2010-09

Integration of Action and Language Knowledge: A Roadmap for Developmental Robotics

Cangelosi, A

http://hdl.handle.net/10026.1/3620

10.1109/TAMD.2010.2053034
IEEE TRANSACTIONS ON AUTONOMOUS MENTAL DEVELOPMENT
IEEE-INST ELECTRICAL ELECTRONICS ENGINEERS INC

Integration of Action and Language Knowledge: A Roadmap for Developmental Robotics

Journal:	*Transactions on Autonomous Mental Development*
Manuscript ID:	Draft
Manuscript Type:	Regular
Date Submitted by the Author:	
Complete List of Authors:	Cangelosi, Angelo; University of Plymouth, School of Computing and Mathematics Metta, Giorgio; Italian Institute of Technology, RBCS; University of Genoa, DIST Sagerer, Gerhard; Bielefeld University, CITEC Nolfi, Stefano; National Reserach Council, ISTC Nehaniv, Chrystopher; University of Hertfordshire, School of Computer Science Fischer, Kerstin; University of Southern Denmark, Department of Business Communication and Information Science Tani, Jun; RIKEN, BSI Sandini, Giulio; Italian Institute of Technology, RBCS Fadiga, Luciano; Italian Institute of Technology, RBCS Wrede, Britta; Bielefeld University, CITEC Rohlfing, Katharina; Bielefeld University, CITEC Tuci, Elio; National Reserach Council, ISTC Dautenhahn, Kerstin; University of Hertfordshire, School of Computer Science Saunders, Joe; University of Hertfordshire, Computer Science; University of Hertfordshire, School of Computer Science Zeschel, Arne; University of Southern Denmark, Department of Business Communication and Information Science
Keywords:	language acquisition through development, motor system and development

Integration of Action and Language Knowledge: A Roadmap for Developmental Robotics

Angelo Cangelosi, Giorgio Metta, Gerhard Sagerer, Stefano Nolfi, Chrystopher Nehaniv, Kerstin Fischer, Jun Tani, Giulio Sandini, Luciano Fadiga, Britta Wrede, Katharina Rohlfing, Elio Tuci, Kerstin Dautenhahn, Joe Saunders, Arne Zeschel

***Abstract*— This position paper proposes that the study of embodied cognitive agents, such as humanoid robots, can advance our understanding of the cognitive development of complex sensorimotor, linguistic and social learning skills. This in turn will benefit the design of cognitive robots capable of learning to handle and manipulate objects and tools autonomously, to cooperate and communicate with other robots and humans, and to adapt their abilities to changing internal, environmental, and social conditions. Four key areas of research challenges are discussed, specifically for the issues related to the understanding of: (i) how agents learn and represent compositional actions; (ii) how agents learn and represent compositional lexicons; (iii) the dynamics of social interaction and learning; and (iv) how compositional action and language representations are integrated to bootstrap the cognitive system. The review of specific issues and progress in these areas is then translated into a practical roadmap based on a series of milestones. These milestones provide a possible set of cognitive robotics goals and test-scenarios, thus acting as a research roadmap for future work on cognitive developmental robotics.**

***Index Terms*— Action learning, Humanoid robot, Language development, Social Learning, Roadmap**

I. INTRODUCTION

THIS paper proposes a developmental robotics approach to the investigation of action and language integration in embodied agents and a research roadmap for future work on the design of sensorimotor, social and linguistic capabilities in humanoid robots. The paper presents a vision of cognitive development in interactive robots that is strongly influenced by recent theoretical and empirical investigations of action and language processing within the fields of neuroscience, psychology, cognitive linguistics. Relying on such evidence on language and action integration in natural cognitive systems, and on the current state of the art in cognitive robotics, the paper identifies and analyses in detail the key research challenges on action learning, language development and social interaction, as well as the issue of how such capabilities are fully integrated. Although the primary target audience of the paper is the cognitive robotics community, as it provides a detailed roadmap for future robotics developments, the article is also relevant to readers from the empirical neural and cognitive sciences, as developmental robotics can serve as a modeling tool to validate theoretical hypothesis (Cangelosi and Parisi, 2002).

The vision proposed in this paper is that research on the integration of action and language knowledge in natural and artificial cognitive systems can benefit from a developmental cognitive robotics approach, as this permits the re-enactment of the gradual process of acquisition of cognitive skills and their integration into an interacting cognitive system. Developmental robotics, also known as epigenetic robotics, or autonomous mental development methodology, is a novel approach to the study of cognitive robots that takes direct inspiration from developmental mechanisms and phenomena studied in children (Lungarella et al. 2003; Cangelosi and Riga 2006; Weng et al. 2001). The methodologies for cognitive development in robots are used to overcome current limitations in robot design. To advance our understanding of cognitive development, this approach proposes the study of artificial embodied agents (e.g. either robots, or simulated robotic agents) able to acquire complex behavioral, cognitive, and linguistic/communicative skills through individual and social learning. Specifically, to investigate action/language integration, it is possible to design cognitive robotic agents

Manuscript received March 2, 2010. This work was supported by the EU Integrating Project “ITALK (214886) within the FP7 ICT programme “Cognitive Systems, Interaction and Robotics”.

A. Cangelosi and T. Belpaeme are with the Centre for Robotics and Neural System of the University of Plymouth, Drake Circus, Plymouth, PL4 8AA, UK (Tel.: +44-1752-586217; fax: +44-1752-586300; e-mail: {acangelosi; tbelpaeme}@plymouth.ac.uk). Contact author is Angelo Cangelosi.

G. Metta, G. Sandini and L. Fadiga are with the Italian Institute of Technology, Genoa, Italy. e-mail: {giorgio.metta;giulio.sandining; luciano.fadiga}@iit.it

G. Sagerer, B. Wrede and K. Rohlfing are with the Applied Computer Science Group at the University of Bielefeld, Germany; e-mail: {sagerer;bwrede;rohlfing}@techfak.uni-bielefeld.de

S. Nolfi and E. Tuci are with the Institute of Cognitive Science and Technology at the National Research Council, Rome, Italy; e-mail: {stefano.nolfi;elio.tuci}@istc.crn.it

C. L. Nehaniv, K. Dautenhahn and J. Saunders are with the Adaptive Systems Research Grout at the University of Hertfordshire, Hatfield, UK; e-mail: {c.l.nehaniv;k.dautenhahn;j.2.saunders}@herts.ac.uk

K. Fischer and A. Zeschel are with the Department of Business Communication and Information Science, University of Southern Denmark , Denmark; e-mail: {kerstin;zeschel}@sitkom.sdu.dk

Jun Tani is with the at Brains Science Institute of RIKEN, Japan; e-mail: tani@brain.riken.jp

capable of learning how to handle and manipulate objects and tools autonomously, to cooperate and communicate with other robots and humans, and to adapt their abilities to changing internal, environmental, and social conditions. The design of object manipulation and communication capabilities should be inspired by interdisciplinary empirical and theoretical investigations of linguistic and cognitive development in children and adults, as well as of experiments with humanoid robots. Such an approach is centered on one main theoretical hypothesis: action, interaction and language develop in parallel and have an impact on each other thus favoring the parallel development of action and social interaction permits the bootstrapping of cognitive development (e.g. Rizzolatti and Arbib 1998). This is possible through the integration and transfer of knowledge and cognitive processes involved in sensorimotor learning and the construction of action categories, imitation and other forms of social learning, the acquisition of grounded conceptual representations and the development of the grammatical structure of language. In addition to advancing our understanding of natural cognition, such a developmental approach towards the integration of action, conceptualization, social interaction and language can have fundamental technological implications for designing communication in robots and overcoming current limitations of natural language interfaces and human-robot communication systems.

This developmental robotics approach to action and language integration is also consistent with related brain-inspired approaches to mental development. For example, computational neuroscience approaches to cognitive development invoke the simultaneous consideration of neural development constraints and how these affect embodiment and cognition factors (Mareschal et al. 2007; Westermann et al. 2006; Weng and Hwang 2006; Weng 2007). For example, Sporns (2007) discusses in detail neurocomputational approaches to studying the role of neuromodulation and value system in developmental robotics.

In short, a complete, embodied cognitive system is needed in order to develop communication skills. The array of skills that are necessary to achieve this goal spans the range from sensorimotor coordination, manipulation, affordance learning to eventually social competencies like imitation, understanding of the goals of others, etc. Any smaller subset of these competencies is not sufficient to develop proper language/communication skills, and further, the development of language clearly bootstraps better motor and affordance learning and/or social learning. The fact that the agent communicates with others improves the acquisition of other skills. By interacting with others agents receive more structured input for learning (imagine a scenario of learning about the use of tools). Generalization across domains is also facilitated by the ability of associating symbolic structures such as those of language.

To follow such a vision, it is necessary to aim at the development of cognitive robotic agents endowed with the following abilities (see also Fig. 1):

- Agents learn to handle objects, individually and collaboratively, through the development of sensorimotor coordination skills and thereby to acquire complex object manipulation capabilities such as making artifacts (tools) and using them to act on other objects and the environment.
- Agents develop an ability to create and use embodied concepts. By embodied concepts we mean internal states grounded in sensory-motor experiences that identify crucial aspects of the environment or of the agent/environmental interaction. Such concepts mediate the agents' motor reactions and are used in communication with other agents. They can be organized in hierarchical representations, such as embodied semiotic schemata, used to plan interaction with the environment. Furthermore, embodied concepts can also be influenced through social and linguistic interaction.
- Agents develop social, behavioral and communicative skills through mechanisms of social learning such as imitation. Interacting with other agents enables the agents to share attention on a particular object or situation in order to cooperate, and to benefit from social adaptation of the partner in order to learn new skills and acquire embodied concepts.
- Agents develop linguistic abilities that allow them to represent situations and to communicate complex meaning via language. They learn relationships between sounds, actions and entities in the world. These relations will facilitate the discovery of word meaning and are a precursor to grammatical comprehension and production. More advanced communication skills develop based on the combination of previously-developed embodied concepts and the development of symbolic and syntactic structures.
- Agents are able to integrate and transfer knowledge acquired from different cognitive domains (perception, action, conceptual and social representations) to support the development of linguistic communication. The co-development, transfer, and integration of knowledge between domains will permit the bootstrapping of the agent's cognitive system.

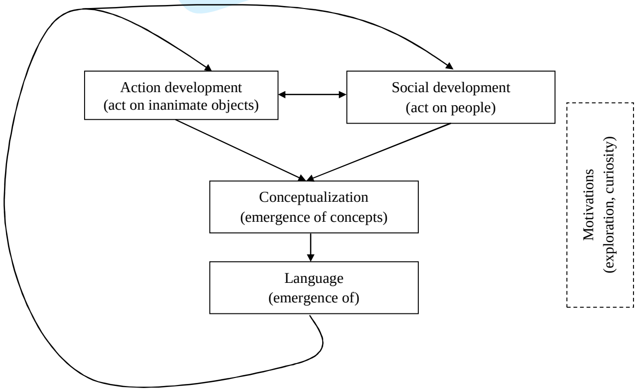

Fig. 1. Connections between the various skills of a developmental cognitive agent. The focus on this paper will be on the aspects more closely related to language and action development (boxes with continuous lines). The diagram also acknowledges the additional contribution of other capabilities related to motivation and affective behavior (dotted box), though they will not be part

of the core discussion in this paper.

Research on the further understanding and design of the above cognitive abilities in natural (children and adults) and artificial (robots) cognitive agents can be centered around four key challenges:

(1) Understanding how agents learn and represent compositional *actions*
(2) Understanding how agents learn and represent compositional *lexicons*
(3) Understanding dynamics of *social* interaction and learning
(4) Understanding how compositional action and language representations are *integrated*

In the following section (section 2) we first provide a brief overview of the state of the art in experimental disciplines investigating embodied cognition and action/language processing in natural cognitive systems (humans and animals) and the state of the art in artificial cognitive systems (robots) models of language learning. This evidence on action language integration has important implications for the design of communication and linguistic capabilities in cognitive systems and robots (Cangelosi et al. 2005, 2008) to progress beyond the state of the art. Sections 3-6 will analyze in detail the specific issues on the four sets of key challenges respectively for action, language, and social learning and for cognitive integration. Additional review of literature on the specific theoretical and empirical work on action, language and social learning will be included within the key challenge sections 3-7. This will further support specific claims and proposals for future developmental robotics investigations in the field. The paper then concludes with the presentation of the research roadmap and a description of key milestones.

II. RELATION TO THE STATE OF THE ART

A. Action and Language Processing in Natural Cognitive Systems

Recent theoretical and experimental research on action and language processing in humans and animals clearly demonstrates the strict interaction and co-dependence between language and action (e.g. Cappa and Perani, 2003; Glenberg and Kaschak, 2002; Pulvermuller et al. 2003; Rizzolatti and Arbib, 1998). In neuroscience, neurophysiology investigations of the mirror neurons system (Fadiga et al., 2000; Gallese et al, 1996) and brain imaging studies on language processing provide an abundance of evidence for intertwined language-action integration. For example, Hauk et al. (2004) used fMRI to show that action words referring to face, arm or leg actions (e.g. to lick, pick, or kick) differentially activate areas along the motor cortex that either were directly adjacent to or overlapped with areas activated by actual movement of the tongue, fingers, or feet. This demonstrates that the referential meaning of action words has a correlate in the somatotopic activation of the motor and premotor cortex. Cappa and Perani (2003) review neuroscience evidence on neural correlates of nouns and verbs. They found a general agreement on the fact that the left temporal neocortex plays a crucial role in lexical-semantic tasks related to the processing of nouns whereas the processing of words related to actions (verbs) involves additional regions of the left dorsolateral prefrontal cortex. Overall, neuroscientific evidence supports a dynamic view of language according to which lexical and grammatical structures of language are processed by distributed neuronal assemblies with cortical topographies that reflect lexical semantics (Pulvermuller 2003). The mastery of fine motor control, such as non-repetitive action sequences involved in making complex tools, is also seen as an ability related to the precursor of Broca's area in the modern brain, which is adjacent to the area that governs fine motor control in the hand. This is consistent with Rizzolatti and Arbib's (1998) hypothesis that area F5 of the monkey's brain, where mirror neurons for manual motor activity have been identified, is homologous to a precursor of Broca's area involved in language processing and speech production and comprehension.

This neuroscience evidence is consistent with growing experimental and theoretical evidence on the role of grounding of language in action and perception (Pecher and Zwaan, 2005; Glenberg and Kashack 2002; Barsalou 1999). Glenberg proposed that the meaning of a sentence is constructed by indexing words or phrases to real objects or perceptual analog symbols for those objects, deriving affordances from the objects and symbols and then meshing the affordances under the guidance of syntax. The direct grounding of language in action knowledge has been recently linked to the mirror neuron system (Glenberg and Gallese, in press). Barsalou (1999) places similar emphasis on perceptual representation for objects and words in his "Perceptual Symbol Systems" account of cognition. For Barsalou, words are associated with schematic memories extracted from perceptual states which become integrated through mental simulators.

Developmental psychology studies based on emergentist and constructivist approaches (e.g. Bowerman and Levinson, 2001; MacWhinney, 2005; Tomasello, 2003) also support a view of cognitive development strongly dependent on the contribution of various cognitive capabilities. They demonstrate the gradual emergence of linguistic constructs built through the child's experience with her social and physical environment. This is consistent with cognitive linguistics approaches (cf. Lakoff, 1987; Langacker, 1987) where syntactic structures and functions, that is, symbolic structures in both lexicon and grammar, are constructed in reference to other cognitive representations.

Another area at the intersection between developmental psychology and cognitive neuroscience that is relevant to cognitive and linguistic development is neuroconstructivism (Sirois et al. 2008; Westermann et al. 2007; Quartz and Sejnowski 1997). This theoretical and experimental framework puts a strong focus on the role of embodiment and brain co-development during cognitive development. It considers the constraints that operate on the development of neural structures that support mental representations and explains cognitive development as a trajectory emerging from the interplay of these constraints. This brain-inspired approach has also been supported by computational models, that have the potential to offer explanations of the interactions between brain and cognitive development (Mareschal et al. 2007; Westermann et al. 2006).

All these studies on action-language integration have important implications for the design of communication and linguistic capabilities in cognitive systems and robots (Cangelosi et al. 2005, 2008). Amongst the various approaches to design communication capabilities in interactive agents, some provide a more integrative vision of language and treat it as an integral part of the whole cognitive system (Cangelosi and Harnad 2000). The agent's linguistic abilities are strictly dependent on, and grounded in, other behaviors and skills. Such a strict action-language interaction supports the bootstrapping of the agent's cognitive system, e.g. through the transfer of properties of action knowledge to that of linguistic representations (and vice versa).

B. Action and Language Learning in Robots

Recent models from cognitive robotics research have addressed some of the issues described above, and contributed to the identification of the open research challenges in language and action research. Before we discuss in detail the key challenges, we review a few of the most interesting contributions.

Deb Roy (2005; Roy et al. 2004) propose the use of conversational robots able to translate complex spoken commands such us "hand me the blue one on your right" into situated actions. These robots are provided with a control architecture that includes a three-dimensional model of the environment (which is updated by the robot on the basis of linguistic, visual, or haptic input) and sensory-motor control programs. This model is consistent with the notion of schemas proposed by Piaget (1954), in which the meaning of words is associated with both perceptual features and motor program. For example, the word 'red' is grounded in the motor program for directing active gaze towards red objects. Similarly, the word 'heavy' is grounded in haptic expectations associated with lifting actions. Objects are represented as bundles of properties tied to a particular location along with encodings of motor affordances for affecting the future location of the bundle.

Dominey, Mallet and Yoshida (2009) designed robotic experiments with robots that, in addition to reacting to language commands issued by the user (which trigger predesigned control programs), are able to acquire on the fly the meaning of new linguistic instructions, as well as new behavioral skills, by grounding the new commands in combinations of pre-existing motor skills. This is achieved during experimental sessions in which the human user and a robot try to cooperatively achieve a shared goal. During these sessions the interaction between the human user and the robot is mediated by two types of linguistic information: (i) linguistic commands (e.g. "open right-hand", "take object-x", "give-me object-y", etc) that trigger contextually independent or dependent behaviors, and (ii) 'meta' commands (e.g. "learn macro-x", "ok", "wait") that structure what the robot is to learn or regulate the human-robot interaction. In another experiment, Dominey and Warneken (2009) designed robots able to cooperate with a human user by sharing intentions with her in a restricted experimental setting. This is achieved by allowing the robot to observe the goal-directed behavior exhibited by a human and then to adopt the plan demonstrated by the user. The robot thus shows both an ability to determine and recognize the intentions of other agents, and an ability to share intentions with the human user. These two skills are at the basis of social learning and imitation in humans, as proposed by Tomasello et al. (2005). These abilities have been realized by providing the robot with a model of the environment, the possibility to represent intentional plans constituted by sequences of actions producing specific effects, and the ability to recognize actions and to attribute them to the robot itself or to a human agent.

Weng (2004) designed a developmental learning architecture that allows a robot to progressively expand its behavioral repertoire while interacting with a human trainer that shapes its behavior. Different learning methods are used, including learning by demonstration (in which the robot learns while the trainer drives the robot's actuators), reinforcement learning (in which the robot learns through a form of trial and error process guided by the positive or negative feedback provided by the trainer), and language learning (in which the robot learns to associate the current sensory states to the action triggered by the trainer through language commands, and also learns to anticipate the next sensations and actions). The approach proposed by Weng is inspired by animal learning, neuroscience evidence, and cognitive science models, aiming to be general enough to be task independent (i.e. to allow the robot to learn any type of task through the same learning methods). This architecture has been successfully implemented, for example, in an humanoid robot that first learns to associate four language commands to four corresponding context-independent behaviors, then learns to associate a fifth language command to a composite action consisting of the execution of the four behaviors acquired previously in sequence (thanks to the mediation of the user that trains the robot by producing the four corresponding language commands after the fifth command), and (eventually) to be able to extinct one of the previously acquired reactions to language commands as a result of negative feedbacks provided by the user (Zhang and Weng, 2007).

Sugita and Tani (2005) developed a model in which a robot acquires the ability to both translate a linguistic command into context-dependent behaviors, and an ability to map sequences of sensory-motor state experienced while producing a given behavior into the corresponding verbal descriptions. More specifically a wheeled robot, provided with a 2DOF arm and a CTRNN controller, is trained through a learning by demonstration method to carry out behavioral and linguistic tasks that consist respectively in: (i) interacting with the three objects presented in its environment through the execution of three different types of behaviors such as "indicate object-x", "touch object-x", and "push object-x", and (ii) processing the corresponding language commands such as predicting the next word forming the corresponding sentence. The two tasks are carried out by two different modules of the neural controller. However these modules co-influence each other through some shared neurons (called parametric bias) that are forced to assume similar states during the execution of the two related tasks. At the end of the training process the robot shows an ability to translate the language commands into the corresponding situated actions as well as an ability to generate the right language output when the robot is forced to produce a given behavior. The fact that the robot reacts appropriately to sentences never experienced during the training process, moreover, demonstrates how it is able to represent the meaning of words and the corresponding behavior in a compositional manner.

Steels, Kaplan and Oudeyer have studied the acquisition of language in both developmental contexts (Steels and Kaplan 2000; Oudeyer and Kaplan 2006) and evolutionary scenarios (Steels 2005b). For example, Oudeyer and Kaplan (2006) investigated the hypothesis that children discover communication as a result of exploring and playing with their environment using a pet robot (Sony AIBO robot) scenario. As a consequence of its own intrinsic motivation, the robot explores this environment by focusing first on non-communicative activities and then discovering the learning potential of certain types of interactive behavior. This motivational capability results in robots acquiring communication skills through vocal interactions without having a specific drive for communication.

The following sections will discuss in detail the key research challenges for cognitive robotics models of action and language integration, also referring to additional literature work addressing the specific research issues.

III. KEY CHALLENGE 1: LEARNING AND REPRESENTATION OF COMPOSITIONAL ACTIONS

The investigation of grasp-related functions in the brain and the successive discovery of the mirror neurons system have changed the perception of the importance of manipulation and its relationship to speech (Rizzolatti and Arbib 1998). Although, the mirror neuron system is the quintessential example of this changed understanding of the neurophysiology of action, the study of the control of action in its entirety revealed modularity and compositionality as key elements of flexible and adaptable behavior generation (Mussa-Ivaldi and Giszter 1992; Mussa-Ivaldi and Bizzi 2000; Rizzolatti et al. 1997; Graziano et al. 1997). The important point here is that areas of the brain that were considered as mere sensorimotor transformation circuits (i.e. changing coordinates or frame of reference) revealed a deeper structure with peculiar characteristics. This deeper structure includes multisensory neurons (e.g. visuo-motor in F5, visuo-haptic-proprioceptive in F4), generalization (the same neuron fires irrespective of the effector used), and compositionality (different areas specialize to different goals –reaching, grasping, etc.– rather than just reflecting a generic somatotopy. This is not a single homunculus, but rather multiple representations of the body with respect to the different action goals. Modularity was discovered in the cerebral cortex but also down to the spinal cord. In a recent experiment (Borroni et al. 2005) the so-called "motor resonance" effect has been demonstrated using the H-reflex technique of the peripheral nerves and transcranial magnetic stimulation (TMS). Additional experiments, such as those in Sakata et al. (1995) showed a link between the "shape" of objects and the actions that can successfully manipulate these objects. Further Gallese et al. (1996) observed neurons in the premotor cortex (area F5) which fire selectively for certain combinations of grasp type and object shape (F5 canonical neurons). It seems that the brain stores a "vocabulary" of actions that can be applied to objects and the mere fixation of a given object activates potential motor acts even if, the monkey in this case, did not move.

This new evidence generated a surge of interest including the cognitive sciences on one side and, the robotics community on the other (see Clark 2001 for a summary). Concepts like that of Gibsonian affordances started to be considered and modeled in robotics (Metta and Fitzpatrick 2003) and the links between imitation and manipulation were explored (Simmons and Demiris 2006; Metta et al. 2006). In this respect, the link between internal models, prediction, and the activation of a mirror-like system was approached in many different ways by using most disparate models (Oztop et al. 2006, Ito et al. 2006, to name a few). Clearly, this effort is even more relevant given the special relationship between mirror neurons, manipulation and language (Fadiga et al. 2002). In the experiment by Fadiga and colleagues (2002), it was possible to measure motor effects when listening to words of different categories in strict congruence with the muscular activation required to pronounce the same set of words, which provides evidence for the presence of a speech-mirror system in humans akin to the grasp mirror system of the monkey. A more recent experiment confirms these findings and enters into the details of the motor resonance effect depending on the phonology versus the frequency of words (Roy et al. 2008). The results indicate that rare words require a stronger activation of the premotor cortex as if the increased difficulty of the task requires reliance on the premotor activation and, conversely, common words are recognized because of a consolidated and larger number of cues which lower the premotor cortex activation.

Further, evidence has accumulated demonstrating the pervasiveness of this principle in several domains, including reaching (e.g. Graziano et al. 1997; Fogassi et al. 1996), attention (Craighero et al. 1999), and motor imagery

(Jeannerod 1997) to name a few. It remains to be considered that none of these skills is innate, but rather they develop through experience and in many cases require several years before reaching maturity (von Hofsten 2004). Aspects like prediction (prospective behavior) and explorative and social motives have to be considered in motor learning since they seem to be crucial also for the engineering of adaptive systems in any meaningful sense. In this respect, it seems that newborns are sensitive to their own and other's motor movements and use these to assess social cues. For example, motion during eye gaze and human facial expressions are used in judging social interaction (Moore et al. 1997; Farroni et al. 2004). Children use these early sensory commodities to bootstrap cognitive development, which includes motor skills. They subsequently go through an extensive period of exploration and development guided by various motivations (including the motivation of exercising the motor system, known as "motor babbling"). This leads to the acquisition of several motor skills like the ability of directing gaze, of coordinating head and eye movements, of coordinating gaze and attention together with reaching and eventually of manipulating the external world via grasping (von Hofsten 2004).

In the light of these results, modular motor control for articulation is a prerequisite for speech in humans, and it can be certainly considered as a prerequisite for speech also in artificial systems. This follows in some sense the approach of Liberman and Mattingly (1985) who first formulated the so called "motor theory of speech perception", which was exactly proposed because of the difficulty of performing artificial speech recognition (ASR) entirely on acoustic analysis. Motor activation and sensory processing seem to be deeply intertwined in the brain (not only in the premotor cortex). Conversely, in robotics, it was possible to demonstrate an improvement due to learning in multisensory (sensorimotor) environments (Metta et al. 2006; Hinton and Nair 2006). Manipulation plays a pivotal role in this picture, sharing a similar "grammatical/hierarchical" structure with language but also owing to the close homology between F5 in the monkey and Broca's in humans (Rizzolatti and Arbib 1998).

The next sections will highlight and discuss some of the main open research issues in action learning that are highly relevant to future cognitive robotics research. Specifically, the focus will be on (i) the properties of generalization and compositionality in action development, (ii) the issues of recursive and (iii) hierarchical motor representations, (iv) the issues in embodied concept representation and (v) the mental representation of concepts during development. These research issues will then be used to identify specific milestones on action learning in the roadmap.

A. *Generalization and Compositionality*

The development of complex action and manipulation capabilities constitute the foundation for the synchronous development of motor, social and linguistic skills. For this it is fundamental to identify the characteristics of action development that are compatible with this scenario and reject those that are mere engineering shortcuts. In particular, two core properties of biological motor control systems are considered: compositionality and generalization.

Compositionality refers to the ability of exploiting the combinatorial explosion of possible actions for creating a space of expressive possibilities that grows exponentially with the number of motor primitives. The human motor system is known to be hierarchically organized (with primitives implemented as low as at the spinal cord level) and it is simultaneously adaptive in recombining the basic primitives into solutions to novel tasks (via sequencing, summation, etc.). The hierarchy is implemented in the brain by exploiting muscle synergies as well as parallel controllers reaching different degrees of sophistication apt to either address the global aspects of a motor task or the fine control required for the use of tools (Rizzolatti and Luppino 2001).

The aspect of generalization is equally crucial. It refers, in this context, to the ability of acquiring (read learning) motor tasks by various means, using any of the body effectors, and even via imagination of the motor execution itself (as for example in Jeannerod 1997). Naïvely, one could assume a common representational framework defined in some task independent system of coordinates. However, at the same time, neuroscience seems to be indicating that representation is effector-dependent (Fogassi et al. 1996). This is clearly a question that needs to be addressed with links to many different aspects of the representation of linguistic constructs (e.g. actions vs. the description of actions).

In artificial systems, this translates into the realization of a modular controller which, on the one hand, combines a limited set of motor primitives in realizing global control strategies, and on the other, learns to finely move single degrees of freedom to affect particular complex motor mappings (similar to what happens in the brain between the control effected by the premotor cortex versus that generated by the primary motor cortex). Simultaneously, the adaptation and estimation of bodily parameters must be considered both on the developmental and on the single task/session timescale. It is then particularly important that artificial systems show these properties if their motor controller has to form a suitable basis for further development in more higher-order cognitive scenarios such as language.

One interesting topic of research concerns the selection of a generic endpoint for subsequent actions (motor invariance) and fast adaptation to disturbances (changes in dynamics, weight, etc.). One example of flexibility in humans is the possibility of dynamically select the end point for subsequent tasks and reducing/increasing the number of degrees of freedom employed given the precision, noise, and other parameters required (e.g. imagine how humans reduce the number of degrees of freedom by laying objects on a table when precision is required such as in inserting a thread into a needle). This flexibility in choosing the effector to use seems fundamental to adaptability and relates to the existence of a peripersonal sensorimotor space (Fogassi et al. 1996). Another example of flexibility in humans is in adapting to added perturbations (e.g. increased weight or changed dynamics). In the latter case, the motor system adapts after a few dozen trials and does it by estimating and modeling the change of dynamics maintaining a very energetically efficient control strategy (for example see Lackner and DiZio 1998).

B. Recursive and Hierarchical Primitives

As previously pointed out, motor and linguistic skills share a relevant structure. Specifically, the modular organization of biological motor systems has been shown to be based on hierarchical recursive structures which have linguistic analogues in grammatical/syntactical structures.

Primitives have been identified in the spinal cord of frogs and rats, thus revealing that a modular structure exists at the movement execution level (the lowest level in the motor hierarchical structure). Interestingly these modules have very simple combinatorial rules (linear superposition) which have led to interesting applications (Wolpert and Kawato 1998).

Higher hierarchical structures seem to play a crucial role in movement planning while still preserving a substantial modularity. As to this concern, there is evidence for the existence of individual cortical substructures which code increasingly higher movement related abstractions. There is evidence supporting the existence of structures coding (1) hand kinematics (Georgopoulos et al. 1982), (2) specific action goal, timing and execution (Rizzolatti et al. 1988), (3) movement sequencing (Carpenter et al. 1999), (4) virtual action descriptions (i.e. actions which do not have a concrete goal yet) (Nakayama et al. 2008) (5) object affordance in terms of correspondences between object and motor prototypes (Murata et al. 1997) and (6) movement recognition (Gallese et al. 1996) (Rizzolatti et al. 1996).

At present, the rules governing the combination of different action executions have been widely studied and have been successfully applied in the area of motor control. Conversely, the rules governing the combination of goals in action planning appear to be more complex and not yet completely understood. Remarkably, these rules seem to be fundamental in order to fully exploit the properties of compositionality and generalization embedded in a modular architecture. Moreover, the “definition” (here to be understood as “development”) of suitable compositional rules appears to be an ideal candidate for providing theoretical insights into the integration of action, social and linguistic skills

C. Hierarchical Learning

The observation that the brain uses hierarchical organizations in various sensory and motor systems has inspired the development of similarly organized artificial systems. Essentially, two different approaches have been followed within this context: a bottom-up approach which falls within the mathematical framework of function approximation and a top-down approach based on the properties of the motor output.

As to bottom up approaches, one of the first to mention is LeNet, which uses a convolution network with multiple layers for handwritten digit recognition (LeCun et al. 1990). More recently, Serre et al. (2007) have developed a computational model of the lower levels of the visual cortex. This model alternates levels of template matching and maximum pooling operations, similar to the role of simple and complex cells as found in the visual cortex (Hubel and Wiesel 1962). This model has shown excellent performance on immediate recognition benchmark problems, whereas extensions have been used for action recognition (Jhuang et al. 2007) and facial expression recognition (Meyers and Wolf 2008). The underlying principle of these systems is to gradually increase both the selectivity of neurons to stimuli along with their invariance to (2D) transformations in a series of processing levels (Giese and Poggio 2003). Further, the receptive field of the neurons increases along the hierarchy. In effect, these hierarchies serve to extract relevant features from the data stream and to combine these in compact, high level representations.

Besides having a biological foundation, hierarchical architectures are also believed to have computational advantages over single layered architectures. Hierarchical architectures trade breadth for depth and can theoretically achieve a logarithmic decrease in the number of neurons needed to learn certain tasks (Bengio and LeCun 2007, Mnih and Hinton 2009). However, hierarchical architectures are notoriously hard to train and may therefore not reach up to their full potential. Hinton et al. proposed a novel learning method for deep belief networks, which is a variant of a multi-layered neural network, to address this problem (Hinton et al. 2006). In this method each layer is trained separately to output a compact and sparse representation of its input distribution. Only the most relevant aspects of the input distribution remain at the top level, therefore facilitating generalization. If used in the opposite direction, i.e. from output to input, then each layer will attempt to reconstruct the original input from the compact output representation. An interesting direction for novel research is to apply these hierarchical learning methods for motor control.

In contrast to bottom up approaches, top down approaches are based on the input/output properties of the motor system. As to this concern, one of the most interesting theoretical results has been proposed by D. M. Wolpert in the framework of multiple paired forward and inverse models (Wolpert and Kawato, 1998). By devising a modular structure which has strong similarities with the modularity present in the cerebellum, it was proposed that multiple forward and inverse models can be simultaneously learnt in order to approximate complex sensory motor mappings (module learning problem). Interestingly it was observed that the problem of choosing the correct subset of inverse models to handle the current context (module selection problem) can initially be solved by exploiting forward model predictions. Simultaneously, these predictions can be used to train suitable responsibility predictors which can be used later to solve the selection problem by exploiting contextual cues only.

New research in cognitive robotics should focus on the acquisition of hierarchical and compositional actions. Typical experimental scenarios might involve robotic agents that use proprioceptive and visual information to actively explore the environment. This will allow agents to build embodied sensorimotor categories of object-body interactions. Actually, such trials have been demonstrated in (Yamashita and Tani 2008). It was shown that a humanoid robot can learn to generate object manipulation behaviors in a compositional way by self-organizing functional hierarchy by which the lower level primitives such as touch/lift/move objects are sequentially combined in the higher level by utilizing inherent time constant differences in the employed dynamic neural network model. However, the experiment was limited in its scalability and lacked developmental aspects. New studies should include more advanced experiments to look at developmental processes of acquiring manipulation action patterns based on combination and sequences of movements. For example, new robotics experiment might start from situations in which robot agent learns to use a tool (e.g. “stick”) to push an object. Other tasks might include a cascade of inter-dependent actions, such as making a composite tool (e.g. combine a stick with a cuboid object – as with the handle and head of a “hammer”) and using this tool on a third object (e.g. to crack open a spherical object – “nut”). Tasks can be inspired by object manipulation and tool making/use observed abilities in primates and humanoids, and their relationship with the development of linguistic capabilities (e.g. Corballis 2002; Greenfield 1991). A possible starting point could be to attempt object manipulation in order to get an agent to relate one object with another in a particular combination, as a young infant would (Tanaka and Tanaka 1982). In conjunction with the research undertaken by Hayashi and Matsuzawa (2003) on the development of spontaneous object manipulation in apes and children, language experiments can focus on the following tasks: (i) Inserting objects into corresponding holes in a box; (ii) Serializing nested cups; (iii) Inserting variously shaped objects into corresponding holes; (iv) Stacking up wooden blocks. A first instance of the experiments could be able to isolate the agent from the human, so as to let it calibrate its joints and hand-eye coordination, recognizing color, form/shapes and moving objects. The second part would be to introduce the agent to a “face to face” situation where a user would use linguistic instructions in order to expand the object “knowledge acquisition”, taking the form of some kind of symbolic play.

D. *Embodied Learning of Representation and Concepts*

A fundamental skill of any cognitive system is the ability to produce a variety of behaviors and to display the behavior that is appropriate to the current individual, social, cultural and environmental circumstances. This will require agents: (1) to reason about past, present and future events, (2) to mediate their motor actions based on this reasoning process and (3) to communicate using a communication system that shares properties with natural language. In order to do this, robots will need to develop and maintain internal categorical states, i.e. ways to store and classify sensory-motor information. To properly interact with the objects and entities in the environment, agents should possess a categorical perception ability which allows them to transform continuous signals perceived by sensory organs into internal states or internal dynamics in which members of the same category resemble one another more than they resemble members of other categories (Harnad 1990). These internal states can be called “embodied concepts” and can be considered as representations grounded in sensory-motor experiences that identify crucial aspects of the environment and/or of the agent/environmental interaction.

In the literature there are two orthogonal approaches to representing concepts in artificial systems: one commonly known as the symbolic approach, the other as the subsymbolic approach. In the symbolic approach, conceptual information is represented as a symbolic expression containing recursive expressions and logical connectors, while in the subsymbolic approach concepts are represented in a continuous domain, for example in connectionist networks or semantic spaces (cf. Gärdenfors, 2000). Both approaches serve their purpose, but none seems to resonate well with human conceptualization. Humans use symbolic knowledge in representations for communication and reasoning (Deacon, 1997), but these symbols are implemented on a neural substrate, which is non-symbolic and imprecise. There have been few attempts to reconcile both, and new research should focus at the design of a conceptual representation which has the precision of logic symbols, but the plasticity of human concepts. This representation should also support the acquisition of concepts through embodied sensorimotor interactions.

Embodied concepts can be immediately related to sensory or motor experiences, such as motor action concepts or visual shape/object concepts, in which case we call them perceptual concepts. On the other hand, concepts can also be indirectly related to perceptual input, in which case we call them abstract embodied concepts (e.g. Wiemer-Hastings and Xu 2005; Barsalou 1999). These concepts are typically hierarchical constructs based on other abstract concepts and perceptual concepts. Categories, in our approach, will be based on commonalities and structure of concepts that exists among items (cf. Rakison and Oakes 2003).

In line with a dynamical system view of cognitive development (Thelen and Smith, 1994), embodied concepts should be conceived at the same time as pre-requisites for the development of behavioral, social, and communicative skill and as the result of the development and co-development of such skills. In this respect, the development of embodied concepts might play the role of a scaffold which enables the development of progressively more complex skills.

An important challenge for cognitive robotics thus consists in identifying how embodied agents can develop and progressively transform their embodied concepts autonomously while they interact directly with the physical and social environment (without human intervention) and while they attempt to develop the requested behavioral skills. This objective can be achieved through experiments studying different aspects of categorization and concept formation, with the goal of progressively integrating into a single setup categorization aspects previously studied in isolation. These experiments require that the robot is left completely free to determine how they interact with the environment in order to perform the categorization task. For example, a robot placed in front of objects (one at a time) varying with respect to their shape, size, and orientation will be trained for the ability to categorize the shape of the object by producing different labels for objects with different shapes. The robot will be rewarded on the basis of its ability to label the shape of the object and will not be asked to produce any specific behavior (i.e. it will be left free to determine how to interact with the objects).

The goal of this research methodology is twofold. On one side, these experiments can pose the basis for the investigation of more complex experimental scenarios in which the development of an ability to linguistically categorize selected features of the environment will be integrated with the development of an ability to display certain behavioral and social skills. On the other side, these experimental scenarios can be used to study the role of active categorical perception and the role of the integration of sensory-motor information over time.

Active categorical perception refers to the fact that in agents which are embodied and situated, the stimuli which are sensed do not depends only on the structure of the environment but also on the agents' motor behavior. This implies that categorization is an active process that requires: (a) the exhibition of a behavior which allows the agents to experience the stimuli that provide the necessary regularities to perceptually categorize the current agent/environmental state, and (b) the development of an ability to internally elaborate the experienced sensory states. The ability to coordinate the sensory and motor process, however, does not only represent a necessity but also an opportunity, since the possibility to alter the experienced sensory stimuli might significantly simplify the perceptual categorization process or might lead to the generation of the regularities that are necessary to perceptually categorize functionally different agent/environmental situation. The goal of this set of experiments, therefore, will be that to identify how such possibility can be exploited, Although pioneering research in this area has provided important theoretical contributions (Chiel and Beer 1997; Scheier et al. 1998; Pfeifer and Scheier 1999; Nolfi and Floreano 2000; O'Regan and Noë 2001; Keijzer, 2001) as well as few preliminary demonstrations of how artificial embodied agents can develop active categorization skills (Nolfi and Marocco 2002; Beer 2003; Nolfi 2005), some themes still deserve substantial further investigations. In particular, open questions concern: (i) the identification of the modalities with which action can facilitate or enable categorical perception, (ii) the identification of how internal categories can be represented, (iii) the identification of the adaptive mechanisms which can lead to the development of two interdependent skills (the ability to act so to favor categorical perception and the ability to categorize perceived sensory-motor information codetermined by agents' motor behavior).

Another important focus of future research on embodied concept learning and representation regards the development of abstract perceptual categories based on regularities distributed over time. The regularities that can be used to categorize functionally different agent/environmental circumstances are not necessarily available within a single sensory pattern and often require an ability to integrate sensory-motor information through time. Consider for example the problem of grasping objects of different shapes on the basis of tactile information or the problem visually recognizing an object by visually exploring it through eye movements. To functionally categorize the nature of these agent/environmental situations, the agent should take into account aspects such as the duration of an event or the sequence with which different events occur. This problem is further complicated by the fact that regularities that should be integrated over time might be distributed at different time scales (e.g. ranging from milliseconds, to seconds or minutes). Recent research in this area has demonstrated how robotic agents can successfully develop categorization abilities and abstract perceptual categories provided that certain pre-requisites are met (Wolpert and Kawato 1998; Nolfi and Tani 1999; Tani and Nolfi 1999; Beer 2003; Sugita and Tani, 2005; Ito et al. 2006; Gigliotta and Nolfi 2008; Yamashita and Tani, 2008). These studies also provide useful hints which might help us to identify the characteristics of the developmental process and of the robots which represent a pre-requisite for the ability to develop abstract concepts. However, whether and how these models can be scaled to more complex scenarios remains an open question which deserves further investigations.

E. Social Learning of Concepts

In order to understand how humans represent knowledge, much can be learned from studying how infants and young children acquire concepts. There are many experimental studies and theories on concept acquisition in young children (Rakison and Oakes, 2003). Children, for example, employ a number of strategies to facilitate concept acquisition, such as mutual exclusivity, where a word is only related to one object in a context and not to others (Markman, 1989), or the preference to bind unfamiliar words with unfamiliar perceptual input: the novel name novel category principle (Mervis and Bertrand, 1994). Also, language seems to play a crucial role in concept acquisition. Although linguistic relativism —the interaction between language and thought— used to be controversial, recent studies have convincingly shown that language and conceptualization do interact in a number of different domains, such as time, space and color (for example (Boroditsky 2001; Gilbert et al., 2006; Gumperz and

Levinson, 1997; Roberson et al., 2005; Winawer et al., 2007), but see Pinker (2007) for a critical note. Although the evidence for the interaction between language and concepts is convincing, it is only recently that the importance of language for the acquisition of concepts has been noted. Choi et al. (1999), for example, show how young children (18-23 months) are already sensitive to linguistic concepts for space (see also Majid et al., 2004). This does not tell whether children actively use language to acquire concepts. However, Xu (2002) shows how 9-month olds use of language can play an important role in learning object concepts and more recently, Plunkett, Hu and Cohen (2008) show how linguistic labels play a causal role in concept learning of 10-month olds.

In the tightly controlled experimental settings of above mentioned psychological studies, children are exposed to unidirectional communication: objects and linguistic labels are presented to the infants and they induce concepts from these experiences. These experimental conditions however do not reflect reality, where children and caretakers engage in a rich interaction with joint attention, referential and indexical pointing, and implicit and explicit feedback. It is expected that rich, cultural interaction is essential to cognition (Tomasello, 1999). New research should explore the influence of rich interaction on the mental development of robots. It has been argued and, to a certain extent, it has been experimentally shown that this tight interaction is bi-modal, involving both language and action and that this occurs from an early age. Locke (2007) reports how 16.5-month old infants significantly join vocalizations and referential points, which would suggest an integrated system.

Concerning the mental representation of categories and concepts, it is important to first distinguish between categories and concepts. For the pragmatic purposes of developmental robotics and cognitive systems, categories are seen as directly related to perceptual experiences and concepts as higher-level representations, based on categories, but possibly also deduced from contextual information without necessarily being related to perceptually grounded categories. Categorization in artificial intelligence and by extension in recent cognitive systems work has often been considered to be a supervised learning task (e.g. Ponce, 2006), whereby pairs of stimuli (often images) and labels are offered to a learning algorithm. In recent years progress has been made in the representation of images, using either local or global features, and in the learning algorithms. However, nearly all focus on passive learning of categories and concepts from annotated data (cf. however (Oudeyer, 2006)). Future research in developmental robotics could explore active learning, in which the learner (in this case the robot or cognitive system) engages in a dyad with its caretaker and actively invites the caretaker to offer it learning experiences while at the same time using the caretaker to refine categorical and conceptual knowledge. This is an extension of classical symbol grounding (see Harnad, 1990). Instead of meaning only being defined in perception of objects in the environment, social and cultural interaction has an equally important influence on meaning. This is known as extended symbol grounding (Belpaeme and Cowley, 2007). The cultural acquisition of categories has been explored in simulation and robotic environments (see for example Steels, 2006; Vogt, 2003) and close parallels have been noted between simulated cultural learning of words and categories and human category acquisition (Belpaeme and Bleys, 2005; Steels and Belpaeme, 2005). However, while extended symbol grounding has not been explored in environments involving both humans and robots (although see Roy, 2005b; Seabra-Lopes and Chauhan, 2007), this offers an exciting opportunity for cognitive systems research, with a possible impact on other disciplines, such as semantic web research and information search technology.

IV. KEY CHALLENGE 2: LEARNING AND REPRESENTATION OF COMPOSITIONAL LEXICONS

In this section we outline what we see as the most important challenges for automatic language learning in cognitive robots. Amongst the various aspects and level of analyses of language (e.g. phonetics, lexical-semantic, syntactic and pragmatics), the discussion below will mostly focus on the issues related to the acquisition of meaning and words and the developmental emergence of syntactic constructs. This restricted focus is justified by the main aim of the paper on the modeling of lexicons acquisition in developmental robots. We begin with a necessarily brief sketch of what needs to be modeled, drawing on state-of-the-art accounts of language acquisition in cognitive linguistics and developmental psychology (IV.A). In section IV.B, we turn to the question of how these findings can inform experimental research in developmental robotics. Section IV.C then presents theoretical and experimental issues on acoustic packaging of action and language knowledge in robot-directed speech, as well as adult- and child-directed speech.

A. Language Acquisition: Insights from Linguistics and Psychology

Recent empiricist approaches to language acquisition (cf. Tomasello 2003 and Goldberg 2006 for surveys) have amassed considerable evidence that natural languages may be learnable without the aid of substantial language-specific cognitive hardwiring (‘Universal Grammar’). Key findings of this ‘usage-based’ approach to language acquisition relate to:

- the crucial role of general cognitive skills of cultural learning and intention reading;
- the grounding of language in both sensorimotor embodiment and social interaction;
- the significance of statistical learning and the distributional structure of children’s linguistic input;
- the item-based nature of early child language;
- the gradual emergence of grammatical abstractions through processes of schematization.

Given a sophisticated capacity for statistical learning (cf. Gómez 2007 for a recent review) as well as the peculiar structural properties of the specialized linguistic input that they receive (Pine 1994; Snow 1994), children are assumed to acquire complex compositional grammars through piecemeal schematizations over a massive body of memorized and categorized chunks of linguistic experience. Grounded in a set of specifically human skills of social cognition (‘shared intentionality’; cf. Tomasello et al. 2005) and closely interwoven with aspects of general cognitive development, the emergence of grammar is thus described as a slow and gradual transition from rote-learning lexical formulae (holophrases) to increasingly abstract (pivot schemas, item-based constructions) and ultimately fully schematic grammatical resources (abstract constructions, i.e. maximally generalized morphosyntactic rules). Syntactic categories of adult language (e.g. ‘determiner’, ‘verb phrase’, ‘infinitival complement clause’ etc.) are assumed to have no correlate in early learner grammars but only to arise during ontogeny (contrary to the ‘continuity assumption’ of nativist linguistic theories; cf. Pinker 1984). Strictly speaking, it is in fact not assumed that the learning process ever reaches an unchanging ‘final state’ at all – instead, linguistic knowledge is seen as constantly adapting to experience, and it is not assumed that speakers will always extract the highest conceivable generalizations from the data (Dabrowska 2004; Zeschel 2007). The co-existence of massive regularity and likewise massive residual idiosyncrasy in the system points to a cognitive architecture that redundantly represents both entrenched linguistic exemplars (memorized tokens of linguistic experience that are sufficiently frequent) and schematizations over such exemplars (as ‘emergent’ generalizations that are immanent in a set of stored instances), thus spanning a continuum from concrete lexical to abstract grammatical structure in a unified representational format (Bybee 2006; Abbot-Smith and Tomasello 2006). Crucially, due to the assumed tight feedback loop between speakers’ linguistic experience and the elements and structure of their internalized linguistic systems, quantitative-distributional properties of the input take centre stage in usage-based approaches to language acquisition.

We suggest that research in cognitive robotics should capitalize on this important aspect of the learning problem for the design of psycholinguistically informed experiments. Specifically, the design of learner input for such experiments should accommodate the following relevant insights into structural properties of child-directed speech (CDS): the linguistic input that children receive is considerably less variegated (i.e. it uses fewer words and constructions than speech directed at adults; cf. Cameron-Faulkner et al. 2003), it is highly stereotypical (words and constructions are used in their most common senses/functions; cf. Karmiloff and Karmiloff-Smith 2001), it is heavily redundant (i.e. strongly repetitive and reformulative; cf. Küntay and Slobin 1996) and also distributionally skewed in terms of word-construction-combinatorics (i.e. abstract constructions are familiarized via disproportionately heavy use of a single prototypical verb in the pattern; cf. Goldberg et al. 2004; Zeschel and Fischer, 2009). At the same time, when it comes to the core question of precisely how and exactly when specifically which kinds of abstractions are formed during language development, many details of learning-based approaches to language acquisition are as yet unresolved. For instance, are generalized constructional schemas only formed after an initial item-based phase of syntactic development, and possibly only after a certain critical mass of relevant ‘verb islands’ has been acquired (Tomasello 1992; Akhtar 1999)? Or are there ‘weak’ representations of such generalizations from very early on in development that just need to accrue salience before they can be evidenced in learner productions (Tomasello and Abbot-Smith 2002; McClure et al. 2006; Abbot-Smith et al. 2008), or primitive semantic structures to be found in CDS that correspond in some way to the grammatical constructions that are to be learned (Tellier, 1999; Fulop 2004; Sato and Saunders, forthcoming)? Is there a facilitating effect of semantic similarity on schema formation (Tomasello 2000; Morris et al. 2000)? Or is transfer of learning in syntax purely form-based (Ninio 2005a, 2005b)? It is by modeling such issues in appropriately designed artificial learners that future simulation studies and grounded robotic experiments that permit a systematic manipulation and full control of all supposedly relevant variables can make a unique contribution to language research within developmental science.

B. *Application to Automatic Language Learning*

Since the 1990s, there has been a sea change towards the use of statistical, corpus-based methods in all areas of computational linguistics, including the computational modeling of language acquisition. Work in this field constitutes a relatively recent addition to the methodological repertoire of developmental science (cf. Cartwright and Brent 1997; Elman 2006; Kaplan et al. 2008), and it has provided support for several important tenets of usage-based theories of language and its acquisition (cf. e.g. Solan et al. 2005; Borensztajn et al. 2008; Alishahi and Stevenson 2008). Also in the community of theoretical computational linguistics, which had traditionally seen the grammar learning problem to be intractable without Universal Grammar in view of Gold's results (Gold 1967), biases in the data such as typically found in CDS are beginning to be recognized as factors that ameliorate learning difficulty (Adriaans 2001; Clark 2004; Elman 2006). However, the algorithms which such approaches use to distil grammars from corpora are usually not only semantically blind, but also provided with certain grammatical information from the outset (e.g. part-of-speech annotation). From a developmental perspective, neither of these two features carries over to human learners – children ground linguistic signs in embodied experience, and they are not assumed to be equipped with adult syntactic categories such as ‘preposition’ or ‘conjunction’ from birth. Moreover, early caretaker-child interaction is restricted to joint attention scenarios (Dominey and Dodane 2004), which is a further property that lacks in these approaches.

By contrast, language research in cognitive robotics (e.g. Steels 2004) not only seeks to ground linguistic symbols in aspects of agents' sensorimotor experience, but also recognizes the need to address various social-cognitive and interactional underpinnings of the learning scenario (such as joint attention or perspective taking) that are beyond the scope of purely structure-oriented approaches to grammar induction from linguistic corpora. Regarding the present focus on the emergence of compositionality from holophrastic formulae, previous research (e.g. Sugita and Tani 2005) has already provided successful demonstrations of small-scale versions of this task: much in the same way that children learn to use holophrases like 'lemme-see!' to express complex meanings like 'show me this object that we are jointly attending to', robot learners can come to associate internally complex utterances with concurrently experienced perceptual-motor patterns, and subsequently break these patterns down to different formal and semantic constituents in a distributionally driven 'blame assignment' process of the type also ascribed to child language learners (Tomasello 2003). However, the compositional patterns acquired in previous robotic experiments on grounded learning are extremely simple and bear little resemblance to natural language grammars. Put differently, robot learning of holophrases with subsequent decomposition and generalization of an underlying argument structure construction constitutes an important prerequisite for higher-order grammar learning, but it is not the ultimate goal in itself. Key challenges that remain to be addressed on the way to truly naturalistic and successful (i.e. quasi-humanlike) language acquisition can be grouped into three categories:

- Social complexity: ultimately, all linguistic skills should be learned in an unsupervised manner from naturalistic social interaction with human communication partners, thus requiring a working implementation of various pre-linguistic (i.e. language-independent) pragmatic prerequisites for human ostensive-inferential communication (Sperber and Wilson 1995; Tomasello et al. 2005).
- Linguistic complexity: ultimately, the system should be able to reanalyze learned expressions as a compacted encoding of many grammaticalized dimensions in parallel (e.g. participant structure, tense, aspect, voice, mood, polarity, information structure, number, case, definiteness and reference tracking/binding to name but a few), and to combine the ensuing multilayered representations iteratively to produce and interpret progressively more complex (recursively embedded) syntactic structures
- Quantitative complexity: ultimately, the learning target should approximate the statistical structure of natural languages as they are actually experienced by a human learner, thus taking experiments from restricted laboratory settings involving just a handful of lexical items and even fewer grammatical patterns to essentially open-ended massive noisy input with naturalistic distributional properties.

For the moment, these objectives remain long-term goals that are beyond the scope of current experiments on grounded language acquisition. In fact, some researchers are skeptical that higher-order grammar learning along these lines can be achieved with current neural network technology at all (Steels 2005b; Steels and De Beule 2006) and advocate the use of symbolic grammar architectures such as Fluid Construction Grammar (FCG; Steels 2005a) and Embodied Construction Grammar (ECG; Bergen and Chang 2005) instead. However, if the initial focus is on the emergence of compositionality in language, action and action-language mappings, reliance on these mechanisms that include them cannot be built into the system as a design principle already, and any language-specific parameterization on which the learning should take place should not be presupposed and should generally be minimized as far as possible.

In sum, the logical next step thus consists in combining learning scenarios to allow for learning on the basis of distributional cues yet connected to real world, embodied experience. The first major challenge involved is thus the development of a suitable learning architecture that allows grammar induction from large amounts of linguistic data that are connected to categorized patterns of sensory-motor experience. It should permit the representation of constructional exemplars both as records of particular observed linguistic tokens and as records of previous successful analyses of these tokens (as implemented in symbolic approaches such as Batali, 2002). In addition, learners must be capable of mapping recognized individual elements in a string as well as properties of their sequential configuration to representations of objects, events and relations obtained from sensory-motor processing. The second major challenge then relates to the identification of suitable reduced-complexity learning scenarios and interactional tasks for robot language learning experiments that nevertheless accommodate relevant properties of the corresponding real-life challenge that children are facing. Starting out from corpus-based identifications of statistical properties of CDS that permit child language learners to extract the system underlying their earliest productively assembled multi-word combinations from the input, useful operationalisations/adaptations of these properties for the necessarily more restricted input of robots in grounded language learning experiments must be devised. Finally, a third major challenge for future research relates to the implementation of various social-cognitive and interactional prerequisites for child language acquisition in which the process of grounded distributional grammar learning is embedded. These include learners' pre-established understanding of the triadic structure of interactions between two interlocutors and an object that is being jointly attended to (Tomasello 1988, 1995; Carpenter et al. 1998a), their understanding of the behavior of others as intentional (Behne et al. 2005a, 2005b; Carpenter et al. 1998b; Tomasello et al. 2005), their understanding of the normative structure of conventional activities such as symbolic communication (Rakoczy 2007; Rakoczy et al. 2008) and their awareness of the cooperative logic of human communication (Liszkowski 2005, 2006; Tomasello et al. 2007). Especially when scaling up from highly restricted experimental settings to learning from more natural kinds of social interaction, the definition of useful operationalisations of these prerequisites constitutes a further important issue on the agenda of automatic language learning research.

Steels (2005) has recently proposed a model of evolutionary stages in the complexity of human language that provides a clear operational definition of qualitative changes in language development that can be easily tested in robotic experiments. If the above challenges are met, it is not only possible to systematically investigate the transition from holophrases to simple compositionality (stage III) in embodied, interactional experiments, but also from sequentially unordered multi-word speech to the item-based constructions of a syntactically structured grammatical language (stage IV) and ultimately to the abstract constructions of Steel's stage V-languages (higher-level constructions encoding the structural systematicity and internal coherence of a grammatical system at large). By investigating these issues along the lines of (and with special attention to unresolved questions in) current usage-based models of language acquisition in linguistics and psychology, such results promise to be of interest also to developmentalists outside the narrower field of cognitive robotics.

C. *Acoustic Packaging*

In developmental research, it has been recently shown that infants can use speech also as a signal structuring visual input. Brand and Baldwin (2005) suggested a tight interaction between speech and actions calling it "prosodic envelopes". This term refers to segments of both, the action and speech stream that reliably coincide. An example would be that important points in the action stream might be highlighted in the speech stream by a change in prosody or a break in an ongoing stream (Brand and Baldwin, 2005). This idea that the presence of a sound signal helps infants to attend to particular units within the action stream was originally proposed and termed acoustic packaging by Hirsh-Pasek and Golinkoff (1996). The authors argue that infants can use this 'acoustic packaging' to achieve a linkage between sounds and events (see also Zukow-Goldring, 2006) and to observe that certain events co-occur with certain sounds, like for example a door being opened with the word "open!". In fact, recently, many authors highlight the benefit of words or labels as signals that highlight the commonalities between objects (Waxman, 1999) and situations (Choi et al., 1999), facilitate object categorization (Balaban and Waxman, 1997; Xu, 2002), have the power to override the perceptual categories of objects (Plunkett, Hu and Cohen, 2008) and reason about physical events (Gertner, Baillargeon, Fisher and Simons, 2009). Thus, specific sound patterns and categories or types of sound patterns are suggested to help infants to get a better sense of the units within the action stream on the one hand. On the other hand the accompanying action provides pragmatic power to the linguistic information making it more perceivable and thus bootstraps language learning processes. In this vein, Gogate and Bahrick (2001) showed that moving an object in synchrony with a label facilitated long-term memory for syllable- object relations in infants as young as 7 months. By providing redundant sensory information (movement and label), selective attention was affected (Gogate and Bahrick 2001). However, Zukow- Goldring and Rader (2005) remind us that synchrony does not always refer to simultaneous occurrence, and that the exact parameters and theoretical background for the notion of synchrony have to be developed in order to understand how nonlinguistic and linguistic information is linked. In this point, it is of interest to investigate:

- how the speech stream overlaps with the action needed to fulfill the task, i.e. which parts of the motions are highlighted by what aspects of speech;
- how is the velocity profile of the action during the performance of the task and does the velocity differ when speech accompanies a motion;
- how do the intonation contours of the speech stream correlate with the action, i.e. when the contours are raising, is there also an up-motion noticeable and which parts of the motions are prosodically highlighted, e.g. by falling or raising contours?
- do the pauses in both channels (speech and motion) coincide?

V. KEY CHALLENGE 3: SOCIAL INTERACTION AND LEARNING

Traditional approaches for the study of communication and learning are based on a metaphor of signal and response (Fogel and Garvey, 2007). Recently, however, interactive and social aspects of learning have been emphasized (e.g. Nehaniv and Dautenhahn, 2007). Accordingly, for language to emerge, a learner – even when not fully able to signal and respond appropriately in an interaction, like a child that does not yet speak or, as investigated in human-machine interaction, a robot that does not function smoothly (Wrede et al., in press) – needs to treated as a partner, to which the other participant will attempt to adapt. Thus, de León (2000: 151) emphasizes that children "by the time they begin to speak, they have already 'emerged' as participants". In this section, we pursue topics that focus on the learning processes within the context of social interaction. It is becoming increasingly clear that children's conceptualization of the external world and their language system are scaffolded by interaction partners who adapt to them (Wood, Bruner and Ross, 1976).

What does this approach mean for a robot that is supposed to learn action and language? Imagine a child that sees a round thing that can roll. Adults call it "ball". What then gives the child a basis for assuming that that "ball" refers to the object and not to the action of rolling? For a long time, this central challenge of language acquisition had been explained in terms of mapping: A word typically has to be mapped either to an object, an action, or a relationship that holds amongst them. This mapping mechanism suggests a link but does not solve the question how the link is actually achieved. As already pointed out by Quine (1960), it is not clear how a child can achieve such mapping, because it is not the case that a child can fully rely on inner mechanisms allowing her or him to map the correct referent (an object or an action) onto a word. In addition, once a link between e.g. an object and a word is established, it is dynamic and can be changed (extended or specified) in the course of further experience. For example, children may map the word "ball" to the action of rolling but can define it more precisely later. Tomasello (2001) attacks the metaphor of mapping as false and suggests instead that learning is not only about cognitive achievement but also about embodied social interaction, in which a person uses a symbol for the purpose of redirecting another person towards the entity that is referred to. Moreover, children understand intangible situational concepts such as 'sleep' or 'breakfast' from a very early age (Tomasello 2003). In this social approach, it is not only the word that is the sole information available to the hearer for the resolution of reference. Also the behavior of the speaker and the circumstances of the situation as well as the hearer's experience contribute to the formation of the concept (Tomasello, 2001; Dausendschön, 2003; Rolf, Hanheide and Rohlfing, submitted). We aim, therefore, at investigating different forms of learning and scaffolding processes that help a learner to resolve reference in an interaction. Since human behavior is variable, scaffolding as a form of tutor behavior varies across persons. This variability causes problems in artificial systems that are expected to react appropriately to, for example, any form of showing an object (like pointing to it, holding it or waving with it) and to learn from examples that differ in certain aspects. Here, our goal is firstly to identify different forms of the tutoring behavior and then to seek for stability i.e. structure on different levels of analysis. As Conversational Analysis shows (Goodwin, 2000; Schegloff, 2007), the variability of human behavior in interaction can be assessed by discerning more general principles of communicational organization such as turn taking behavior. It is our goal to investigate such principles of organization in order to cope with variability in multimodal behavior.

Nevertheless, as for children, a robot's acquisition of language will necessarily reflect many characteristics of the linguistic behavior of those particular persons with whom it interacts (Saunders et al, submitted). Many properties of language development comprise evidence of mechanisms consistent with recent research in neuroscience proposing dual pathways, dorsal and ventral, e.g. in processing of articulation vs. processing of meaning (Saur et al., 2009). For instance, before they are able to use language to manipulate the intentions of others in the social world around them, infants are already learning to recognize word forms through interaction with their carers (Swingley, 2009). Moreover, the roles of mechanisms of intersubjectivity (Trevarthen, 1979, 1999) such as timing, turn-taking, or joint attentional reference (Tomasello, 2003) will scaffold and shape language acquisition in a social context.

The next sections will look at some of the most important issues in social learning and interaction in cognitive robots. In particular the focus will be: (i) contingency and synchrony in social interaction, (ii) cognitive architectures for intermodal learning, (iii) the scaffolding of behavioral, linguistic and conceptual competencies through social interaction, and finally, (v) a list of the main open research challenges.

A. Intermodal Learning: Contingency and Synchrony

Our perspective on developmental learning is based on the idea that learning is driven primarily through interaction with persons as well as the ambient environment (Saunders et al., 2007a; Saunders et al., 2009; Wrede, et al., 2009). This idea is supported by Csibra and Gergeley (2006) and Zukow-Goldring (2006), who state that learning through imitation is limited because the observed action does not always reveal its meaning. First-person experience as well as social scaffolding may be necessary to acquire certain behavioral competencies (Saunders et al., 2007a). In order to understand an action, a learner will typically need to be provided with additional information given by a teacher who demonstrates what is crucial: the goal, the means and – most importantly – the constraints of a task (Zukow-Goldring, 2006). The tutor, on the other hand, has to make sure that the learner is receptive, and thus ready to learn. They both follow certain interactive regularities. Such interactive rules have been assessed in terms of "grounding" (e.g. by Clark 1992) on a more abstract level but also in terms of "turn-taking" or "contingency" on a more perceptual level. With this sequential organization of an interaction, more systematicity can be derived from the variability of the behavior.

Clark (1992) provided one of the first grounding models with the claim that every individual contribution to a discourse has to be registered by the listener; that is, the listener has to provide a signal of understanding in order for both participants to add the content to their pool of commonly shared information and beliefs ("common ground"). On a more perceptual level, the term contingency refers to a temporal sequence of behavior and reaction, and it has been shown that it plays an important role in the process of developmental learning (e.g. Kindermann, 1993; Gergeley and Watson, 1999; Markova and Legerstee, 2006). In the literature, there is an agreement that contingency is an important factor in the cognitive development of infants – as researched, e.g., within the still face paradigm (e.g. Tronick et al., 1978; Muir and Lee, 2003). There is evidence that parents intuitively produce contingent actions, e.g. mothers have been shown to decrease their level of contingency with their infant's increase of development for a certain task (Kindermann, 1993). Infants have been shown to develop a sensitivity to contingent interactions around 3 months of age (Striano et al., 2005), and typically by the middle of the first year infants begin to move from canonical babbling towards syllable production related to their carers' speech (Vihman and Depaolis, 2000). This development is rooted in contingent interactions with adults. On this basis, infants not only detect contingency but also expect and try to elicit it (Okanda and Itakura, 2006). Thus, infants prefer persons who are and have previously been interacting contingently with them (Bigelow and Birch, 1999).

Against this experimental background, we argue that in order to pursue a social interaction, a system needs to be equipped with mechanisms that detect and produce contingent behavior. Tanaka and his colleagues (2007) have shown that when a system produces a contingent behavior, it gains more attention. The authors provided such a system to kindergarden children and found out that toddlers socialized with this system for a sustained period of time. This suggests strongly that the capability of producing a contingent behavior facilitates human-robot interaction. Yet, for a system to learn form a human, it is necessary that it not only can produce contingent behavior but also detect it. This can be achieved in gathering features that tutoring behavior exhibits in different modalities (Rohlfing et al., 2006). These features will guide the development of tutoring spotter for human-robot interaction systems. This will enable the system to pay attention to an ostensive action and the crucial parts or circumstances, which is helpful in resolving the question of what and when to imitate (Nehaniv and Dautenhahn, 2000).

Mechanisms that detect (and produce) contingency can be a precursor of later dialogical competencies as described in the framework of grounding. While contingency mainly describes a temporal pattern, where one event occurs as an answer to a previous one, grounding relies on semantic information in the sense that one event (or speech act) needs to be grounded by an interaction partner through a signal of understanding.

In recent developmental research, the problem of grounding a symbol has been assessed by analysing intersensory relations between multimodal signals. The idea is that e.g. words as acoustically perceived signal and actions as visually perceived signal may become paired by the shared temporal synchrony (Bahrick et al., 2004). In experimental settings, infants have been shown to learn a label for a new object more easily when the verbal referent was uttered in synchrony with a movement of the named object. In contrast, the name of an object being moved out of sync was not learned (Gogate and Bahrick, 2001). While temporal synchrony has been described as a means to provide "invariance", we are at the same time analysing the variability of the tutor behavior in order to better understand how tutors structure their actions towards infants. Here we follow the idea of "acoustic packaging" (see section IV.C of this paper) that has been pushed forward in experimental work by Brand and Tapscott (2007). Following Hirsh-Pasek and Golinkoff (1996), they suggested that acoustic information, typically in the form of narration, overlaps with action sequences and provides infants with a bottom-up guide to find structure within events. Brand and Tapscott's (2007) results support this idea indicating that infants appear to bind sequences of (sub)actions together based on their co-occurrence with speech. That is, given an action sequence and a verbal utterance overlapping with only part of this sequence, infants are likely to interpret only those action sequences as belonging together that fall within the range of the verbal utterance.

B. Intermodal Learning Architecture

Synchrony and contingency are two of the fundamental phenomena in tutoring and social learning. While there is a growing body of research on the phenomenon of synchrony, there exist only few models of synchrony on an artificial system (Prince et al., 2004; Kose-Bagci et al., 2009; Broz et al. 2009; Rolf et al, submitted). Based on current results reported in literature, models have to address the following questions:

- What is synchrony (in terms of a higher level and temporal structure as well as correlation measure)? (Definition)
- What are the entities that synchrony works on? (Segmentation)
- How can it be detected in the interaction? (Recognition)
- What functions does it serve? (Model)
- How does it vary in different speakers with their way of "acoustic packaging" and different situations (Analysis)
- What is the role of the different modalities (e.g. does vision provide primarily spatial information whereas auditory synchrony is more related to temporal structure?) and how do they interplay?

Currently, the scientific debate (Workshop on Intermodal Action Structuring, in ZiF, Bielefeld in July 2008) seems to converge towards a consensus that the important criteria for synchrony are (1) temporal co-occurrence of an event in different modalities and (2) a correlation between the characteristics of these events. In contrast, "inverse synchrony", meaning that events in two modalities show a temporally exactly disjunct distribution – such as a sequence of speech being followed by a speech pause with a sound of noise that is deliberately being framed by the tutor's utterance – does not constitute an instance of synchrony but rather describes the characteristics of causality or – within the context of interaction – contingency.

The importance of contingency has been recognized by computer scientists and there exist already some computational models for contingency (e.g. Movellan, 2005; Di Paolo et al., 2008). However, these models tend to be focused on a single modality and rigidly limited to specific concrete applications where an "event" has been clearly defined (e.g. Auvray et al., 2006). In order to foster research with respect to developmental learning on robots, the following questions need to be addressed in the near future:

- What is contingency (in terms of temporal structure as well as with respect to semantic content, if any)? (Definition)
- What are the entities that contingency works on? (Segmentation)

- How can contingency be detected in the interaction? (Recognition)
- What functions does it serve? (Model)
- How is it related to further sequential organization of interaction such as turn-taking? (Analysis)
- What is the role of the different modalities and how do they play together?

Against this background knowledge about synchrony and contingency within the framework of developmental robotics, the question of how these two phenomena are interwoven can be tackled. Our current hypothesis is that in order for an infant to learn new actions she or he can rely (1) on structured information provided by the tutor through the application of synchrony as well as acoustic packaging, and (2) on grounding on a more semantic and contingency on a more perceptual level.

Since we assume a continuous mutual adjustment (e.g. Fogel and Garvey, 2007; Wrede et al., 2009) between participants in the process of learning, it is important to investigate the role that contingency plays in the tutor's behavior with respect to synchrony. For instance, it might be the case that it is the infant, through her or his own feedback, who is actually designing the way the tutor is structuring the demonstrated action. The second issue regards the interdependence between the development of contingency and synchrony. Here we aim to understand how synchronous behavior can be a basis for contingent behavior. We are convinced that experiments of human-robot interaction, coupled with observations of parent-children tutoring situations, can shed light on these topics. In addition, the application of learning through interaction paradigms (Wrede et al., in press; Kose-Bagci et al., 2010) can help further robotic research to approach recognition or interaction capabilities (e.g. automatic speech recognition or dialog / contingency mechanisms), as it allows as it allows the analysis of more modalities (e.g. gaze, facial expressions for more socially related functions and hand movements / gestures for more task oriented functions), to develop new methodologies and to conduct evaluation cycles facilitating technical improvement.

C. Scaffolding of Behavioral, Linguistic and Conceptual Competencies

In learning to use language to communicate and manipulate the world around them, human children benefit from a positive feedback loop involving individual learning (by interacting with their hands and bodies with objects around them), social learning (via close interaction with parents and others), and gradual acquisition of linguistic competencies. This feedback cycle supports the scaffolding of increasingly complex skill learning and linguistic development giving the child ever greater mastery of its social and physical environment, as well as supporting the development of cognitive and conceptual capabilities that would seem impossible without language. To realize communication in robots a similar kind of feedback cycle supporting the scaffolding of behavioral, linguistic and conceptual competencies will be required. Such a realization will not only allow better understanding of possible mechanisms for such learning in humans, but also to achieve similar competencies in artificial agents and robots (even if they are not acquired by exactly the same routes).

Social interaction may also allow meaning to be grounded in early childhood language through shared referential inference in pragmatic interactions, whereby shared reference provides the necessary statistical bias to allow focused learning to take place. In order to create appropriate conditions for language learning in robots it would therefore be necessary to expose the robot to similar physical and social contexts. This might be achieved via an interaction environment between a human and a robot where shared intentional-referencing and the associations between physical, visual and speech modalities can be experienced by the robot. In fact the bias of the learning context may require the human interaction partner to treat the robot as an intentional being, even though the robot may have no intentional capability (Cowley, 2008). The output of such studies if combined to yield word or holophrase structures grounded in the robot's own actions and modalities, e.g. as in (Saunders et al., submitted), would provide scaffolding for further proto-grammatical usage-based learning. This requires interaction with the physical and social environment involving human feedback to bootstrap developing linguistic competencies. These structures could then form the basis for further studies on language acquisition, including the emergence of negation (see below) and more complex grammar.

A possible direction (Saunders et al, 2009) for achieving such competencies is to study mechanisms whereby robots or other synthetic agents are expected to exhibit:

- holophrase learning
- segmentation of utterances down to word level
- the grounding of words and lexicon usage frames in action and object learning via physical interactions
- the bootstrapping of simple usage-based proto-grammatical structure via human scaffolding and feedback.

D. Negation

The emergence of various forms of negation (Nehaniv et. al. 2007; Förster et al., in press) through the mechanisms of communicative social interaction is considered to have been an extremely important qualifier in the emergence of symbolic representation capabilities. Very early in the language development of children negative speech acts emerge, such as the rejective and holophrastic "No!", e.g. to refuse certain food or a particular activity. Other functions of negation in early child language include nonexistence, prohibition, denial, inability, failure, ignorance, expressing the violation of a norm, and inferential negation (Choi, 1988).

The mentioned examples show that the various functions of early negation are not necessarily related to each other and that the term encompasses a set of functions that is remarkably larger in scope than the well known negation of propositions in particular. Which function a particular case of negation has is obviously highly context-dependent in more than one sense. It depends on the linguistic context on one hand but also on the situational context. An artificial agent that is supposed to appropriate negative humanlike speech acts therefore cannot derive the meaning of these utterances through a simple lexical analysis. It has to take into account the situation in which the dialogue takes place (joint attentional frame). Current models either choose the representation of objects (Roy, 2005b) or actions (Saunders et al. 2007) as basic representational building blocks. Different functions of negation tend to operate on the other hand more on objects (nonexistence) or more on actions (rejection, prohibition), which suggests that the support for certain forms of negation may be rather weak in each of these existing models. Thus, for achieving the emergence of the full range of early negation, ways have to be found to bypass these difficulties.

Future studies should consider questions such as: (1) Which features must be supported by frameworks for grounded language learning and imitative learning to enable the representation and production of speech acts that involve negation? (2) To what degree and in which form must motivation in the robotic platform be modeled for this purpose, as the majority of early negative speech acts are acts of volition and not acts of description? (3) Can negation emerge as purely syntactical construction or is it necessary to modify the underlying grounding mechanism?

E. Open and Challenging Research Questions in Social Learning and Language

Insights of Wittgenstein (1953) and Millikan (2004), and more constructively Steels (1998, 2007), suggest that to understand signaling and linguistic behavior, one needs to take into account usage in its pragmatic embodied social context. The learning of communicative signaling and linguistic systems (at the ontogenetic, diachronic, and evolutionary levels) are moreover shaped, not only by details of perception and embodiment, e.g. Cangelosi and Parisi (1998), but also by details of transmission, sources of error and variability, as well as feedback and repair mechanisms e.g. (Steels, 1998, Smith et. al., 2003, Wray 1998)).

The overall approach is to understand constructively what mechanisms could be responsible for the ontogeny of linguistic competencies. That is, for such a constructive theory of language to be successful it is necessary to build an instantiation that exhibits the phenomenon to be explained, and, moreover, different constructive mechanisms could be assessed against each other by comparing what they actually generate. Preferably these constructivist evaluation test-beds must involve learning in embodied social interactions with humans and physical interactions with rest of the robot's environment.

Open and challenging research questions in this area include:

- To what extent can the methods be scaled for human-like acquisition of linguistic abilities?
- What `cognitive' capabilities are necessary for recruitment in the development of human-like linguistic competencies?
- Is it necessary to build in universal mechanisms for categorization and generalization, propositional logic, predication, compositional syntax, etc?
- Can these emerge from more elementary processes, such as Hebbian learning, 'chunking', sequential processing and locality principles or more general cognitive capacities such as perspective taking; action hierarchies; expectation, prospection and refusal?
- How can different types of linguistic negation be acquired by a robot or synthetic agent?
- To what extent are these mechanisms for the development of linguistic abilities universal, i.e. applicable for any given target natural language?
- What are appropriate semiotic frameworks for pragmatic acquisition of language usage (e.g. fluid construction grammar in Steels and Wellens, 2006, embodied construction grammar in Bergen and Chang, 2005, or dynamic syntax in Kempson et.al. 2001)?
- To what extent are purported explanations consistent not only with individual ontogeny of linguistic capabilities but also with diachronic (transmission) and evolutionary (philogenetic) considerations?

VI. KEY CHALLENGE 4: PUTTING ACTION AND LANGUAGE TOGETHER AGAIN

The three sections above have considered, in part independently, the key research issues on action learning, lexicon acquisition and social interactions. However, as discussed in the introduction, and as supported by neuroscientific and psychological evidence, cognitive development and general cognitive processing are based on the strict interaction and co-dependence between language and action. This section focuses on the research issues that specifically address the form of language/action interaction and the phenomena underpinning it. Initially the focus is on research based on neurorobotic models for investigating the neural representations of action and language. We then consider cognitive robotics approaches to the psychological phenomena of language grounding in action. Finally, we consider the phylogenetic dimension of cognition evolution and how robotics models can help us investigating the contribution of action cognition in the origins of language.

A. Neural Representations of Action and Language Knowledge

Neuropsychological and neuroscientific literature on language processing in the brain is quite extensive and consistently demonstrates the close integration of action and language processing (Pulvermuller 2003). For example, various studies have analyzed the neural correlates of the processing of various word classes and the verb-noun dissociation in patients. In Cappa and Perani (2003) a review of the neuroscience studies on the neural processing of verbs and nouns is presented. The authors found a general agreement on the fact that the left temporal neocortex plays a crucial role in lexical-semantic tasks related to the processing of nouns whereas the processing of words related to actions (verbs) involves additional regions of the left dorsolateral prefrontal cortex. For example, in the well known neuropsychological study on verbs and noun processing, Damasio and Tranel (1993) reported that most of the patients with selective disorders of noun retrieval had lesions in the left temporal lobe. Instead, verb impairment was associated with damage on the left prefrontal cortex. In a PET study, Martin and colleagues (1995) compared color naming (nouns) and action naming (verbs). They observed a selective activation for color naming of the left fronto-parietal cortex, the middle temporal gyrus, and the cerebellum. Perani, Cappa et al. (1999) also used PET for the processing of concrete and abstract verbs and nouns in Italian. Results indicated that left dorsolateral frontal and lateral temporal cortex were activated only by verbs. In the comparison of abstract and concrete words, only abstract word processing was associated with selective activation of the right temporal pole and amygdala and the bilateral inferior frontal cortex. Finally, in evoked potential studies it was reported that there is selective activation of the frontal lobes for action words (Preissl, Pulvermueller et al., 1995). This difference is related to the semantic content of words rather than to grammatical differences, since no difference was observed between action verbs and nouns with a strong action association (Pulvermuller, Mohr and Schliechert, 1999).

Brain simulation models, such as those of computational neuroscience, have rarely focused on complex linguistic behavior, except for a few studies (e.g., Just et al. 1999). This is due to the complexity of the various linguistic functions (speech processing, lexical and semantic knowledge, syntax) to be included in a model. However, brain simulation models have been commonly developed for a variety of behavioral and cognitive abilities, such as vision, memory, and motor control. More recently, in such models the method of synthetic brain imaging (Arbib et al. 2000; Horwitz et al. 1999) has permitted a more strict integration of experimental data and computational models and a direct comparison of performance in artificial and natural brains. In addition, cognitive models based on neuro-cognitive robots can be used to investigate the neural correlates of motor and linguistic behavior. In Cangelosi and Parisi (2004) a computational model of action and language learning is proposed that specifically looks at action/language integration. This model if based on simulated robots (i.e. agents with 2D robotic arm for manipulating objects) that are evolved for their ability to (a) manipulate objects such as a vertical and a horizontal bar, and (b) to learn lexicons describing the respective agent’s interaction with the objects. The agent’s motor and linguistic behavior is controlled by an artificial neural network. We study the consequences in the network's internal functional organization of learning to process different classes of words. Agents are selected for reproduction according to their ability to manipulate objects and to understand nouns (objects’ names) and verbs (manipulation tasks). Synthetic brain imaging techniques (Arbib et al. 2002) are then used to examine the functional organization of the neural networks. Results show that nouns produce more integrated neural activity in the sensory processing hidden layer, while verbs produce more integrated synaptic activity in the layer where sensory information is integrated with proprioceptive input. Such findings are qualitatively compared with human brain imaging data (Cappa and Perani 2003) that indicate that nouns activate more the posterior areas of the brain related to sensory and associative processing while verbs activate more the anterior motor areas.

These results indicate how neuro-robotic models, directly constrained on known neuroscientific and psychological phenomena, can be used to directly address some of the open questions on the neural representations of action and language knowledge. In particular, future developmental robotics studies based on neuro-robotics agents can be used in the computational modeling of issues such as (i) qualitative and quantitative differences in the neural representations of action and language concepts, (ii) amount of overlap/difference between motor representation patterns and linguistic neural activations, (iii) graduality of motor representation components in various syntactic classes and (iv) developmental timescale and dynamics in the acquisition of motor and linguistic concepts.

B. Action Bases of Language Processing

Psycholinguistic data on Action-Compatibility Effects (ACE) during language comprehension tasks (Glenberg and Kaschak, 2002) support an embodied theory of language that strictly relates the meaning of sentences to human action and motor affordances. Glenberg and Robertson (2000) have proposed the Indexical Hypothesis to explain the detailed interaction of language and action knowledge. This suggests that sentences are understood by creating a simulation of the actions that underlie them. When reading a sentence, the first process is to index words and phrases to objects in the environment or to analogical perceptual symbols. The second process is deriving affordances from the object or perceptual symbol. Finally, the third process is to mesh the affordances into a coherent set of actions. The meshing process is guided by the syntax of the sentence being processed. This suggests a parallel between syntax and action. Syntax has the role of combining linguistic components into an acceptable sentence. Motor control has the role of combining movements to produce the desired action. Moreover, Glenberg (personal communication) suggests that syntax emerges from using linguistic elements to guide mechanisms of motor control to produce effective action or a simulation of it. Such a view is compatible with construction grammar hypothesis that suggests that linguistic knowledge consists of a collection of symbolic form-meaning pairs reflecting, amongst other things, action roles and properties.

Developmental robotics experiments can be used to specifically investigate language grounding and action-compatibility effects in syntax processing. Robots can initially be trained to acquire an action repertoire producing various motor affordance representations and constructs (e.g. give-object-to, receive-object-from, lift-object etc.). In parallel the robots will learn the names of actions and objects name. Further testing of the robot responses to ACE-like situations, and systematic analyses of the robot's internal (e.g. neural patterns controlling the robot motor and linguistic behavior) can provide insights on the fine mechanisms linking microaffordance action representations with language.

C. *Evolutionary Origins of Action and Language Compositionality*

The relationship between language and action is particularly important when we consider the striking similarities and parallels that have been demonstrated to exist between the linguistic structure and the organization of action knowledge. As discussed in section 3, action knowledge can be organized into compositional and hierarchical components. Language has two core characteristics: Compositionality and Recursion. Compositionality refers to the fact that a series of basic linguistic components (i.e. word categories such as nouns, verbs, adjectives etc.) can be combined together to construct meaningful sentences. Recursion refers to the fact that these words and sentences can be recursively combined to express new sentences and meanings. These mechanisms create a parallel between the structure of language and that of meaning (including sensorimotor representations). When considering such remarkable similarities between language and action, some fundamental questions arise: Why do language and action share such hierarchical and compositional structure and properties? Is there a univocal relationship between them (e.g. the structure of action influences that of language, or vice versa), or do they affect each other in a reciprocal way? Do these two abilities share common evolutionary, and/or developmental, processes?

These scientific questions will be investigated through new robotic experiments based on the combination of evolutionary algorithms and ontogenetic/developmental learning algorithms. These experiments will be based on robotic simulations due to time constraints involved in evolutionary computation (i.e. parallel testing of many robots within one generation, to be repeated for hundred of selection/reproduction cycles). Experiment will directly address some of the language origins hypotheses on action/language interaction. For example, one study will consider Corballis (2002) hypothesis that language evolved from the primates' ability to use and make tools and the corresponding cognitive representation that such a compositional behavior requires. Evolutionary simulations will first look at the evolution of tool use and object manipulation capabilities. Subsequently, agents will be allowed to communicate about their action and object repertoire. The analysis of evolutionary advantages in pre-evolving object manipulation capability will be considered. Another simulation will consider Greenfield's (1991) study on sequential sorting behavior and its relationship to language and motor development (evolutionary and ontogenetic). Children use different dominant strategies in sequential tasks such as nesting cups, e.g. from an early "pot" strategy (move one cup at a time) to a later "subassembly" strategy (moved pairs or triples of stacked cups). Greenfield suggests that language and sorting task processes are built upon an initially common neurological foundation, which then divides into separate specialized areas as development progresses. Such a hypothesis will be studied in simulation on the manipulations of the topology of the neural network controlling the agents' linguistic and motor behavior. Simulations will provide further insights on the evolutionary relationship between action and language structure, as well as providing new methodologies for the combination of evolutionary and ontogenetic learning mechanisms in communicating cognitive systems.

VII. 7. A ROADMAP FOR FUTURE RESEARCH

The above research issues constitute some of the key challenges for research in developmental cognitive robotics, in particular regarding ongoing and future work on linguistic communication between robots and human-robot interaction. Other core issues in developmental robotics regard additional linguistic/communicative capabilities, such as new developments in phonetic and articulatory systems, or new insights in concept acquisition and the influence of language on the process, as well as additional cognitive and behavioral abilities. These include research on motivation and emotions, on perception and action, on social interaction, and on higher-order cognitive skills such as decision making and planning.

In addition to research specifically addressing individual cognitive skills and their interaction, other core cognitive robotics research issues regard general cognitive capabilities. In particular, two main challenges regard the further development of learning techniques (e.g. development of new, scalable learning algorithms) and the design of brain-inspired techniques for robot control.

If we consider future advancements on developmental robotics and the parallel progresses in the various cognitive and behavioral capabilities, we can identify a potential sequence of milestones for what regards specifically research on action and language learning and integration (Table 1). These milestones provide a possible set of goals and test-scenarios, thus acting as a research roadmap for future work on cognitive robotics. That it, we do not intend to propose a fully defined and rigid sequential list of milestones, especially as there will be overlap of cognitive capabilities development in the transition between milestones/stages. We rather want to suggest specific experimental test scenarios and target cognitive capabilities that should be studied in future developmental robotics research. These experimental scenarios can also be used to evaluate the progress in the various milestones.

For practical reasons, milestones are grouped along a temporal scale from the next two 2, 4 and 6-8 months, to a more distant times scale of 10, 15 and 20 years' perspective. The descriptions of the closest (2-8 years) three milestones will be more extensive that those for the more distance milestones (10 years and over), as it is very difficult to foresee now the detailed development for longer term goals.

<MILESTONES TABLE ABOUT HERE>

A. *Milestone for Action Learning Research*

This section gives an overview of the six milestones on action learning. We will describe in more details the first three milestones given current state of the art and related foreseeable advancements in action learning research. The remaining longer term milestones will be briefly introduced, as their detailed specification will depend much on actual achievements in the preceding 2-8 years of research.

Action Learning Milestone I (~ next 2 years). The first milestone, crucial to human development, has to do with the acquisition of the simplest possible actions. Actions here are intended not as simple movements and, therefore, we are not considering a purely motor – read muscular – aspect, but rather a complete sensorimotor primitive. We see action (as opposed to movement or reflexes) as goal-directed movements, initiated by a motivated subject and exploiting prospective capabilities (predicting the future course of the movement) – see (von Hofsten, 2004). This difference is important because it shifts the focus of observation from the control of the muscles to the connection between a goal, a motive and predictive information (e.g. the context of action execution). Actions are in a sense defined by the "goal" not by how the goal is achieved – that is, grasping can happen with the left or right hand as well as with the mouth. This is why the capacity of categorizing, perceiving objects, events and states parallels the development of action (primitives). Developmental psychology supports this view as in e.g. (Woodward, 1998) together with neurophysiology as summarized in (Jeannerod, 1997). It is also evident that in humans, these abilities are pre-linguistic (e.g. reaching develops at around m3, early grasping and manipulation soon after – m4-5 –, the hand is adjusted to the object's size at around m9 and they're finally integrated in a single smooth action at around m13 of age). It is worth noting that in human infants, action develops from pre-existing basic structuring – both of the motor system (de Vries et al., 1982) and of the somatotopy of the sensory system (Johnson, 1997; Quartz and Sejnowski, 1997; von der Malsburg and Singer, 1988). This prestructuring seems to emerge from very specific mechanisms already in operation in the fetus. Similarly, some basic knowledge about objects (e.g. that motion boundaries are representative of objects), about numbers (e.g. one vs. two, quantities) and about others (the presence of other people) seems to be available to the newborn (Spelke, 2000).

This step, fundamental to human development, seems to be also necessary in building a robot that develops. Here, our hypothetical milestone has to include: the ability to detect objects (though not necessarily their identity), to gaze (although not as smoothly as in adults), reach and clasp the hand around the object. These abilities are supported by an improvement in the ability to predict internal dynamics (self-generated forces), sitting (thus freeing the hands from their support function) and by an improvement in vision (binocular disparity develops by m3 or so), smooth pursuit becomes fully operational and by an increased social interaction (correct hemisphere of gaze). On the computational side, achieving a similar milestone requires methods for learning that show certain "good properties" like incremental learning, bounded memory and representation complexity and that provide certain guarantees (formal) of convergence. Ideally, we would like to combine full online methods with the good properties of convergence of batch methods, although typically online methods are evaluated by the number of mistakes (to be bound) rather than convergence which lacks of clear significance (Bengio and LeCun, 2007).

Action Learning Milestone II (~ next 4 years). Our second milestone refers to the flexible acquisition of action patterns and their combination to achieve more complex goals. Evidence from neurophysiology shows that this is the case also in the brain – for example, in non-human primates the flexible use of actions with respect to external visual cues has been demonstrated (Fogassi et al., 2005; 1998) . Mirror responses have been found in the parietal cortex that depend on the goal of the action (e.g. eat vs. place) as a function of the presence of certain objects (e.g. a tray for placing instructs the monkey to execute a place). Some neurons in this area start responding before the hand action becomes unambiguous showing that the extra visual cue (the tray) determines their activation. In a sense, the other's intention is encoded in the presence of the specific context (exemplified by the tray). For developmental robots the possibility of exploiting external or self-generated forces together with the flexible reuse of motion primitives is one step forward towards the acquisition of a "grammar" of action (or a vocabulary of actions as described by (Fadiga et al., 2000)). Here many different methods have been proposed in robotics, in particular, to represent complex actions as subactions and to combine them smoothly. These range from the use of multiple forward-inverse models as in the well-known MOSAIC method (Haruno et al., 2001) and the more recent HAMMER (Demiris and Khadhouri, 2006) to trajectory decomposition as in Billard et al. (2004) or in (Chakravarthy and Kompella, 2003) using a formalism derived from catastrophe theory. The problem of exploiting self-generated forces has been addressed recently by Nori et al. (Nori et al., 2009) and requires the autonomous acquisition of dynamical models of the body. This skill also requires "developmental learning" methods that can operate in high-dimensional spaces as in e.g. (Schaal et al, 2000). An important element in the definition of motor primitives, their combination, and generation of action is the detection of affordances. The term affordance was originally used by James J. Gibson (Gibson, 1977) to refer to “action possibilities” on a certain objects, with reference to the actor’s capabilities. More recently, neural responses which can be made analogous to the perception of affordances have been found in the monkey (Gallese et al., 1996) and computational approaches were formulated in robotics (Metta and Fitzpatrick, 2003). It is possible to build formal models of affordances and relate learning, detection and imitation. This approach has been pioneered in models of the mirror neurons (Metta et al., 2006) and extended recently to include various modalities including word-object associations as in (Krunic et al., 2009). Bayesian methods form a very natural formalization of affordance learning by taking into account the uncertainty of the physical interaction between effectors and objects as well as the multiple action possibilities provided by objects to complex manipulation (e.g. with multiple fingers).

Action Learning Milestone III (~ next 6-8 years). The third milestone regards the processes when social (imitation) learning word to object association starts to develop. Simultaneously it is possible to imagine simple syntactic associations between actions and objects via the affordance mechanism discussed above. At this stage, around the onset of the first single world-single object associations, infants are perfect at reaching and getting possess of objects, in detouring around barriers and in separating the “line of sight” from the “line of reach” thus effectively enabling interaction in complex scenarios (Diamond, 1981). While social behaviors can be already seen in newborns, at this stage (12m), infants acquire the ability to use pointing for sharing attention or requesting an object. Requests can be more subtle as asking for the object name, or information about the object. Some studies show that pointing at 12 months predicts speech production rates at 24 months (Camaioni et al., 1991) and that the combination of pointing and a word which differs from the object signed precedes two-word sentences, the first grammatical construction (Goldin-Meadow and Butcher, 2003).

Action Learning Milestone IV (~ next 10 years). This longer term milestones refers to (i) the acquisition of action generalization rules through social learning and (ii) the development of an ability to correlate action and language generalization capabilities though the sharing of representation and rules. For action generalization rules we refer here to the development of higher-order representation of action constructs that share common sensorimotor actuators and strategies.

Action Learning Milestone V (~ next 15 years). One component of this milestone refers to the acquisition of the ability to generalize over goals. Once the robot has developed goal-directed behavior for a larger set of independent goals, we expect robots to acquire generalization capability for goals that share the same action and social roles. This milestone also focuses on further extension and enhancement of the shared action/language integration system. For example we expect research to focus on the development of higher-order cognitive abilities to correlate recursive and composite actions with recursive syntactic construct.

Action Learning Milestone VI (~ 20+ years). This milestone regards further development of an open-ended capability to learn rich action repertoires based on complex social and linguistic descriptions, as also detailed in the Milestones VI of the language and social learning components.

B. *Milestone for Language Learning Research*

We propose to address these issues with incremental increases of the complexity of the learning architecture, scenario and task:

Language Learning Milestone I (~ next 2 years). This milestone documents the general feasibility of adopting a grounded neural network approach to learning an elementary repertoire of lexical items and productive basic sentence types (argument structure constructions) and provides a precise empirical characterization of the initial learning target, i.e. children's actual experience with the most basic English sentence types and their most common realizations in the input. In addition, work in this period lays the computational foundations for embodied robotic learning of the investigated patterns in restricted learning by demonstration tasks. Specifically, the consortium will present a demonstration of abstract grammatical construction learning that proceeds from the acquisition of holistic utterance-scene pairs over the segmentation of recurrent constitutive elements of the acquired holophrases to their compositional recombination (i.e. generalization).

Language Learning Milestone II (~ next 4 years). The milestone scales the lexicon up to multiple grammatical constructions that are acquired in parallel, ultimately embracing all five of the basic sentence type/argument structure constructions of English and the event types that are associated with their prototypical uses.

Language Learning Milestone III (~ next 6-8 years). This introduces implementations of the most elementary socio-cognitive/pragmatic capabilities that are required for simple linguistic interactions (e.g. joint attention, perspective taking, turn taking). With these capabilities in place, language learning experiments can shift from learning by demonstration to more naturalistic forms of language learning from social interaction (albeit initially confined to fairly rigidly restricted language games proceeding by fixed protocols).

Language Learning Milestone IV (~ next 10 years). This milestone marks a progressive diversification of the linguistic resources employed, as well as a more naturalistic approximation of their actual quantitative proportions in children's linguistic input, extending current learning architectures progressively to combine grounded learning with large scale distributional learning. Using corpora of child-directed speech as an empirical yardstick, more and more words and constructions are fed into the still restricted/non-spontaneous tutor-learner interaction according to distributional patterns extracted from naturally occurring child-directed speech.

Language Learning Milestone V (~ next 15 years). This relates to advanced skills of social cognition that must eventually be incorporated into robotic systems at some point or other (however simplified) if serious progress towards human-like communicative capabilities is to be made: these higher-level prerequisites for ostensive-inferential communication include such complex and contextually contingent capabilities as action recognition, goal inference, belief ascription and everything else that is commonly subsumed under the notion of "shared intentionality" (Tomasello et al. 2005). In general, the more aspects of these distinctly human traits can be adapted and rebuilt in artificial systems, the more open-ended the learner's capacity for flexible intelligent interaction during language learning tasks and communication experiments will be.

Language Learning Milestone VI (~ next 20+ years). Finally, to the extent that all of the above has been integrated more or less successfully into a running system, milestone VI marks the stepwise addition of further grammatical and distributional complexity in order to further approximate the real-life challenge facing child language learners. Among other things, this additional complexity may relate to such dimensions as the relation between speech act participants and the proposition expressed (with the grammatical correlate sentence mood), the relation between speech act time and event time (grammatical reflex: tense) or the conceptualization of event structure and event sequencing (grammar: aspect). Likewise, the input used for pertinent learning experiments should increasingly resemble the quantitative properties of naturally occurring child-directed speech. In this, milestone VI marks incremental increases both in the grammatical and in the quantitative complexity of learners' linguistic input, thus paving the way to progressively open-ended interactional scenarios for grounded language learning experiments.

C. Milestone for Social Learning Research

Social Learning Milestone I (~ next 2 years). The first target in social research involves studying and implementing non-verbal social cues for language and skill learning. The second target is modeling holophrase acquisition via intermodal learning; this entails sensitivity to aspects of acoustic packaging (cf. Sec. IV.C). The first target attempts to exploit biased learning via a form of rudimentary intentional reference. This can be achieved via joint attention between robot and human whereby the robot responds to gaze direction, mirroring and turn-taking in the interaction with the human interaction partner. The non-verbal clues direct robot attention to the actions or objects. Language acquisition proceeds by associating the robot's focus of attention (including its full sensorimotor feedback) with salient aspects of the human's speech modality.

The second challenge regards the modeling of holophrase acquisition via intermodal learning. This particularly refers to the implementation of the acoustic packaging that automatically permits the division of a sequence of events into units and thus there is synchrony between language and events.

Social Learning Milestone II (~ next 4 years). The roadmap development in a 4-year perspective within the social learning scenarios expects that an ability to detect and exploit tutoring interactions will be developed in humanoid robots. This would be achieved by extending and enhancing the developments in previous milestones. Scaffolded learning of hierarchical behaviors in social interaction and the learning of grammar and vocabulary complement and enhance each other. Additionally further research on joint intentional framing and referential intent should be carried out together with the basic ideas for acquisition of negation usage of various types (e.g. refusal, absence, prohibition, propositional denial). Most of the latter require some modeling of motivation (volition and affect) on the part of the robot, as well as temporal scope encompassing memories and habits.

Tutoring plays an important role in understanding actions. Research would consider how tutoring could be used for learning, how complex actions could be structured, which kind of units could be observed and how speech/sound signals (acoustic packaging) could be modeled. Studies would also be carried out to extend previous research in order to establish how to enhance rudimentary intentional reference to more sophisticated mechanisms for joint intentional framing and referential intent. This would take into account both interaction partners' gaze, speech, gesture and motion clues. A further outcome of this milestone would be the acquisition of the meaningful usage of many forms of negation. Negation has been considered as a primarily grammatical phenomenon. However negation appears to be quite varied and emerges long before the production of grammatical utterances in young children. The part of the roadmap would lead to a better understanding of how negation fits into developmental learning and with the rest of language acquisition.

Social Learning Milestone III (~ next 6-8 years). At this stage we would expect that research will build on previous achievements to focus on two main areas of social learning and language. Firstly the development of architectures capable of exploiting pragmatic skills such as sequential interactional organization (contingency, turn-taking) and use of prosody for grammatical learning and secondly being able to harness Model/Rival (M/R) learning, motivational systems and predictive models of social interaction. Prosodic bias occurring in speech directed at infants could be associated with gestural indications to not only highlight key parts of speech but also provide clues to the grammatical nature of language in the interaction.

A key issue in language research is also that of individuating participants and the acquisition of pronoun and anaphora usage and grammatical agreement based, e.g., on person and number and, in some languages, gender. For example, to understand that "I" means the speaker need not necessarily arise in pure two-way interaction (one interaction partner might use "I" to refer to themselves but not to the other partner) , however "I" can be obtained from 3-way interaction. Furthermore it has been shown from animal studies that a 3-way interaction (introducing a rival who also acts as a model for functional use of utterances) accelerates (language) learning. Further investigations of the role of these interaction phenomena are necessary.

Social Learning Milestone IV (~ next 10 years). The 10 year goal would be to exploit interactions of prosody, internal motivation, inter-subjectivity and pragmatics in language acquisition and dialogue whilst developing architectures based on intermodal learning and sensitivity to a tutor.

Social Learning Milestone V (~ next 15 years). A longer term goal would be that of temporally extended understanding of the social motivations and intentions of other minds, context, and (auto)biographic and narrative (re)construction. Thus rather than focusing and responding to events occurring in the immediate moment the robot language learner expands their scope to encompass a wider temporal horizon. This necessarily would require the development of mechanisms to cope with extended context including both the robot's own history and the ability to construct such events in relation to an interaction partner. We would envisage therefore the development of first systems that are capable of social learning and sequential organization of interaction in specific scenarios.

Social Learning Milestone VI (~ 20+ years). A very long term goal would be the development of systems that are capable of social learning and pragmatic organization of interaction related to grammar, language, and behavior in various open-ended scenarios. Clearly this would build of the achievements of earlier parts of the roadmap.

D. Milestone for Cognitive Integration Research

All previous milestones, though grouped for sake of clarity in the three research challenge areas of action, language and social learning, already include foreseen development that imply the integration of the tree cognitive capabilities. In the section below we will list additional future progress milestones not explicitly discussed in the previous section.

Cognitive Integration Milestone I (~ next 2 years). This milestone explicitly refers to the development of robotics cognitive models able to integrate basic action and naming representations into emergence shared representation roles for both actions and names, implicitly integrating the capabilities discussed in the previous set of milestones. For example, we expect here that any experiment of the learning of labels for individual objects and action categories is implicitly linked, and integrated with, the experiment on the acquisition of new motor primitives and their application to object manipulation contexts. This integration assumes the sharing of internal representation and processes for both sensorimotor and linguistic knowledge. And we expect that such a progress in the acquisition of new action and language concepts is always developed in a social learning and imitation context.

Cognitive Integration Milestone II (~ next 4 years). A further area of research achievable in a four-year perspective will be the simulation of embodiment phenomena in language learning robots such as the Action-Language Compatibility effects (Glenberg and Kashark 2002; Tucker and Ellis 2004). Another milestone regards the development of evolutionary models demonstrating the co-evolution of action and language skills for simple grounded lexicons and simple syntactic constructs (e.g. agent-verb-patient, agent-verb-preposition).

Cognitive Integration Milestone III (~ next 6-8 years). Expected ongoing progress on the development of large-scale computational neuroscience models could lead to the application of these brain models to robotics action and language integration systems. This would for example build up on previous milestone reproducing behavioral action-language compatibility effects to computational neuroscience models investigating fine neural mechanism explaining facilitation and inhibition effects in multiple object scenarios (Ellis et al. 2007).

Cognitive Integration Milestone IV (~ next 10 years). This longer-term milestone refers to the development of general-purpose grammatical constructions for the creation of new complex motor and perceptual concepts. As specified in the language milestone IV section, at this stage we expect a progressive diversification of the linguistic resources and acquisition of large scale distributional learning. In this integrative milestone the focus in on how more advanced sensorimotor knowledge systems and richer social factors can help this complexification of the linguistic system.

Cognitive Integration Milestone V (~ next 15 years). New developments consequent to the acquisition or large lexicons and syntactic capabilities will allow the testing in robotics models of challenging research issues in embodiment literature. For example, the sensorimotor grounding of abstract concepts is a challenge for embodiment theory of cognition (Barsalou 1999; Andrews et al. 2009; Kousta et al. 1999). Embodied theories should be able to explain the contribution of sensorimotor and affective knowledge can explain the acquisition of abstract concepts, such as happiness and beauty, or non-semantic words such as the function words “to” and “and”.

Cognitive Integration Milestone VI (~ 20+ years). This longer term milestone refers to robotics experiments that can demonstrate the acquisition of open repertoires of compositional actions and lexicons sharing natural language properties. This could include emergent syntactic properties such as morphology, tense and case agreement.

VIII. CONCLUSION

Overall, our vision for cognitive robotics research on action and language integration within the social learning context proposes the combination of a developmental approach to embodied machine learning with usage-based models of natural language acquisition (Tomasello 2003) and construction-based theories of grammar (Goldberg 1995, 2006; Langacker 2008). In this, it subscribes to basic tenets of cognitive-linguistic theories of child language acquisition such as the assumption that language learning

- does not require substantial innate grammatical and

sensorimotor hardwiring;
- is grounded in recurrent patterns of embodied experience and situated social interaction;
- builds on a set of pre-acquired social cognitive capabilities that are required for cooperative ostensive-inferential communication in general;
- proceeds through tacit distributional analysis of a noisy but also richly structured linguistic input.

In order to implement these assumptions in a concrete agenda that can serve as an experimental roadmap and testbed for pertinent developmental research, we proposed that three key scientific challenges must be met:

- the development of scalable language processing and learning architectures that can (in principle) handle the full combinatorial complexity of natural language;
- the development of suitable implementations of basic social cognitive prerequisites for language acquisition as identified by experimental research in developmental psychology;
- the development of empirically substantiated characterizations of the actual learning target and its stepwise appropriation by the learner as determined by empirical research on child language acquisition.

Consistently with the above developmental principles, in this paper we have identified a series of core research challenges in the different areas of action, language and social learning, as well as challenges regarding their integration leading to the bootstrap of further cognitive and linguistic capabilities. These principles have been translated in a practical roadmap based on a series of research milestones within the next 20 year perspective. These milestones provide a possible set of goals and test-scenarios, thus acting as a research roadmap for future work on cognitive robotics. Although we do not propose that these milestones to be a rigid set of fully defined and fully sequential research goals, they can however provide operational definitions of research objectives for the next two decades of research. This milestone list, together with other proposals on language development stages (see for example Steels, 2005b, grammaticalization stages), can contribute to the evaluation of advances for future developmental cognitive robotics research (e.g. Cangelosi et al., 2008).

ACKNOWLEDGEMENTS

The author would like to acknowledge the contribution to this paper to the whole team of the ITALK project. In particular, we wish to thank the following people for comments and contributions to some of the sections in the paper: Davide Marocco, Francesco Nori, Vadim Tikhanoff, Alessandra Sciutti, Arjan Gijsberts, Karola Pitsch, Lars Schillingmann, Marco Mirolli, Caroline Lyon, Frank Foerster and Yo Sato.

REFERENCES

[1] Abbot-Smith K., and Tomasello M., “Exemplar-learning and schematization in a usage-based account of syntactic acquisition,” *The Linguistic Review*, vol. 23, pp. 275-290, 2006.

[2] Abbot-Smith K., Lieven E.V.M., and Tomasello M., “Graded representations in the acquisition of English and German transitive constructions,” *Cognitive Development*, vol. 23, pp. 48-66, 2008.

[3] Adriaans P., “Learning shallow context-free languages under simple distributions,” in *Algebras, Diagrams and Decisions in Language, Logic and Computation*, Copestake et al., Eds. CSLI, 2001.

[4] Akhtar N., “Acquiring word order: Evidence for data-driven learning of syntactic structure,” *Journal of Child Language*, vol. 26(2), pp. 339-356, 1999.

[5] Alishahi A., and Stevenson S., “A computational model of early argument structure acquisition,” *Cognitive Science*, vol. 32, pp. 789-834, 2008.

[6] Andrews M., Vigliocco G., Vinson D.P., “Integrating experiential and distributional data to learn semantic representations,” *Psychological Review*, vol. 116(3), pp. 463-498, 2009.

[7] Arbib M.A., Billard A., Iacoboni M., and Oztop E., “Synthetic brain imaging: grasping, mirror neurons and imitation,” *Neural Networks*, vol. 13, pp. 975-997, 2000.

[8] Auvray M., Lenay C. and Stewar J., “The attribution of intentionality in a simulated environment: the case of minimalist devices,” in *Proceedings of the 10th Meeting of the Association for the Scientific Study of Consciousness*, UK, June, 2006.

[9] Bahrick L. E., Lickliter R. and Flom R., “Intersensory redundancy guides the development of selective attention, perception and cognition in infancy,” *Current Directions in Psychological Science*, vol. 13, pp. 99-102, 2004.

[10] Balaban M. T., and Waxman S. R. “Do words facilitate object categorization in nine-month-old infants?” *Journal of Experimental Child Psychology*, vol. 64, pp. 3-26, 1997.

[11] Barsalou L. W., “Perceptual symbol systems,” *Behavioral and Brain Sciences*, vol. 22(4), pp. 577-660, 1999.

[12] Batali J., “The negotiation and acquisition of recursive grammars as a result of competition among exemplars,” in *Linguistic Evolution through Language Acquisition: Formal and Computational Models*. Briscoe, T. Ed., Cambridge: Cambridge University Press, 2002.

[13] Beer R.D., “The dynamics of active categorical perception in an evolved model agent,” *Adaptive Behavior*, vol. 11, pp. 209–243, 2003.

[14] Behne T., Carpenter M. and Tomasello M., “One-year-olds comprehend the communicative intentions behind gestures in a hiding game,” *Developmental Science*, vol. 8, pp. 492-499, 2005b.

[15] Behne T., Carpenter M., Call J., and Tomasello M., “Unwilling versus unable: Infants' understanding of intentional action,” *Developmental Psychology*, 41, pp. 328-337, 2005a.

[16] Belpaeme T., and Bleys J., “Explaining universal color categories through a constrained acquisition process,” *Adaptive Behavior*, vol. 13(4), pp. 293-310, 2005.

[17] Belpaeme T., and Cowley S. J., “Extended Symbol Grounding,” *Interaction Studies*, vol. 8(1), pp. 1-6, 2007.

[18] Bengio Y., and LeCun Y., “Scaling learning algorithms towards AI,” in *Large-Scale Kernel Machines*, L. Bottou, O. Chapelle, D. DeCoste and J. Weston, Eds., Cambridge, MA, USA: MIT Press, 2007.

[19] Bergen B., and Chang N., “Embodied Construction Grammar in Simulation-Based Language Understanding,” in *Construction Grammars: Cognitive grounding and theoretical extensions*, J-O Östman, and M. Fried, Eds., John Benjamins, 2005.

[20] Bigelow A.E., and Birch S.A.J., “The effects of contingency in previous interactions on infants' preference for social partners,” *Infant Behavior and Development*, vol. 22(3), pp. 367-382, 1999.

[21] Billard A., Epars Y., Calinon S., Cheng G., and Schaal S., “Discovering Optimal Imitation Strategies,” *Robotics and Autonomous Systems*, (Special Issue: Robot Learning from Demonstration) vol. 47(2-3), pp. 69-77, 2004.

[22] Borensztajn G., Zuidema J., and Bod R.. “Children’s grammars grow more abstract with age – Evidence from an automatic procedure for identifying the productive units of language,” in *Proceedings CogSci 2008*, Washington D.C, 2008.

[23] Boroditsky L., "Does language shape thought? English and Mandarin speakers' conceptions of time," *Cognitive Psychology*, vol. 43, pp. 1-22, 2001.

[24] Borroni P., Montagna M., et al., "Cyclic time course of motor excitability modulation during the observation of a cyclic hand movement," *Brain Research*, vol. 1065, pp. 115-124, 2005.

[25] Bowerman M., and Levinson S., *Language Acquisition and Conceptual Development*. Cambridge: Cambridge University Press, 2001.

[26] Brand R. J., and Baldwin D. A., Paper presented at the *X. International Congress for Studies in Child Language (IASCL)*, Berlin, Germany, 2005.

[27] Brand R. J., and Tapscott S., "Acoustic packaging of action sequences by infants," *Infancy*, vol. 12(1), pp. 321-332, 2007.

[28] Broz F., Kose-Bagci H., Nehaniv C.L., and Dautenhahn K., "Learning behavior for a social interaction game with a childlike humanoid robot", in *Social Learning in Interactive Scenarios Workshop*, *Humanoids 2009*, Paris, France, 7 December, 2009.

[29] Bybee J., "From usage to grammar: the mind's response to repetition," *Language*, vol. 82, 4, pp. 711-733, 2006.

[30] Camaioni L., Caselli M. C., Longbardi E., and Volterra V., "A parent report instrument for early language assessment," *First Language*, vol. 11, pp. 345-360, 1991.

[31] Cameron-Faulkner T., Lieven E.V., and Tomasello M., "A construction based analysis of child directed speech," *Cognitive Science*, vol. 27, pp. 843-873, 2003.

[32] Cangelosi A., Belpaeme T., Sandini G., Metta G., Fadiga L., Sagerer G., Rohlfing K., Wrede B., Nolfi S., Parisi D., Nehaniv C., Dautenhahn K., Saunders J., Fischer K., Tani J. and Roy D., "The ITALK project: Integration and transfer of action and language knowledge," *Proceedings of Third ACM/IEEE International Conference on Human Robot Interaction (HRI 2008)*, Amsterdam, 12-15 March 2008. .

[33] Cangelosi A., and Harnad S., "The adaptive advantage of symbolic theft over sensorimotor toil: Grounding language in perceptual categories," *Evolution of Communication*, vol. 4(1), pp. 117-142, 2000.

[34] Cangelosi A., and Parisi D., "The emergence of a language in an evolving population of neural networks," *Connection Science*, vol. 10(2), pp. 83-97, 1998.

[35] Cangelosi A., and Parisi D. (Eds.), *Simulating the Evolution of Language*. London: Springer, 2002.

[36] Cangelosi A., and Parisi D., "The processing of verbs and nouns in neural networks: Insights from synthetic brain imaging," *Brain and Language*, vol. 9(2), pp. 401-4082004.

[37] Cangelosi A., and Riga T., "An embodied model for sensorimotor grounding and grounding transfer: Experiments with epigenetic robots," *Cognitive Science*, vol. (4), pp. 673-689, 2006.

[38] Cangelosi A., Bugmann G., and Borisyuk R. (Eds.), *Modeling Language, Cognition and Action.* Singapore: World Scientific, 2005.

[39] Cappa S.F., and Perani D., "The neural correlates of noun and verb processing," *Journal of Neurolinguistics*, vol. 16 (2-3), pp. 183-1892003.

[40] Carpenter M., Akhtar N., and Tomasello M., "Sixteen-month-old infants differentially imitate intentional and accidental actions," *Infant Behavior and Development*, vol. 21, pp. 315-30, 1998b.

[41] Carpenter M., Nagell K., and Tomasello M., "Social cognition, joint attention, and communicative competence from 9 to 15 months of age," *Monographs of the Society for Research in Child Development*, vol. 63, 4, 1998a.

[42] Cartwright T.A., and Brent M.R., "Syntactic categorization in early language acquisition: Formalizing the role of distributional analysis," *Cognition*, vol. 63, pp. 121-170, 1997.

[43] Chakravarthy V. S., and Kompella B., "The shape of handwritten characters," *Pattern Recognition Letters*, vol. 24(12), pp, 1901-1913, 2003.

[44] Chiel H.J., and Beer R.D., "The brain has a body: Adaptive behavior emerges from interactions of nervous system, body and environment," *Trends in Neurosciences*, vol. 20, pp. 553-557, 1997.

[45] Choi S., McDonough L., Bowerman M., and Mandler J. M., "Early sensitivity to language-specific spatial categories in English and Korean," *Cognitive Development*, vol. 14(2), pp. 241-268, 1999.

[46] Choi S., "The semantic development of negation: a cross linguistic longitudinal study," *Journal of Child Language*, 15(3), pp. 517-531, 1988.

[47] Clark A., "Reasons, robots and the extended mind," *Mind and Language*, vol. 16(2), pp. 121–145, 2001.

[48] Clark A., "Grammatical inference and first language acquisition," in *Workshop on Psycho-computational Models of Human Language Acquisition*, Geneva, Switzerland, 2004.

[49] Clark H., *Arenas of Language Use*. Chicago: University of Chicago Press, 1992.

[50] Corballis M.C., *From Hand to Mouth: The Origins of Language*. Princeton University Press, 2002.

[51] Cowley S.J., "Robots – the new linguistic informants?" *Connection Science*, vol. 20(4), pp. 349-369, 2008.

[52] Craighero L., Fadiga L., et al., "Movement for perception: a motor-visual attentional effect," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 25, pp. 1673-1692, 1999.

[53] Csibra G., and Geregly G., "Social learning and social cognition: The case for pedagogy," in *Attention and Performance XXI*, M.H. Johnson and Y. Munakata, Eds., Oxford: Oxford University Press, 2006, pp. 249-274.

[54] Dabrowska E., *Language, Mind and Brain: Some Psychological and Neurological Constraints on Theories of Grammar*. Edinburgh: Edinburgh University Press, 2004.

[55] Damasio A. R., and Tranel D., "Nouns and verbs are retrieved with differently distributed neural systems," *Proceedings of the National Academy of Sciences*, vol. 90, pp. 4957–4960, 1993.

[56] Dausendschön-Gay U., "Producing and learning to produce utterances in social interaction," *Eurosla Yearbook*, vol. 3, pp. 207-228, 2003.

[57] de León L., "The emergent participant: Interactive patterns in the socialization of Tzotzil (Mayan) infants," *Journal of Linguistic Anthropology*, vol. 8, pp. 131-161, 2000.

[58] de Vries J. I. P., Visser G. H. A., and Prechtl H. F. R., "The emergence of fetal behaviour. i. qualitative aspects," *Early Human Development*, vol. 23, pp. 159-191, 1982.

[59] Deacon T.W., *The Symbolic Species: The Coevolution of Language and Human Brain*. London: Penguin, 1997.

[60] Demiris Y., and Khadhouri B., "Hierarchical attentive multiple models for execution and recognition (HAMMER)," *Robotics and Autonomous Systems*, vol. 54, pp. 361-369, 2006.

[61] Di Paolo E.A., Rohde M., and Iizuka H., "Sensitivity to social contingency or stability of interaction? Modeling the dynamics of perceptual crossing," *New Ideas in Psychology* (Special Issue on Dynamics and Psychology), vol. 26, pp. 278-294, 2008.

[62] Diamond A., "Retrieval of an object from an open box: The development of visual-tactile control of reaching in the first year of life," *Society of Research in Child Development Abstracts*, vol. 78(3), 1981.

[63] Dominey P.F., Mallet A., and Yoshida E., "Real-time spoken-language programming for cooperative interaction with a humanoid apprentice," *International Journal of Humanoid Robotics*, vol. 6 (2), pp. 147-171, 2009.

[64] Dominey P. F., and Warneken F., "The basis of shared intentions in human and robot cognition," *New Ideas in Psychology*, in press, 2009.

[65] Dominey P.F. and Dodane C., "Indeterminacy in language acquisition: the role of child directed speech and joint attention," *Journal of Neurolinguistics, 17*, vol. 2-3: pp. 121-145, 2004.

[66] Elman J., Computational approaches to language acquisition. In K. Brown, ed. *Encyclopedia of Language and Linguistics*, Second Edition, vol. 2,. Oxford: Elsevier, 2006, pp. 726-732.

[67] Fadiga L., Fogassi L., Gallese V., and Rizzolatti G., "Visuomotor neurons: ambiguity of the discharge or 'motor' perception?" *International Journal of Psychophysiology*, vol. 35(2-3), pp. 165-177, 2000.

[68] Fadiga L., Craighero L., et al., "Speech listening specifically modulates the excitability of tongue muscles: a TMS study," *European Journal of Neuroscience*, vol. 15(2), pp. 399-402, 2002.

[69] Farroni T., Johnson M.H., and Csibra G., "Mechanisms of eye gaze perception during infancy," *Journal of Cognitive Neuroscience*, vol. 16(8), pp. 1320-1326, 2004.

[70] Fogassi L., Ferrari P. F., Gesierich B., Rozzi S., Chersi F., and Rizzolatti, G., "Parietal lobe: From action organization to intention understanding," *Science*, 308(4), pp. 662-667, 2005.

[71] Fogassi L., Gallese V., Fadiga L., and Rizzolatti G., "Neurons responding to the sight of goal directed hand/arm actions in the parietal

area PF (7b) of the macaque monkey," *Society of Neuroscience Abstracts*, vol. 24, pp. 257.255, 1998.

[72] Fogassi L., Gallese V., et al., "Coding of peripersonal space in inferior premotor cortex (area F4)," *Journal of Neurophysiology*, vol. 76, pp. 141-157, 1996.

[73] Fogel A., and Garvey A., "Alive communication," *Infant Behavior and Development* vol. 30, pp. 251-257, 2007.

[74] Förster F., Nehaniv C. L., and Saunders J., "Robots that Say 'No'," in *Proceedings of European Conference on Artificial Life 2009 (ECAL 2009)*, Springer Lecture Notes in Artificial Intelligence, in press..

[75] Fulop S., "Semantic bootstrapping in type-logical grammar," *Journal of Logic, Language and Information*, vol. 14, pp. 49–862005.

[76] Gallese V., Fadiga L., Fogassi L., and Rizzolatti G., "Action recognition in the premotor cortex," *Brain*, vol. 19(2), pp. 593-609, 1996.

[77] Gardenfors P., (2002). *Conceptual Spaces: The Geometry of Thought*, MIT Press..

[78] Georgopoulos A.P., Kalaska J. F., Caminiti R., and Massey J.T., "On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex," *Journal of Neuroscience*, vol. 2(11), pp. 1527-37, 1982.

[79] Gertner Y., Baillargeon R., Fisher C., and Simons J. D., "Language facilitates Infants' Physical Reasoning", Paper presented at the *Biennial Meeting of the Society of Research in Child Development*, Denver, USA, 2009.

[80] Gibson J. J., "The theory of affordances," In R. Shaw and J. Bransford (Eds.), *Perceiving, acting and knowing: toward an ecological psychology*, Hillsdale: Lawrence Erlbaum, 1977, pp. 67-82.

[81] Giese M., and Poggio T., "Neural mechanisms for the recognition of biological movements and action," *Nature Review Neuroscience*, vol. 4, pp. 179–192, 2003.

[82] Gigliotta O., and Nolfi S., "On the coupling between agent internal and agent/environmental dynamics: Development of spatial representations in evolving autonomous robots," *Adaptive Behavior*, vol. 16, pp. 148-165, 2008.

[83] Gilbert A., Regier T., Kay P., and Ivry R., "Whorf hypothesis is supported in the right visual field but not the left," *Proceedings of the National Academy of Sciences*, vol. 103(2), pp. 489-494, 2006.

[84] Glenberg A.M., and Gallese V., "Action-based Language: A theory of language acquisition, comprehension, and production," in preparation.

[85] Glenberg A.M., and Kaschak M., "Grounding language in action," *Psychonomic Bulletin and Review*, vol. 9(3), pp. 558-565, 2002.

[86] Glenberg A.M., and Robertson D.A., "Symbol grounding and meaning: A comparison of high-dimensional and embodied theories of meaning," *Journal of Memory and Language*, vol. 43(3), pp. 379-401, 2000.

[87] Gogate L. J., and Bahrick L. E., "Intersensory redundancy and 7-month-old infants' memory for arbitrary syllable-object relations," *Infancy*, vol. 2, pp. 219-231, 2001.

[88] Gold E. M., "Language identification in the limit," *Information and Control*, vol. 10(5), pp. 447 – 474, 1967.

[89] Goldberg A.E., *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press, 2006.

[90] Goldberg A.E., Casenhiser D.M., and Sethuraman N., "Learning argument structure generalizations," *Cognitive Linguistics*, 15(3), pp. 289-316, 2004.

[91] Goldin-Meadow S., and Butcher C., "Pointing towards two-word speech in young children," in *Pointing: where language, culture, and cognition meet*, S. Kita, Ed. Mahwah, NJ.: Erlbaum Ass, 2003, pp. 85-187.

[92] Gómez R., "Statistical learning in infant language development," in *The Oxford Handbook of Psycholinguistics*, M. G. Gaskell, Ed. Oxford: Oxford University Press. 2007, pp. 601-615.

[93] Goodwin C., "Action and embodiment within situated human interaction," *Journal of Pragmatics*, vol. 32, pp. 1489-1522, 2000.

[94] Graziano M.S.A., X. Hu, et al., "Visuo-spatial properties of ventral premotor cortex," *Journal of Neurophysiology*, vol. 77, pp. 2268-2292, 1997.

[95] Greenfield P. M., "Language, tools, and brain: The ontogeny and phylogeny of hierarchically organized sequential behavior," *Behavioral and Brain Sciences*, vol. 14(4), pp. 531-551, 1991.

[96] Gumperz J. J., and Levinson S. C., "Rethinking linguistic relativity," *Current Anthropology*, vol. 32, pp. 613-623, 1997.

[97] Harnad S., "The symbol grounding problem," *Physica D*, vol. 42, pp. 335-346, 1990.

[98] Haruno M., Wolpert D. M., and Kawato M., "MOSAIC Model for sensorimotor learning and control," *Neural Computation*, vol. 13, pp. 2201-2220, 2001.

[99] Hauk O., Johnsrude I., and Pulvermuller F., "Somatotopic representation of action words in human motor and premotor cortex," *Neuron*, vol. 41(2), pp. 301-307, 2004.

[100] Hayashi M., and Matsuzawa T., "Cognitive development in object manipulation by infant chimpanzees," *Animal Cognition*, vol. 6(4), pp. 225-233, 2003.

[101] Hinton G. E., and Nair V., "Inferring motor programs from images of handwritten digits," *Advances in Neural Information Processing Systems, 18*, MIT Press, 2006.

[102] Hinton G. E., Osindero S., and Teh Y.-W., "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18(7), 1527–1554..

[103] Hirsh-Pasek K., and Golinkoff R. M. (1996). *The Origins of Grammar: Evidence from Early Language Comprehension*. Cambridge, MA: MIT Press, 2006.

[104] Horwitz B., Tagamets M.-A., and McIntosh A.R., "Neural modeling, functional brain imaging, and cognition," *Trends in Cognitive Science*, vol. 3, pp. 91-98, 1999.

[105] Hubel D. H., and Wiesel T.N., "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *Journal of Physiology*, vol. 160, pp. 106–154, 1962.

[106] Ito M., Noda K., Hoshino Y., and J. Tani, "Dynamic and interactive generation of object handling behaviors by a small humanoid robot using a dynamic neural network model," *Neural Networks*, vol. 19, pp. 323-337, 2006b.

[107] Ito M., Noda K., et al., "Dynamic and interactive generation of object handling behaviors by a small humanoid robot using a dynamic neural network model," *Neural Networks*, vol. 19(2), pp. 323-337, 2006a.

[108] Jeannerod M., *The Cognitive Neuroscience of Action*. Cambrige, MA and Oxford UK, Blackwell Publishers Inc, 1997.

[109] Jhuang H., Serre T., Wolf L., and Poggio T., "A biologically inspired system for action recognition," in *Proceedings of the Eleventh IEEE International Conference on Computer Vision (ICCV)*, 2007.

[110] Johnson, M. H., *Developmental Cognitive Neuroscience (3 ed. Vol. 1)*. Malden, MA and Oxford UK: Balckwell Publisher Inc, 1997.

[111] Just M.A., Carpenter P.A., and Varma S., "Computational modeling of high-level cognition and brain function," *Human Brain Mapping*, vol. 8, 128-136, 1999.

[112] Kaplan F., Oudeyer P-Y. and Bergen B., "Computational models in the debate over language learnability," *Infant and Child Development* vol. 17, 1, pp. 55-802008.

[113] Karmiloff K. and Karmiloff-Smith A., *Pathways to language: From foetus to adolescent. Developing Child Series*, Harvard University Press, 2001.

[114] Keijzer F., "Representation and behavior," London, UK: MIT Press, 2001.

[115] Kempson R.M., Meyer-Viol W., and Gabbay D.M., *Dynamic Syntax: The Flow of Language Understanding*. Blackwell, 2001.

[116] Kindermann T. A., "Fostering independence in mother-child interactions: longitudinal changes in contingency patterns as children grow competent in developmental tasks," *International Journal of Behavioral Development* vol. 16(4), pp. 513-535, 1993.

[117] Kose-Bagci H., Broz F., Shen Q., Dautenhahn K., and Nehaniv C.L., "As time goes by: representing and reasoning timing in the human-robot interaction studies," in *AAAI - Spring Symposium 2010: It's All in the Timing: Representing and Reasoning About Time in Interactive Behavior*, Stanford University, Palo Alto, California, 2010.

[118] Kose-Bagci H., Dautenhahn K., Syrdal D.S., and Nehaniv C.L., "Drum-mate: Interaction dynamics and gestures in human-humanoid drumming experiments," *Connection Science*. in press.

[119] Kousta S., Vigliocco G., Vinson D., and Andrews M., "Happiness is… an abstract word: The role of affect in abstract knowledge representation," *Proceedings of the 31st Meeting of the Cognitive Science Society*2009.

[120] Krunic V., Salvi G., Bernardino A., Montesano L., and Santos-Victor J., "Affordance based word-to-meaning association," Paper presented at the *International Conference on Robotics and Automation 2009* (ICRA 2009), Kobe, Japan, 2009.

[121]Küntay A., and Slobin D.I., “Listening to a Turkish mother: Some puzzles for acquisition,” in *Social Interaction, Social Context, and Language. Essays in Honor of Susan Ervin-Tripp*, D. Slobin et al. Eds. Mahwah, N.J.: Erlbaum, 1996, pp. 265-286.

[122]Lackner J. R., and DiZio P., “Adaptation in a rotating artificial gravity environment,” *Brain Research Reviews*, vol. 28, pp. 194-202, 1998.

[123]Lakoff G., *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. Chicago: University of Chicago Press, 1987.

[124]Langacker R. W., *Foundations of Cognitive Grammar*. Stanford, Calif.: Stanford University Press, 1987.

[125]LeCun Y., Matan O., Boser B., Denker J. S., Henderson D., Howard R. E., Hubbard W., Jackel L. D., and Baird H. S., “Handwritten zip code recognition with multilayer networks,” in *Proc. of the International Conference on Pattern Recognition*, vol. II, IEEE, Atlantic City, 1990, pp. 35–40.

[126]Liberman A. M., and Mattingly I.G., “The motor theory of speech perception revised,” *Cognition*, vol. 21(1), pp. 1-36, 1985.

[127]Liszkowski U., “Human twelve-month-olds point co-operatively to share interest with and provide information for a communicative partner,” *Gesture*, vol. 5, pp. 135-154, 2005.

[128]Liszkowski U., “Infant pointing at twelve months: Communicative goals, motives, and social-cognitive abilities,” in *The roots of human sociality: Culture, cognition, and interaction*. N. Enfield and S. Levinson, Eds. Oxford: Berg, 2006, pp. 153-178.

[129]Locke J. L., “Bimodal signaling in infancy: Motor behavior, reference, and the evolution of spoken language,” *Interaction Studies*, vol. 8(1), pp. 159-175, 2007.

[130]Lungarella M., Metta G., Pfeifer R. and Sandini G., “Developmental robotics: A survey,” *Connection Science*, vol. 15(4), pp. 151-1902003.

[131]Lupyan G., Rakison D.H., and McClelland J.L., “Language is not just for talking: labels facilitate learning of novel categories,” *Psychological Science*, vol. 18(12), pp. 1077-1083, 2007.

[132]MacWhinney B., “The emergence of linguistic form in time,” *Connection Science*, vol. 17(3-4), pp. 191-211, 2005.

[133]Majid A., Bowerman M., Kita S., Haun D. B. M., and Levinson S. C., “Can language restructure cognition? The case for space,” *Trends in Cognitive Sciences*, vol. 8(3), pp. 108-114, 2004.

[134]Mareschal D., Johnson M., Sirios S., Spratling M., Thomas M. S. C., and Westermann G., *Neuroconstructivism Volume 1: How the brain constructs cognition*. Oxford: Oxford University Press, 2007a.

[135]Mareschal D., Sirios S., and Westermann G., *Neuroconstructivism Volume 2: Perspectives and Prospects*. Oxford: Oxford University Press, 2007b.

[136]Markman E.M., *Categorization and Naming in Children*. Cambridge, MA: MIT Press, 1989.

[137]Markova G., and Legerstee M., “Contingency, imitation, and affect sharing: Foundations of infants' social awareness,” *Developmental Psychology*, vol. 42, pp. 132-141, 2006.

[138]Martin A., Haxby J.V., Lalonde F.M., Wiggs C.L., and Ungerleider L.G., “Discrete cortical regions associated with knowledge of color and knowledge of action,” *Science*, vol. 270, pp. 102-105, 1995.

[139]McClure C., Pine J., and Lieven E. “Investigating the abstractness of children's early knowledge of argument structure,” *Journal of Child Language* vol. 33, 4, pp. 693-720, 2006.

[140]Mervis C. B., and Bertrand J., “Acquisition of the novel name--nameless category (N3C) principle,” *Child Development*, 65(6), pp. 1646-1662, 1994.

[141]Metta G., and Fitzpatrick P., “Early Integration of Vision and Manipulation,” *Adaptive Behavior*, vol. 11(2), pp. 109-128, 2003.

[142]Metta G., Sandini G., Natale L., Craighero L., and Fadiga L., “Understanding mirror neurons: a bio-robotic approach,” *Interaction Studies*, special issue on Epigenetic Robotics, vol. 7(2), pp. 197–232, 2006.

[143]Meyers E., and Wolf L., “Using Biologically Inspired Visual Features for Face Processing,” *International Journal of Computer Vision*, vol. 76(1), pp. 93–104, 2008.

[144]R.G. Millikan, *Varieties of Meaning*. MIT Press, 2004.

[145]Mirolli M., and Parisi D., “Towards a Vygotskyan cognitive robotics: the role of language as a cognitive tool,” *New Ideas in Psychology* in press.

[146]Mnih A., and Hinton G. E., “A scalable hierarchical distributed language model,” *Advances in Neural Information Processing Systems, 21*, MIT Press, Cambridge, MA, 2009.

[147]Moore C., Angelopoulos M., and Bennett P., “The role of movement in the development of joint visual attention,” *Infant Behavior and Development*, vol. 20(1), pp. 83-92, 1997.

[148]Morris W.C., Cottrell G.W., and Elman J., “A connectionist simulation of the empirical acquisition of grammatical relations,” in Wermter, S. and Sun, R. eds. *Hybrid Neural Symbolic Integration*. Berlin: Springer, pp. 175-193, 2000.

[149]Movellan J. R., “An infomax controller for real time detection of social contingency,” in *Proceedings of 2005 4th IEEE International Conference on Development and Learning*, pp. 19-24, 2005.

[150]Muir D., and Lee K., “The still-face effect: methodological issues and new applications,” *Journal of Infancy*. vol. 4(4), pp. 483-491, 2003.

[151]Murata A., Fadiga L., Fogassi L., Gallese V., Raos V., Rizzolatti G., “Object representation in the ventral premotor cortex (area F5) of the monkey,” *Journal Neurophysiology* vol. 78(4), pp. 2226-30, 1997.

[152]Mussa-Ivaldi F. A., and Bizzi E., “Motor Learning through the Combination of Primitives,” *Philosophical Transaction of the Royal Society B: Biological Sciences*, vol. 355(1404), 1755-1769, 2000.

[153]Mussa-Ivaldi F. A., and Giszter S. F., “Vector field approximation: a computational paradigm for motor control and learning,” *Biological Cybernetics*, vol. 67, pp. 491-500, 1992.

[154]Nakayama Y., Yamagata T., Tanji J., and Hoshi E., “Transformation of a virtual action plan into a motor plan in the premotor cortex,” *Journal of Neuroscience*, vol. 28(41), pp. 10287-10297, 2008.

[155]Nehaniv C. L., and Dautenhahn K. (Eds.), *Imitation and Social Learning in Robots, Humans and Animals: Behavioural, Social and Communicative Dimensions*, Cambridge University Press, 2007.

[156]Nehaniv C. L., and Dautenhahn, K., “Of hummingbirds and helicopters: An algebraic framework for interdisciplinary studies of imitation and its applications,” in *Interdisciplinary Approaches to Robot Learning*, Y. Demiris and A. Birk, Eds., World Scientific Series in Robotics and Intelligent Systems, 2000.

[157]Nehaniv C.L., Lyon C., and Cangelosi A., “Current work and open problems: A road-map for research into the emergence of communication and language,” in *Emergence of Communication and Language*, C. Lyon, C. L. Nehaniv, and A. Cangelosi, Eds., Springer, 2007, pp. 1-27.

[158]Ninio A., “Testing the role of semantic similarity in syntactic development,” *Journal of Child Language*, vol. 32, pp. 35-61, 2005a.

[159]Ninio A., “Accelerated learning without semantic similarity: indirect objects,” *Cognitive Linguistics*, vol. 16, 3, pp. 531-556, 2005b.

[160]Nolfi S., “Categories formation in self-organizing embodied agents,” in *Handbook of Categorization in Cognitive Science*. H. Cohen and C. Lefebvre, Eds., Elsevier. 2005, pp. 869-889.

[161]Nolfi S., and Floreano D., *Evolutionary Robotics: The Biology, Intelligence, and Technology of Self-Organizing Machines*. Cambridge, MA: MIT Press/Bradford Books, 2000.

[162]Nolfi S. and Marocco D., “Active perception: A sensorimotor account of object categorization,” In *From Animals to Animats 7: Proceedings of the VII International Conference on Simulation of Adaptive Behavior*, B. Hallam et al., Eds., Cambridge, MA: MIT Press, 2002, pp. 266-271.

[163]Nolfi S., and Tani J., “Extracting regularities in space and time through a cascade of prediction networks: The case of a mobile robot navigating in a structured environment,” *Connection Science*, vol. 11(2), pp. 129-152, 1999.

[164]Nori F., Sandini G., and Konczak J., “Can imprecise internal motor models explain the ataxic hand trajectories during reaching in young infants?” in *Ninth International Conference on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems (EPIROB)*, Venice, Italy, 2009.

[165]O’Regan J.K., and Noë A., “A sensorimotor account of vision and visual consciousness,” *Behavioral and Brain Sciences*, 5, 2001.

[166]Okanda M., and Itakura S., “Development of contingency: How infants become sensitive to contingency?” in *Proceedings of the XVth Biennial International Conference on Infant Studies*, Kyoto, Japan, 2006.

[167]Oudeyer P-Y., *Self-organization in the Evolution of Speech*, Oxford University Press, 2006..

[168]Oudeyer .-Y., Kaplan F. “Discovering communication,” *Connection Science*, 18(2), pp. 189-206, 2006.

[169]Oudeyer P-Y., Kaplan F., and Hafner V., "Intrinsic motivation systems for autonomous mental development," *IEEE Transactions on Evolutionary Computation*, vol. 11, pp. 265–286, 2007.

[170]Oztop E., Kawato M., et al. (2006). "Mirror neurons and imitation: A computationally guided review," *Neural Networks*, vol. 19(3), pp. 254-271..

[171]Pecher D., and Zwaan R.A., *Grounding Cognition: The Role of Perception and Action in Memory, Language, and Thinking*. Cambridge University Press, 2005.

[172]Perani D., Cappa S. F., Schnur T., Tettamanti M., Collina S., Rosa M. M., and Fazio F., "The neural correlates of verb and noun processing: A PET study," *Brain*, vol. 122, pp. 2337-44, 1999.

[173]Pfeifer R., and Scheier C., *Understanding Intelligence*, Cambridge, MA: MIT Press, 1999.

[174]Piaget J., *The Construction of Reality in the Child*, Ballentine, 1954.

[175]Pine J. M., "The language of primary caregivers". in *Input and Interaction in Language Acquisition*, Gallaway, C. and Richards, B.J. Eds. Cambridge: Cambridge University Press, 1994, pp. 15-37..

[176]Pinker S., *Language learnability and language development. Cambridge*, Mass.: Harvard University Press, 1984.

[177]Pinker S., *The stuff of thought: language as a window into human nature*. New York: Viking, 2007.

[178]Plunkett K., Hu J.F., and Cohen L.B., "Labels can override perceptual categories in early infancy," *Cognition*, vol. 106, 665, 2008.

[179]Ponce J. (Ed.), *Toward category-level object recognition*. Berlin, New York: Springer, 2006.

[180]Preissl H., Pulvermüller F., Lutzenberger W., and Birbaumer N. "Evoked potentials distinguish between nouns and verbs," *Neuroscience Letters*, vol. 197, pp. 81-83, 1995.

[181]Prince C. G., Hollich G. J., Helder G. J., Mislivec E. J., Reddy A., Salunke S. and Memon N., "Taking synchrony seriously: A perceptual-level model of infant synchrony detection," in *Proceedings of the Fourth International Workshop on Epigenetic Robotics*, 2004, pp. 89–96.

[182]Pulvermuller F., *The Neuroscience of Language. On Brain Circuits of Words and Serial Order*, Cambridge University Press, 2003.

[183]Pulvermuller F., Hauk O., Shtyrov Y., Johnsrude I., Nikulin V., and Ilmoniemi R., "Interactions of language and actions," *Psychophysiology*, vol. 40, 70, 2003.

[184]Pulvermuller F., Mohr B., Schleichert H. and Veit R. "Operant conditioning of left-hemispheric slow cortical potentials and its effect on word processing," *Biological Psychology*, vol. 53, pp. 177-215, 2000.

[185]Quartz S. R., and Sejnowski T. J., "The neural basis of cognitive development: A constructivist manifesto," *Behavioral and Brain Sciences*, vol. 20, pp. 537-596, 1997.

[186]Quine W., *Word and Object*. Cambridge M.A.: MIT Press, 1960.

[187]Rakison D.H., and Oakes L.M., *Early Category and Concept Development: Making Sense of the Blooming, Buzzing Confusion*. London: Oxford University Press, 2003.

[188]Rakoczy H, Warneken F., and Tomasello M. "The sources of normativity: Young children's awareness of the normative structure of games," *Developmental Psychology*, 44, 3, pp. 875 – 881, 2008.

[189]Rakoczy H., "Play, games, and the development of collective intentionality," *New Directions for Child and Adolescent Development*, vol. 115, pp. 53-67, 2007.

[190]Rizzolatti G., Fadiga L., Gallese V., and Fogassi L., "Premotor cortex and the recognition of motor actions," *Cognitive Brain Research*, vol. 3(2), pp. 131-41, 1996.

[191]Rizzolatti G., and Arbib M., "Language within our grasp," *Trends in Neuroscience*, vol. 21, pp. 188-194, 1998.

[192]Rizzolatti G. and Luppino G., "The cortical motor system," *Neuron*. vol. 31, pp. 889-901, 2001.

[193]Rizzolatti G., Camarda R., Fogassi L., Gentilucci M., Luppino G., and Matelli M., "Functional organization of inferior area 6 in the macaque monkey. II. Area F5 and the control of distal movements," *Experimental Brain Res*earch, vol. 71(3), pp. 491-507, 1988.

[194]Rizzolatti G., Fadiga L., et al., "The space around us," *Science*, vol. 277, pp. 190-191, 1997.

[195]Rizzolatti G., Fogassi L., et al., "Parietal cortex: from sight to action," *Current Opinion Neurobiology*, vol. 7(4), 562-567, 1997.

[196]Roberson D., Davidoff J., Davies I. R. L., and Shapiro L. R., "Color categories: Evidence for the cultural relativity hypothesis," *Cognitive Psychology*, vol. 50(4), pp. 378-411, 2005.

[197]Rohlfing K.J., Fritsch J., Wrede B. and Jungmann T., "How can multimodal cues from child-directed interaction reduce learning complexity in robots?" *Advanced Robotics*, vol. 20(10), pp. 1183-1199, 2006.

[198]Rolf M., Hanheide M., and Rohlfing K.J., "Attention via synchrony. Making use of multimodal cues in social learning," submitted.

[199]Roy A., Craighero L., et al., "Phonological and lexical motor facilitation during speech listening: A transcranial magnetic stimulation study," *Journal of Physiology - Paris* vol. 102, pp. 101-105, 2008.

[200]Roy D., "Grounding words in perception and action: insights from computational models," *Trends in Cognitive Science*, vol. 9(8), pp. 389-396, 2005a.

[201]Roy D., "Semiotic schemas: A framework for grounding language in action and perception," *Artificial Intelligence*, vol. 167(1-2), pp. 170-205, 2005b.

[202]Roy D., Hsiao K-Y., and Mavridis N., "Mental imagery for a conversational robot," *IEEE Transactions on System Man and Cybernetics, Part B Cybernetics* vol. 34, pp. 1374–1383, 2004.

[203]Sakata H., Taira M., et al., "Neural mechanisms of visual guidance of hand actions in the parietal cortex of the monkey," *Cerebral Cortex*, vol. 5(5), pp. 429-438, 1995.

[204]Sato Y and Saunders J., "Semantic bootstrapping in embodied robots," To appear in *Proceedings of EVOLANG*, Utrecht. forthcoming.

[205]Saunders J., Nehaniv C.L., and Lyon C., "Robot learning of lexical semantics from sensorimotor interaction and the unrestricted speech of human tutors".

[206]Saunders J., Lyon C., Forster F., Nehaniv C.L., and Dautenhahn K., "A constructivist approach to robot language learning via simulated babbling and holophrase extraction," in *Proc. 2nd International IEEE Symposium on Artificial Life*, Nashville, Tennessee, USA - 30 March-2 April 2009, 2009.

[207]Saunders J., Nehaniv C.L., and Dautenhahn K., "Experimental comparisons of observational learning mechanisms for movement imitation in mobile robots," *Interaction Studies*, vol. 8(2), pp. 307-3352007a.

[208]Saunders J., Nehaniv C.L., Dautenhahn K., and Alissandrakis A., "Self-imitation and environmental scaffolding for robot teaching," *International Journal of Advanced Robotic Systems*, vol. 4(1), pp. 109-124, 2007b.

[209]Saur D., Kreher B.W., Schnell S., Kümmerer D., Kellmeyer P., Vry M.-S., Umarova R., Musso M., Glauche V., Abel S., Huber W., Rijntjes M., Hennig J., and Weille C., "Ventral and dorsal pathways for language" *Proceedings of the National Academy of Science*, vol. 105(46), pp. 18035–18040, 2008.

[210]Schaal S., Atkeson C., and Vijayakumar S., "Real-time robot learning with locally weighted statistical learning," in *International Conference on Robotics and Automation*, San Francisco, 2000.

[211]Schegloff E. A., *Sequence Organisation in Interaction. A Primer in Conversation Analysis*. Cambridge University Press, 2007.

[212]Scheier C., Pfeifer R. and Kunyioshi Y., Embedded neural networks: exploiting constraints. *Neural Networks*, vol. 11, pp. 1551-1596, 1998.

[213]Seabra Lopes L., and Chauhan A., "How many words can my robot learn? An approach and experiments with one-class learning," *Interaction Studies*, vol. 8(1), pp. 53-81, 2007.

[214]Serre T., Wolf L., Bileschi S., Riesenhuber M., and Poggio T., "Robust object recognition with cortex-like mechanisms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29(3), pp. 411–426, 2007.

[215]Simmons G. and Demiris Y., "Object grasping using the minimum variance model," *Biological Cybernetics*, vol. 94(5), pp. 393-40, 2006.

[216]Sirois S., Spratling M., Thomas M. S. C., Westermann G., Mareschal D., and Johnson M. H., "Precis of neuroconstructivism: How the brain constructs cognition," *Behavioral and Brain Sciences*, vol. 31, pp. 321-356, 2008.

[217]Smith K., Brighton H., and Kirby S., "Complex systems in language evolution: the cultural emergence of compositional structure," *Advances in Complex Systems*, vol. 6(4), 537-558, 2003.

[218]Snow C.E. "Beginning from baby talk: Twenty years of research on input and interaction," in Gallaway, C. and Richards, B.J., eds. *Input*

and Interaction in Language Acquisition, 3-12. Cambridge: Cambridge University Press, 1994.

[219]Solan Z., Horn D., Ruppin E., and Edelman S., “Unsupervised learning of natural languages,” *Proceedings of the National Academy of Science*, vol. 102, pp. 11629-11634, 2005.

[220]Spelke E. S., “Core knowledge,” *American Psychologist*, vol. 55, pp. 1233-1243, 2000.

[221]Sperber D., and Wilson D., *Relevance: Communication and Cognition*. Oxford, Blackwell, 1995.

[222]Sporns O., “What neuro-robotic models can tell us about neural and cognitive development,” in *Neuroconstructivism: Perspectives and Prospects*, Mareschal, D., Sirois, S., Westermann, G. and Johnson, M.H., Eds., Oxford University Press, Oxford, UK, 2007, pp. 179-204.

[223]Steels L., “The origins of syntax in visually grounded robotic agents,” *Artificial Intelligence*, vol. 103 (1-2), 133-156, 1998.

[224]Steels L., “Constructivist development of grounded construction grammars,” in *Proc. Ann. Meeting Assoc. for Computational Linguistics Conf*, Scott, D., Daelemans, W. and Walker, M. eds., Barcelona: ACL, 2004, pp. 9-16.

[225]Steels L., “The role of Construction Grammar in Fluid Language Grounding,” Unpublished manuscript, 2005a.

[226]Steels L., “The emergence and evolution of linguistic structure: from lexical to grammatical communication systems,” *Connection Science*, vol. 17(3-4), pp. 213-230, 2005b.

[227]Steels L., “Experiments on the emergence of human communication,” *Trends in Cognitive Sciences*, 10(8), pp. 347-349, 2006.

[228]Steels L., “The Recruitment Theory of Language,” in *Emergence of Communication and Language*, C. Lyon, C. L. Nehaniv, and A. Cangelosi, Eds. Springer, 2007, pp. 129-150.

[229]Steels L., and De Beule J., “Unify and merge in fluid construction grammar,” in *Symbol Grounding and Beyond: Proceedings of the Third International Symposium on the Emergence and Evolution of Linguistic Communication. Lecture Notes in Computer Science (LNCS/LNAI) 4211*, Vogt, P. et al. Eds. Berlin: Springer, 2006, pp 197-223.

[230]Steels L., and Belpaeme T., “Coordinating perceptually grounded categories through language. a case study for color,” *Behavioral and Brain Sciences*, vol. 24(8), pp. 469-529, 2005.

[231]Steels L. and Kaplan F. () AIBO's first words: The social learning of language and meaning. *Evolution of Communication*, vol. 4(1), pp. 3-32, 2001.

[232]Striano T., Henning A. and Stahl D., “Sensitivity to social contingencies between 1 and 3 months of age,” *Developmental Science*, vol. 8, pp. 509-518, 2005.

[233]Sugita Y., and Tani J., “Learning semantic combinatoriality from the interaction between linguistic and behavioral processes,” *Adaptive Behavior*, vol. 13(1), pp. 33-52, 2005.

[234]Swingley D., “Contributions of infant word learning to language development,” *Philosophical Transactions of Royal Society. B*, vol. 364, pp. 3617 - 3632, Dec 2009.

[235]Tanaka F., Cicourel A., and Movellan J. R., “Socialization between toddlers and robots at an early childhood education center,” *Proceedings of the National Academy of Sciences*, vol. 104, pp. 17954-17958, 2007.

[236]Tanaka M., and Tanaka S., *Developmental diagnosis of human infants: from 6 to 18 months* [in Japanese]. Otsuki, Tokyo, 1982.

[237]Tani J., and Nolfi S., “Learning to perceive the world as articulated: An approach for hierarchical learning in sensory-motor systems,” *Neural Networks*, vol. 12, pp. 1131-1141, 1999.

[238]Tellier I., “Towards a semantic-based theory of language learning,” *Proceedings of 12th Amsterdam Colloquium*, pp. 217-222, 1999.

[239]Thelen E., and Smith L., A *Dynamic Systems Approach to the Development of Cognition and Action* (Cognitive Psychology Series). Cambridge, MA: MIT Press1994.

[240]Tomasello M., “The role of joint attentional processes in early language development,” *Language Sciences*, vol. 10, pp. 69-88, 1988.

[241]Tomasello M., *First verbs: A case study in early grammatical development*. Cambridge: Cambridge University Press, 1992.

[242]Tomasello M., “Joint attention as social cognition,” in *Joint Attention: Its Origins and Role in Development*. C. Moore and P. Dunham, Eds. Mahwah: Lawrence Erlbaum, 1995.

[243]Tomasello M., *The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard University Press, 1999.

[244]Tomasello M., “The item-based nature of children’s early syntactic development,” *Trends in Cognitive Sciences*, vol. 4(4), pp. 156-163, 2000.

[245]Tomasello M., “Could we please lose the mapping metaphor, please?” *Behavioral and Brain Sciences*, vol. 24, pp. 1119–1120, 2001.

[246]Tomasello M., *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University, 2003.

[247]Tomasello M., and Abbot-Smith, K., “A tale of two theories: response to Fisher,” *Cognition* vol. 83, pp. 207-214, 2002.

[248]Tomasello M., Carpenter M. and Lizskowski U., “A new look at infant pointing,” *Child Development* vol. 78, pp. 705-22, 2007.

[249]Tomasello M., Carpenter M., Call J., Behne T., and Moll H. Y., “Understanding and sharing intentions: the origins of cultural cognition,” *Behavioral and Brain Sciences*, vol. 28, pp. 675–735, 2005.

[250]Trevarthen C., “Communication and cooperation in early infancy: a description of primary intersubjectivity,” in *Before Speech: The Beginning of Interpersonal Communication*, M. Bullowa, Ed., Cambridge University Press, 1979.

[251]Trevarthen C., “Musicality and the intrinsic motive pulse: evidence from human psychobiology and infant communication” (special issue), *Musicae Scientae*, pp. 155–215, 1999.

[252]Tronick E., Als H., Adamson L., Wise S., and Brazelton T.B., “The infants’ response to entrapment between contradictory messages in face-to-face interactions,” *Journal of the American Academy of Child Psychiatry*, vol. 17, pp. 1–13, 1978.

[253]Vihman M.M., and Depaolis R.A., “The role of mimesis in infant language development: evidence from phylogeny?” in *The Evolutionary Emergence of Language*, C. Knight, M. Studdert-Kennedy and J. R. Hurford, Eds. Cambridge University Press, 2000.

[254]Vogt P., “Anchoring of semiotic symbols,” *Robotics and Autonomous Systems*, vol. 43(2), pp. 109-120, 2003.

[255]von der Malsburg C., and Singer W., “Principles of cortical network organisations,” in *Neurobiology of the Neocortex*, P. Rakic and W. Singer, Eds. London: John Wiley and Sons Ltd, 1988, pp. 69–99.

[256]von Hofsten C., “An action perspective on motor development,” *Trends in Cognitive Sciences*, vol. 8(6), pp. 266-272, 2004.

[257]Waxman S. and Markow D., “Words as invitations to form categories: Evidence from 12 to 13-month-old infants,” *Cognitive Psychology*, vol. 29(3), pp. 257-302, 1995.

[258]Waxman S. R., “Specifying the scope of 13-month-olds’ expectations for novel words,” *Cognition*, vol. 70, pp. B35–B50, 1999.

[259]Weng J., and Hwang H., “From neural networks to the brain: Autonomous mental development. *IEEE Computational Intelligence Magazine*,” vol. 1(3), pp. 15-31, 2006..

[260]Weng J., “Developmental robotics: theory and experiments,” *International Journal of Humanoid Robotics*, vol. 1(2), 2004.

[261]Weng J., “On developmental mental architectures,” *Neurocomputing*, vol. 70(13-15), pp. 2303-2323, 2007.

[262]Weng J., McClelland J., Pentland A., Sporns O., Stockman I., Sur M., and Thelen E., “Autonomous mental development by robots and animals,” *Science*, vol. 291, pp. 599–600, 2001.

[263]Westermann G., Mareschal D., Johnson M.H., Sirois S., Spratling M. and Thomas M., “Neuroconstructivism,” *Developmental Science*, vol. 10(1), pp 75–83, 2007.

[264]Westermann G., Sirois S., Shultz T.R., and Mareschal D., “Modeling developmental cognitive neuroscience,” *Trends in Cognitive Sciences*, vol. 10 (5), pp. 227–233, 2006.

[265]Wiemer-Hastings K., and Xu X., “Content differences for abstract and concrete concept,” *Cognitive Science*, vol. 29(5), pp. 719-736, 2005.

[266]Winawer J., Witthoft N., Frank M. C., Wu L., Wade A. R., and Boroditsky L., “Russian blues reveal effects of language on color discrimination,” *Proceedings of the National Academy of Sciences*, vol. 104(19), pp. 7780-7785, 2007.

[267]Wittgenstein L., *Philosophical Investigations (Philosophische Untersuchungen)*. German with English translation by G.E.M. Anscombe, 3rd ed. Basil Blackwell, 1968, (first published 1953), 1953.

[268]Wolpert D.M., and Kawato M., “Multiple paired forward and inverse models for motor control,” *Neural Networks*, vol. 11, pp. 1317–1329, 1998.

[269]Wood D., Bruner J. S., and Ross G., “The role of tutoring in problem solving,” *Journal of Child Psychology and Psychiatry*, vol. 17, pp. 89-100, 1976.

[270] Woodward A. L., “Infant selectively encode the goal object of an actor's reach,” *Cognition*, vol. 69, pp. 1-34, 1998.

[271] Wray A., “Protolanguage as a holistic system for social interaction,” *Language and Communication*, vol. 18, pp. 47-67, 1998.

[272] Wrede B., Kopp S., Rohlfing K.J., Lohse M., and Muhl C., “Appropriate feedback,” *Journal of Pragmatics*. in press.

[273] Wrede B., Rohlfing K. J., Hanheide M., and Sagerer G., “Towards Learning by Interacting,” in *Creating Brain-Like Intelligence, Lecture Notes in Artificial Intelligence 5436*, B. Sendhoff et al. Eds. Berlin Heidelberg: Springer-Verlag, 2009, pp. 139-150.

[274] Xu F., “The role of language in acquiring kind concepts in infancy,” *Cognition*, vol. 85, pp. 223–250, 2002.

[275] Yamashita Y., and Tani J., “Emergence of functional hierarchy in a multiple timescale neural network model: a humanoid robot experiment,” *PLoS Computational Biology*, vol. 4, 11, 2008.

[276] Yoshida H., and Smith L.B. “Linguistic cues enhance the learning of perceptual cues,” *Psychological Science*, vol. 16(2), pp. 90-95, 2005.

[277] Zeschel, A. (2007). *Delexicalisation patterns: a corpus-based approach to incipient productivity in ‘fixed expressions.* Doctoral dissertation, Universität Bremen..

[278] Zeschel A., and Fischer K., “Constructivist grammar classifications for grounded language learning,” *Deliverable 3.1, ITALK project*, www.italkproject.org, 2009.

[279] Zhang Y., and Weng J., “Task Transfer by a Developmental Robot,” *IEEE Transactions on Evolutionary Computation*, vol. 11(2), pp. 226-248, 2007.

[280] Zukow-Goldring P., “Assisted imitation: Affordances, effectivities, and the mirror system in early language development,” in *From Action to Language*, Arbib, M.A. Ed. Cambridge: CUP, 2006, pp. 469-500.
.

Angelo Cangelosi is Professor of Artificial Intelligence and Cognition at the University of Plymouth (UK), where he leads the Centre for Robotics and Neural Systems. He obtained a PhD in psychology and computational modeling in 1997 from the University of Genoa, whilst also working as visiting scholar at the National Research Council (Rome), the University of California San Diego (USA) and the University of Southampton (UK). He has produced more than 140 scientific publications, and has been awarded numerous research grants from UK and international funding agencies for a total of over £10Million.

Cangelosi is co-Editor-in-Chief of the “Interaction Studies” journal and serves in the editorial board of the “Journal of Neurolinguistics”, “Connection Science” and “Frontiers in Neurorobotics”. He has chaired various international conferences, including the 6th International Conference on the Evolution of Language (Rome 2006) and the 9th Neural Computation and Psychology Workshop (Plymouth, 2004). He has been invited/keynote speaker at numerous conferences and workshops. Books edited by Cangelosi include “Simulating the Evolution of Language” (2002, Springer, co-edited with D. Parisi) and “Emergence of Communication and Language” (2007, Springer; co-edited with C. Lyon and C.L. Nehaniv).

He is coordinator of the EU FP7 project “ITALK: Integration and Transfer of Action and Language Knowledge in Robot”, a multi-partner research project on action and language learning on the humanoid robot iCub. He also coordinates the “RobotDoC” Marie Curie doctoral network in developmental robotics, and the UK Cognitive Systems Foresight project “VALUE” co-funded by EPSRC, ESRC and BBSRC.

All the other co-authors are members of the ITALK project.
.

Action learning	Developmental learning of simple actions (primitives) Capacity to categories and name objects, events and states Ability to detect objects, gaze, reach and clasp the hand around the object	Acquisition of hierarchical and compositional actions	Learning the association between syntactic constructions and composite actions via social learning	Social based acquisition of action generalization rules Ability to correlate action and language generalization capabilities	Acquisition of the ability to generalize over goals Ability to correlate recursive /composite actions with recursive linguistic expressions	Ability to learn rich action repertoires based on social/linguistic descriptions
Language learning	Grounded acquisition, decomposition and generalization of simple transitive holophrases in learning by demonstration tasks	Grounded acquisition, decomposition and generalization of the five basic argument structure constructions of English from holophrastic instances in learning by demonstration tasks	Grounded interactive language learning games in simple joint attention scenarios based on the implementation of elementary socio-cognitive/pragmatic capabilities	Learning from increasingly more complex/diversified linguistic input within progressively less restricted learner-tutor interactions	Progressively more human-like cooperative ostensive-inferential communication based on the implementation of more advanced socio-cognitive/pragmatic capabilities	Learning progressively more complex grammars from quantitatively naturalistic input
Social learning	Harnessing of elementary non-verbal social cues (gaze, turn-taking, mirroring etc) to enhance social learning for language and skill acquisition Modeling holophrase acquisition via intermodal learning (acoustic packaging)	Development of a tutor spotter for social learning scenarios Joint intentional framing and referential intent Acquisition of negation usage of various types (eg refusal, absence, prohibition, propositional denial)	Development of architectures capable of exploiting pragmatic skills such as sequential interactional organization (contingency, turn-taking) and use of prosody for grammatical learning Harnessing of Model/Rival (M/R) learning, motivational systems and predictive social interaction	Exploiting interactions of prosody, internal motivation, inter-subjectivity and pragmatics in language acquisition and dialogue Developing architectures based on intermodal learning and sensitivity to a tutor	Temporally extended understanding of the social motivations and intentions of other minds, context, and (auto)biographic and narrative (re)construction Development of first systems that are capable of social learning and sequential organization of interaction in specific scenarios	Development of systems that are capable of social learning and pragmatic organization of interaction in various scenarios
Cognitive integration	Integration of basic action and naming representations and emergence of shared representation roles for both actions and names	Simulation of Action-Language Compatibility effects Co-evolution of action and language skills for simple grounded lexicons	Computational neuroscience models of action and language integration	Use of general purpose grammatical constructions for the creation of new complex motor and perceptual concepts	Scalable lexicons of abstract concepts based on the developmental acquisition of a grounding kernel	Acquisition of open repertoires of compositional actions and lexicons sharing natural language properties
	Next 2 Years	**Next 4 Years**	**Next 6-8 Years**	**Next 10 Years**	**Next 15 Years**	**Next 20 Years**
TIME						