional sensor modalities have been used by Grollman in [28] where symbol grounding in robot perception is considered through a data-driven approach deriving categories from robot sensor data that include infrared, sonar and data from a time-of-flight distance camera. Isomap nonlinear dimension reduction and Bayesian clustering (Gaussian mixture models) with model identification techniques are used to discover categories. Trials in various indoor and outdoor environments with different sensor modalities are presented and the learned categories are then used to classify new sensor data.

2.1 Perceptual Anchoring

A special case of Symbol Grounding is the connection of sensor data coming from physical objects to higher level symbolic information that refers to those objects. The process of creating and maintaining this connection is called Anchoring and has been formally defined by Coradeschi in [17] and then in [18].

The use of anchoring in planning, recovery planning and solving of ambibuities is explored in works of Karlsson and Broxvall [5, 6, 35] Anchoring with other sensor modalities like olfaction is explored in works of Loutfi and Broxvall [7, 40–42] while the integration of high-level conceptual knowledge on a single agent, via the combination of a fully-fledged Knowledge Representation and Reasoning (KR&R) system with the anchoring framework and more specifically, the use of semantic knowledge and common-sense information so as to enable reasoning about the perceived objects at the conceptual level has been considered by Lemaignan and Daoutis in [22, 38]. Cooperative anchoring among robots in a robot soccer application is presented by LeBlanc in [37] while multi-agent anchoring in a smart home environment is presented in works of Broxvall and Daoutis [7, 20].

A framework for computing the spatial relations between anchors is presented by Melchert in [43–45] where a set of binary spatial relations were used to provide object descriptions. Human interaction is used to disambiguate between visually similar objects. Similarly in [46] an approach to establish joint object reference is formulated by Moratz. The object recognition approach assigns natural categories (e.g. "desk", "chair", "table") to new objects based on their functional design, relations (e.g. "the briefcase to the left of the chair") are then established allowing users to refer to objects which cannot be classified reliably by the recognition system alone.

Anchoring has also been used by Lemaignan [39] to enable a grounded and shared model of the world that is suitable for dialogue understanding. Realistic human-robot interactions are considered that deal with complex, partially unknown human environments and a fully embodied (with arms, head,...) autonomous robot that manipulates a large range of household objects. A knowledge base models the beliefs of the robot and also every other cognitive agent the robot interacts with. A framework is also presented to extract symbolic facts from complex real scenes. The robot builds a 3D model of the world on-line by merging different sensor modalities. It computes spatial relations between perceived objects in realtime and the system allows virtually viewing of the same scene from different points of view.

A different approach to anchoring is presented by Heintz in [30,32,33] where anchoring is considered in the context of unmanned aerial vehicles. In their stream-based hierarchical anchoring framework, a classification hierarchy is associated with expressive conditions for hypothesizing the type and identity of an object given streams of temporally tagged sensor data. A metric spatio-temporal logic is used to represent the conditions which are efficiently evaluated over these streams using a progression-based technique. The anchoring process constructs and maintains a set of *object linkage structures* representing the best possible hypotheses at any time. Each hypothesis can be incrementally generalized or narrowed down as new sensor data arrives. Symbols can be associated with an object at any level of classification, permitting symbolic reasoning on different levels of abstraction.

Additional approaches of anchoring are presented in a special issue on Anchoring published by the Robotics and Autonomous Systems Journal. In [64] an overview of the GLAIR approach to anchoring is outlined by Shapiro where abstract symbolic terms that denote an agent's mental entities are anchored to the lower-level structures used by the embodied agent to operate in the real (or simulated) world. In [75] the anchoring problem is approached by Vogt using semiotic symbols defined by a triadic relation between forms, meanings and referents. Anchors are formed between these three elements and a robotic experiment based on adaptive language games is presented that illustrates how the anchoring of semiotic symbols can be achieved in a bottom-up fashion. Person tracking using anchoring has been investigated by Fritsch in [25] where laser range data is used to extract the legs of a person while camera images from the upper body part are used for extracting the faces. The results of the different percepts, which originate from the same person are combined in one anchor for the person.

An interesting application of Anchoring is in the field of topological maps and in particular the investigation of the connection of symbolic information to spatial information. Work in this area has been presented by Galindo in [26, 27] where a multi-hierarchical approach is used to acquire semantic information from a mobile robot sensors for navigation tasks. The spatial information is anchored to the semantic information and the approach is validated via experiments where a mobile robot uses and infers new semantic information from its environment, improving its operation. Similarly Elmogy in [23] investigates how a topological map is generated to describe relationships among features of the environment in a more abstract form to be used in a robot navigation system. A language for instructing the robot to execute a route in an indoor environment is presented where an instruction interpreter processes a route description and generates its equivalent symbolic and topological map representations. Finally Blodow in [4] uses semantic mapping in kitchen environments to help performing manipulation tasks.

3 Grounding Words in Action

The research group headed by Cangelosi has been working in cognitive robotics models using the humanoid robot iCub. In [66, 67] a cognitive robotics model is described in which the linguistic input provided by the experimenter guides the autonomous organization of the knowledge of the iCub. A hierarchical organization of concepts is used for the acquisition of abstract words. Higher-order concepts are grounded using basic concepts and actions that are directly grounded in sensorimotor experiences. The method used is a recurrent neural network that permits the learning of higher-order concepts based on temporal sequences of action primitives. In [11] a review of cognitive agent and developmental robotics models of the grounding of language is presented. Three models are discussed: a multi-agent simulation of language evolution, a simulated robotic agent model for symbol grounding transfer, and a model of language comprehension in the humanoid robot iCub. The complexity of the agent's sensorimotor and cognitive system gradually increases in the three models. In previous works [13, 15] the combination of cognitive robotics with neural modeling methodologies is also considered to demonstrate how the language acquired by robotic agents can be directly grounded in action representations, in particular language learning simulations show that robots are able to acquire new action concepts via linguistic instructions. Finally in [14] an embodied model for the grounding of language in action is presented and experimented on epigenetic robots. Epigenetic robots have an integrative vision of language in which linguistic abilities are strictly dependent on and grounded in other behaviors and skills. Experiments done with simulated robots show that higher order behavioral abilities can be autonomously built on previously grounded basic action categories following linguistic interaction with human users.

Another approach to learning of actions is presented by Oladell in [54] where representational complexity is managed using a symbolic feature representation generated via policies, affordances and goals. The approach is demonstrated in a simulation environment with a robot arm and camera. Learning tasks revolve around lift, move, and drop and the policies are learnt using QLearning. The agent learns new policies, affordances and goals and adds them to the dictionary. After each addition, the best common sub-structure is extracted.

Learning of meanings of both action and substantive words is presented by Tellex in [70] where a probabilistic approach is used to learn word meanings from large corpora of examples and use those meanings to find good groundings in the external world. The framework handles complex linguistic structures such as referring expressions (for example, "the door across from the elevators") and multiargument verbs (for example, "put the pallet on the truck") by dynamically instantiating a conditional probabilistic graphical model that factors according to the compositional and hierarchical structure of a natural language phrase.

4 Social Symbol Grounding

A recent line of research in Symbol Grounding is Social Symbol Grounding. As defined by Cangelosi [10] the social symbol grounding considers the next step after the connections between the sensor data and symbols for individual agents are achieved, that is how can these connections be shared among many agents. Several approaches have been presented to address this issue. Heintz in [31] presents a distributed information fusion system for collaborative UAVs. In [65] Steels examines if a perceptually grounded categorical repertoire can become sufficiently shared among the members of a population to allow successful communication, using color categorization as a case study. Several models are proposed that are inspired by alternative hypotheses of human categorization. He has proposed various robotic models of the emergence of communication based on the languages games for the Talking Heads experiments and the AIBO and QRIO robots.The paper argues that the collective choice of a shared repertoire must integrate multiple constraints, including constraints coming from communication. Similarly Fontanari in [24] use language games to study evolution of compositional lexicons. In [77] the New Ties project is presented. The project aims at evolving a virtual simulated cultural society where the agents evolve a communication system that is grounded in their interactions with their virtual environment and with other individuals. An hybrid model of language learning involving joint attention, feedback, cross-situational learning and the principle of contrast is investigated. A number of experiments are carried out in simulation showing that levels of communicative accuracy better than chance evolve quite rapidly and that accuracy is mainly achieved by the joint attention and cross-situational learning mechanisms while feedback and the principle of contrast contribute less. As mentioned in the introduction the social symbol grounding problem is a difficult problem to solve, because an unknown word can -theoretically- refer to an infinite number of objects. Vogt investigated in [77] a number of heuristics from child language acquisition literature that help to reduce this indeterminacy: joint attention, principle of contrast and corrective feedback. In [76] mutual exclusivity, and a few potential dialogues that help to reduce referential indeterminacy have been also implemented. In [68] it is argued that the primary motivation for an agent to construct a symbol-meaning mapping is to solve a tasks, in particular it is investigated how agents learn to solve multiple tasks and extract cumulative knowledge that helps them to solve each new task more quickly and accurately.

The relevance of joint attention as found by [77] and the motivation to solve a joint task [68] indicate the relevance of taking cues arising from the current situation into account. Even more, Belpaeme & Cowley [3] argue that the symbol grounding problem as defined by Harnad [29] has to be extended to incorporate the process of language acquisition itself as language facilitates the acquisition of meaning [1].

Indeed, studies of parent-infant interaction indicate that parents help their infants to understand not simply the relationship between a symbol and a referent, but rather by making sense of a whole situation to them. They do so by presenting re-curring patterns of interaction that facilitate further learning of new items or actions in similar situations. Recurring patterns (or "pragmatic frames") contain important pragmatic information that help to decode the semantic information. More specifically, frames provide "predictable, recurrent interactive structures" [52] (p. 171) that scaffold the childs emerging understanding [72] as new linguistic labels will be perceived as a new slot within a familiar routine. Some robotic approaches already try to model these interactional cues by establishing frames to achieve Joint Attention through mutual gaze [47], guiding attention through saliency-based strategies [48], or to establish a temporal alignment through synchrony-based strategies [63] or the elicitation of contingent feedback [8]. These frames provide more information than the simple establishment of symbol-referent associations. Rather, they contain - among others - information about semantic roles (e.g. agent-patient relations but also about the nature of goals or constraints of actions, as well as success and failure) and thus semantic-syntactic relations which are important to enable generalisation to new situations. Understanding is thus seen as a continuous process rather than a (static) representation that establishes associations between symbols and internal sensori-motor concepts.

The process of how joint understanding of a shared situation can be achieved has also been formulated in a more formalised way through the step-wise process of "grounding" [16] which describes 4 levels (attention, signal decoding, semantic processing, intention recognition) that need to be grounded in order to achieve mutual understanding.

However, while well founded in infant development, these concepts yet lack the proof-of-concept that they indeed facilitate language learning by providing relevant information that - if taken into account - would significantly influence the learning dynamics. If these considerations hold true, this would mean to re-consider Cangelosi's definition of social symbol grounding as a second step that enables to share connections between percepts and symbols [10]: one would have to consider the social symbol grounding problem as the initial step that facilitates the acquistion of language and meaning without which no such relations can be learned.

5 Grounding symbols in the Semantic Web

Recent trends examine the Symbol Grounding Problem in the context of web technologies and specifically the semantic web. In [34] Semantic Web technologies are used by Johnston for grounding robotic systems. In particular the OBOC robotic software system including an ontology-based vision subsystem is presented. OBOC has been tested and evaluated in the robot soccer domain. The grounding of knowledge for everyday tasks using the World Wide Web has been considered by Nyga and Tenorth in [53, 71] while a first attempt of an extension of the anchoring framework to handle the grounding and integrate symbolic and perceptual data that are available on the web is outlined by Daoutis in [21].

The problem of giving semantics to the semantic web is considered by Cregan in [19]. The paper argues that the symbol grounding problem is of relevance for the Semantic Web as inappropriate correspondence between symbol and referent can result in logically valid but meaningless inferences. In fact ontology languages can provide a means to relate data items to each other in logically well-defined ways, but they are intricate "'castles in the air"' without a pragmatic semantics linking them in a systematic and unambiguous way to the real world entities they represent.

6 Conclusions

This short review presents recent work in Symbol Grounding that is focused on the use of symbol grounding in robotics and intelligent systems applications. The field is clearly very active and many articles have been published in recent years. This is a consequence of the current trends of integrating robots and distributed systems in unstructured and dynamic environments. Such environments require a flexible handling of knowledge and the connection of symbolic and sensory information to be able to successfully operate. In addition systems where humans have an active role are becoming more common. Here symbol grounding is essential to insure meaningful natural language communication. Finally the use of the web as a source of information about objects and their properties is providing new opportunities to access a very large and updated storage of data, both symbolic and visual. The use of symbol grounding to connect the information in the web to real data is maybe the most important challenge for the field.

Acknowledgements We would like to thank Tony Belpaeme, Fredrik Heintz, Sven Albrecht, Angelo Cangelosi, Paul Vogt, Katerina Pastra and Sverin Lemaignan for their helpful comments to improve the article and make it more complete.

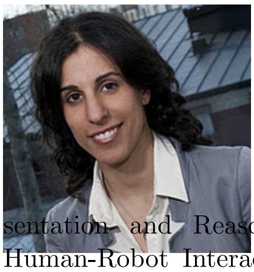
References

1. S. Harnad A. Cangelosie. The adaptive advantage of symbolic theft over sensorimotor toil: Grounding language in perceptual categories. *Evolution of Communication*, 4(1):117–142, 2001.
2. T. Belpaeme, S.J. Cowley, and K. MacDorman, editors. *Symbol grounding*. John Benjamins, Amsterdam, 2009.
3. Tony Belpaeme and Stephen J. Cowley. Extending symbol grounding. *Interaction Studies*, 8(1):1–16, 2007.
4. Nico Blodow, Lucian Cosmin Goron, Zoltan-Csaba Marton, Dejan Pangercic, Thomas Rühr, Moritz Tenorth, and Michael Beetz. Autonomous semantic mapping for robots performing everyday manipulation tasks in kitchen environments. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2011.
5. M. Broxvall, S. Coradeschi, L. Karlsson, and A. Saffiotti. Have another look on failures and recovery planning in perceptual anchoring. In *AAAI Workshop - Technical Report*, volume WS-04-03, pages 63–70, 2004.
6. M. Broxvall, S. Coradeschi, L. Karlsson, and A. Saffiotti. Recovery planning for ambiguous cases in perceptual anchoring. In *Proceedings of the National Conference on Artificial Intelligence*, volume 3, pages 1254–1260, 2005.
7. M. Broxvall, S. Coradeschi, A. Loutfi, and A. Saffiotti. An ecological approach to odour recognition in intelligent environments. In *Proceedings - IEEE International Conference on Robotics and Automation*, volume 2006, pages 2066–2071, 2006.
8. N. J. Butko and J. R. Movellan. Detecting contingencies: An infomax approach. *Neural Networks*, 23(8-9):973–984, 2010.
9. A. Cangelosi. Symbol grounding in connectionist and adaptive agent models. In *Lecture Notes in Computer Science*, volume 3526, pages 69–74, 2005.
10. A. Cangelosi. The grounding and sharing of symbols. *Pragmatics and Cognition*, 14(2):275–285, 2006.
11. A. Cangelosi. Grounding language in action and perception: From cognitive agents to humanoid robots. *Physics of Life Reviews*, 7(2):139–151, 2010.

12. A. Cangelosi. Solutions and open challenges for the symbol grounding problem. *International Journal of Signs and Semiotic Systems*, 1(1):49–54, 2011.
13. A. Cangelosi, E. Hourdakis, and V. Tikhanoff. Language acquisition and symbol grounding transfer with neural networks and cognitive robots. In *IEEE International Conference on Neural Networks - Conference Proceedings*, pages 1576–1582, 2006.
14. A. Cangelosi and T. Riga. An embodied model for sensorimotor grounding and grounding transfer: Experiments with epigenetic robots. *Cognitive Science*, 30(4):673–689, 2006.
15. A.a Cangelosi, V.a Tikhanoff, J.F.b Fontanari, and E.c Hourdakis. Integrating language and cognition: A cognitive robotics approach. *IEEE Computational Intelligence Magazine*, 2(3):65–70, 2007.
16. H. Clark. *Using language*. Cambridge University Press, 1996.
17. S. Coradeschi and A. Saffiotti. Anchoring symbols to sensor data: Preliminary report. In *In Proc. of the 17th AAAI Conf*, pages 129–135. AAAI Press, 2000.
18. S. Coradeschi and A. Saffiotti. An introduction to the anchoring problem. *Robotics and Autonomous Systems*, 43(2-3):85–96, 2003.
19. A.M. Cregan. Symbol grounding for the semantic web. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 4519 LNCS, pages 429–442, 2007.
20. M. Daoutis, S. Coradeschi, and A. Loutfi. Cooperative knowledge based perceptual anchoring. *International Journal on Artificial Intelligence Tools*, 21(3):44–87, 2012.
21. M. Daoutis, S. Coradeschi, and A. Loutfi. Towards concept anchoring for cognitive robots. *Intelligent Service Robotics*, 5(4):213–228, 2012.
22. M. Daoutis, S. Coradeshi, and A. Loutfi. Grounding commonsense knowledge in intelligent systems. *Journal of Ambient Intelligence and Smart Environments*, 1(4):311–321, 2009.
23. M. Elmogy, C. Habel, and J. Zhang. Multimodal cognitive interface for robot navigation. *Cognitive Processing*, 12(1):53–65, 2011.
24. J.F.a Fontanari and L.I.b c Perlovsky. Evolving compositionality in evolutionary language games. *IEEE Transactions on Evolutionary Computation*, 11(6):758–769, 2007. cited By (since 1996) 24.
25. J. Fritsch, M. Kleinehagenbrock, S. Lang, T. Plötz, G. A. Fink, and G. Sagerer. Multi-modal anchoring for human-robot-interaction. *Robotics and Autonomous Systems, Special issue on Anchoring Symbols to Sensor Data in Single and Multiple Robot Systems*, 43(2–3):133–147, 2003.
26. C. Galindo, J.-A. Fernandez-Madrigal, J. Gonzalez, A. Saffiotti, and P. Buschka. Life-long optimization of the symbolic model of indoor environments for a mobile robot. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 37(5):1290–1304, 2007.
27. C.a Galindo, A.a Saffiotti, S.a Coradeschi, P.a Buschka, J.A.b Fernandez-Madrigal, and J.b Gonzalez. Multi-hierarchical semantic maps for mobile robotics. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS*, pages 3492–3497, 2005.
28. D.H. Grollman, O.C. Jenkins, and F. Wood. Discovering natural kinds of robot sensory experiences in unstructured environments. *Journal of Field Robotics*, 23(11-12):1077–1089, 2006.
29. S. Harnad. The symbol grounding problem. *Physica D: Nonlinear Phenomena*, 42(1-3):335–346, 1990.
30. Fredrik Heintz and Patrick Doherty. Managing dynamic object structures using hypothesis generation and validation. In *Proceedings of the AAAI Workshop on Anchoring Symbols to Sensor Data*, 2004.
31. Fredrik Heintz and Patrick Doherty. Federated dyknow, a distributed information fusion system for collaborative UAVs. In *Proceedings of the International Conference on Control, Automation, Robotics and Vision (ICARCV)*, 2010.
32. Fredrik Heintz, Jonas Kvarnström, and Patrick Doherty. A stream-based hierarchical anchoring framework. In *Proceedings of IROS*, 2009.
33. Fredrik Heintz, Jonas Kvarnström, and Patrick Doherty. Bridging the sense-reasoning gap: DyKnow – stream-based middleware for knowledge processing. *Advanced Engineering Informatics*, 24(1):14–26, January 2010.
34. B. Johnston, F. Yang, R. Mendoza, X. Chen, and M.-A. Williams. Ontology based object categorization for robots. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5345 LNAI:219–231, 2008.
35. L. Karlsson, A. Bouguerra, M. Broxvall, S. Coradeschi, and A. Saffiotti. To secure an anchor - a recovery planning approach to ambiguity in perceptual anchoring. *AI Communications*, 21(1):1–14, 2008.
36. J. Kittler, M. Shevchenko, and D. Windridge. Visual bootstrapping for unsupervised symbol grounding. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4179 LNCS:1037–1046, 2006.
37. K. LeBlanc and A. Saffiotti. Cooperative information fusion in a network robot system. In *Proc of the Int Conf on Robot Communication and Coordination (RoboComm)*, Athens, Greece, 2007. Online at http://www.aass.oru.se/˜asaffio/.
38. S. Lemaignan, R. Alami, A. K. Pandey, M. Warnier, and J. Guitton. *Bridges between the Methodological and Practical Work of the Robotics and Cognitive Systems Communities - From Sensors to Concepts*, chapter Towards Grounding Human-Robot Interaction. Intelligent Systems Reference Library. Springer Publishing, 2012.
39. S. Lemaignan, R. Ros, E.A. Sisbot, R. Alami, and M. Beetz. Grounding the interaction: Anchoring situated discourse in everyday human-robot interaction. *International Journal of Social Robotics*, 4(2):181–199, 2012.
40. A. Loutfi and S. Coradeschi. Improving odour analysis through human-robot cooperation. In *Proceedings - IEEE International Conference on Robotics and Automation*, volume 2005, pages 4443–4449, 2005.
41. A. Loutfi and S. Coradeschi. Smell, think and act: A cognitive robot discriminating odours. *Autonomous Robots*, 20(3):239–249, 2006.
42. A. Loutfi and S. Coradeschi. Odor recognition for intelligent systems. *IEEE Intelligent Systems*, 23(1):41–48, 2008.
43. A. Loutfi, S. Coradeschi, M. Daoutis, and J. Melchert. Using knowledge representation for perceptual anchoring in a robotic system. *International Journal on Artificial Intelligence Tools*, 17(5):925–944, 2008.
44. J. Melchert, S. Coradeschi, and A. Loutfi. Knowledge representation and reasoning for perceptual anchoring. In *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI*, volume 1, pages 129–136, 2007.

45. J. Melchert, S. Coradeschi, and A. Loutfi. Spatial relations for perceptual anchoring. In *AISB'07: Artificial and Ambient Intelligence*, pages 459–463, 2007.
46. R. Moratz. Intuitive linguistic joint object reference in human-robot interaction: Human spatial reference systems and function-based categorisation for symbol grounding. In *Proceedings of the National Conference on Artificial Intelligence*, volume 2, pages 1483–1488, 2006.
47. Yukie Nagai. Learning for joint attention helped by functional development. *Advanced Robotics*, 20(10):1165–1181, 2006.
48. Yukie Nagai and Katharina J. Rohlfing. Computational analysis of motionese: Towards scaffolding robot action learning. *IEEE Transactions on Autonomous Mental Development*, 1(1):44–54, 2009.
49. T. Nakamura, T. Araki, T. Nagai, and N. Iwahashi. Grounding of word meanings in latent dirichlet allocation-based multimodal concepts. *Advanced Robotics*, 25(17):2189–2206, 2011.
50. T. Nakamura, T. Nagai, and N. Iwahashi. Grounding of word meanings in multimodal concepts using lda. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2009*, pages 3943–3948, 2009.
51. C.J. Needham, P.E. Santos, D.R. Magee, V. Devin, D.C. Hogg, and A.G. Cohn. Protocols from perceptual observations. *Artificial Intelligence*, 167(1-2):103–136, 2005. cited By (since 1996) 19.
52. A. Ninio and C. Snow. *Pragmatic Development*. WestviewPress, 1996.
53. Daniel Nyga, Moritz Tenorth, and Michael Beetz. Understanding and executing instructions for everyday manipulation tasks from the world wide web. In *ICRA*, 2009.
54. M. Oladell and M. Huber. Symbol generation and grounding for reinforcement learning agents using affordances and dictionary compression. In *Proceedings of the 25th International Florida Artificial Intelligence Research Society Conference, FLAIRS-25*, pages 132–135, 2012.
55. K. Pastra. Viewing vision-language integration as a double-grounding case. In *AAAI Fall Symposium - Technical Report*, volume FS-04-01, pages 62–69, 2004.
56. K. Pastra. *Vision-language integration: A double-grounding case*. PhD thesis, Department of Computer Science, University of Sheffield, 2005.
57. K. Pastra. Praxicon: the development of a grounding resource. In *Proceedings of the International Workshop on Human-Computer Conversation, Bellagio, Italy*, 2008.
58. K. Pastra. From lexicon to praxicon: language-action-image semantic relations. In *K. Potagas and I. Evdokimidis (eds), Conversations on Language and Action, Aiginiteion Series*, 2010.
59. K. Pastra, Y. Aloimonos, K. Pastra, and Y. Aloimonos. The minimalist grammar of action. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1585):103–117, 2012.
60. K. Pastra and Y. Aloimonos (eds.). *Technical Report of the AAAI 2011 Workshop on "Language-Action Tools for Cognitive Artificial Agents: Integrating Vision, Action and Language"*. AAAI, 2010.
61. K. Pastra, E. Balta, P. Dimitrakis, and G. Karakatsiotis. Embodied language processing: a new generation of language technology. In *Proceedings of the AAAI 2011 International Workshop on Language-Action Tools for Cognitive Artificial Agents: Integrating Vision, Action and Language*, 2011.
62. K. Pastra, P. Dimitrakis, E. Balta, and G. Karakatsiotis. Praxicon and its language-related modules. In *Proceedings of Companion Volume of the 6th Hellenic Conference on Artificial Intelligence (SETN)*, pages 27–32, 2010.
63. L. Schillingmann, B. Wrede, and K. J. Rohlfing. A computational model of acoustic packaging. *IEEE Transactions on Autonomous Mental Development*, 1(4), 2009.
64. S.C. Shapiro and H.O. Ismail. Anchoring in a grounded layered architecture with integrated reasoning. *Robotics and Autonomous Systems*, 43(2-3):97–108, 2003.
65. L. Steels and T. Belpaeme. Coordinating perceptually grounded categories through language: A case study for colour. *Behavioral and Brain Sciences*, 28(4):469–489, 2005.
66. F. Stramandinoli, A. Cangelosi, and D. Marocco. Towards the grounding of abstract words: A neural network model for cognitive robots. In *Proceedings of the International Joint Conference on Neural Networks*, pages 467–474, 2011.
67. F. Stramandinoli, D. Marocco, and A. Cangelosi. The grounding of higher order concepts in action and language: A cognitive robotics model. *Neural Networks*, 32:165–173, 2012.
68. S. Swarup, K. Lakkaraju, S.R. Ray, and L. Gasser. Symbol grounding through cumulative learning. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4211 LNAI:180–191, 2006.
69. M. Taddeo and L. Floridi. Solving the symbol grounding problem: A critical review of fifteen years of research. *Journal of Experimental and Theoretical Artificial Intelligence*, 17(4):419–445, 2005.
70. S. Tellex, T. Kollar, S. Dickerson, M.R. Walter, A.G. Banerjee, S. Teller, and N. Roy. Approaching the symbol grounding problem with probabilistic graphical models. *AI Magazine*, 32(4):64–76, 2011.
71. Moritz Tenorth and Michael Beetz. Towards practical and grounded knowledge representation systems for autonomous household robots. In *ICRA*, 2009.
72. M. Tomasello. *Constructing language: A usage-based theory of language acquisition*. Cambridge, 2003.
73. M. Vavrecka, I. Farkas, and L. Lhotska. Bio-inspired model of spatial cognition. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 7062 LNCS(PART 1):443–450, 2011.
74. P. Vogt. The physical symbol grounding problem. *Cognitive Systems Research*, 3(3):429–457, 2002.
75. P. Vogt. Anchoring of semiotic symbols. *Robotics and Autonomous Systems*, 43(2-3):109–120, 2003.
76. P. Vogt and E. Haasdijk. Modelling social learning of language and skills. *Artificial Life*, 16(4):289–309, 2010.
77. P.A.a Vogt and F.a b Divina. Social symbol grounding and language evolution. *Interaction Studies*, 8(1):31–52, 2007.
78. C. Yu and D.H. Ballard. On the integration of grounding language and learning objects. In *Proceedings of the National Conference on Artificial Intelligence*, pages 488–493, 2004.

Silvia Coradeschi is a professor at the AASS center at the School of Science and Technology of rebro University. Her main current research interests are in: Mobile robotics, Human Robot Interaction, Environmental monitoring Intelligent systems for medical applications, and Intelligent homes for elderly autonomous living. She has defined the concept of anchoring symbols to sensor data and has contributed with both theoretical and applicative studies to it. In particular she has formally defined anchoring and she has applied it in olfaction and in intelligent home envoronments.



Amy Loutfi is an associate professor at Örebro University, AASS. Her general interests include robotics and intelligent systems and more specifically: Machine Olfaction including Mobile Robot Olfaction, Knowledge Representation and Reasoning for Sensor Systems and Human-Robot Interaction and Social Robotic Telepresence. She has been working in anchoring sysmbols to sensors data both in the olfaction domain and in the domain of intelligent homes.

Since 2010 **Britta Wrede** is interim head of the Applied Informatics Group at Bielefeld University and since 2008 head of the research group Hybrid Society of the CoR-Lab. She has been working on human-robot dialog modeling, emotion recognition and modeling in HRI, developmentally inspired speech recognition approaches, visual attention modeling, the analysis of tutoring behavior towards children and robots and the modeling of the perception of multi-modal tutoring behavior for learning. She is Principal Investigator in several EU projects (ITALK, RobotDoc, Humavips) and national projects funded by DFG (CRC 673 Alignment in Communication; TATAS The Automatic Temporal Alignment of Speech), DLR (Sozirob The Robot as Fitness Coach) and BMBF (DESIRE Deutsche Service Robotik Initiative). Her research is driven by the question how to equip robots with a better understanding of their environment and is strongly inspired by human development. It follows the hypothesis that learning needs to be embedded in social interaction.