External Symbol Grounding Workshop 2006

3 and 4 July 2006

Plymouth, United Kingdom

BOOK OF ABSTRACTS AND PAPERS

The Distributed Language Group

euCognition
www.euCognition.org

The University of Plymouth

External Symbol Grounding workshop 2006

3 and 4 July 2006
Plymouth, United Kingdom

Welcome to Plymouth and to the External Symbol Grounding workshop 2006. ESG2006 is an international workshop for research on grounding external signs and symbols, and is the successor to the first Distributed Language Group's Conference on Cognitive Dynamics and the Language Sciences, held at Sidney Sussex College, Cambridge on 9-11 September 2005.

ESG2006 brings together scholars from a number of disciplines who view language and cognition as linking what goes on in the head with causal processes that are intersubjective, multimodal, affect-laden, and organised by historically rooted customs and artefacts. The main topic of the workshop will be to consider how symbol grounding can be reconsidered when language is viewed as a dynamical process rooted in both culture and biology. Research related to robotic or computer modelling of symbol grounding, psychological and linguistic viewpoints on cognitive development and semiotic dynamics are of great interest.

The workshop brings together linguists, psychologists, ethologists and social biologists, social and cognitive neuroscientists, philosophers, computer scientists, and roboticists for an intense two days of presenting and discussing (potentially incompatible) views. We hope that the workshop's limited size and its informal setting will encourage interaction.

The work presented at the workshop is eligible for publication in a special issue of Interaction Studies. Detailed instructions can be found on the workshop's webpage at http://www.tech.plym.ac.uk/SoCCE/ESG2006/.

The workshop is kindly sponsored by the euCognition (the European Network for the Advancement of Artificial Cognitive Systems) to support two international speakers and to reduce the registration fees. Also, Stephen J. Cowley and the Distributed Language Group he is leading deserve special mention for proposing and setting off the organisation of ESG2006 here in Plymouth.

We hope you enjoy the workshop, and trust it will be a highly productive and sociable event.

Tony Belpaeme
Karl MacDorman

ESG2006 workshop organisers

Tony Belpaeme
Karl MacDorman
Stephen J. Cowley
Angelo Cangelosi

Workshop secretariat

Short Course Unit, Faculty of Technology, University of Plymouth

Internal supports for external symbol grounding

Michael L. Anderson
University of Maryland, College Park

Traditionally, the symbol grounding problem--or, at least, the solution to the symbol grounding problem--has been conceived in terms of an individual, isolated symbol user (generally a robot, but sometimes a baby) having to make the appropriate connections not just between a symbol and other symbols, but between symbols, perceptions, and actions. Recently, however, more attention has been paid to the fact that grounding symbols--learning a language, for instance--is not solipsistic but social; one grounds symbols in cooperation with others.

There are many implications of this shift in emphasis, implications for the nature of language and of cognition, and for the design of robots and machine learning algorithms. Perhaps the most important immediate effect is to bring the conception of the symbol grounding problem more in line with the long-recognized fact that meanings are not private, but shared.or, as some of the work in this area puts it, symbols and their meanings are "distributed". One predictable result of this has been to shift research attention away from the individual agent, and on to groups of agents, and the dynamics of those groups. This is well and good, but we should not suppose that the problem can be entirely cast (much less solved) at the level of social processes. The individual symbol user still has internal mechanisms that must be understood. That being said, the recognition of the importance (even the centrality) of social interactions in the symbol grounding process has important implications for how we should understand the nature of the individual, internal mechanisms that support the symbol grounding process. What must they be like to allow the individual to participate in the collective practice of symbol grounding? What sorts of mechanisms support the individual-social interface? What are the grounds of intersubjective interpretation, and what makes normativity possible? It is on these questions that will focus in my talk. I will present my work on the bodily and behavioral roots of intentionality in terms that I hope will be useful to the broader project of situating symbol grounding in social/cultural context.

More specifically I will ask (and answer) the following questions: (1) What is the nature and content of mental representations, such that they are suited to public interpretation? (2) What accounts for the intersubjective interpretability of human behavior? (3) How is it possible to ground the normativity of content in a public, external world, and thereby avoid a shared conceptual internalism (linguistic relativity)? Because my general approach to answering these questions involves highlighting the ways in which mental representation, intersubjective interpretation, and normativity depend upon the body and behavior, I also ask (4) What is the functional structure of the brain that supports the grounding of higher-order content (e.g. language) in simple bodily capacities?

Briefly, my answers to these questions are

(1) According to the guidance theory of representation, (Rosenberg and Anderson 2004; Anderson and Rosenberg in press) content is to be cashed out in terms of action-guidance, and intentionality is grounded in the natural directeness of action. The guidance theory offers a way of fixing representational content that gives causal and evolutionary history only an indirect (non-necessary) role, and allows for an account of representational error, expressed in terms of failure

of action, that does not rely on any such notions as proper function, ideal perceptual/epistemic conditions, or normal circumstances. Importantly, the failure of an action is an event detectable by, and thus available to, not just the agent, but also the agent's compatriots.

(2) Relying primarily on the work of Shaun Gallagher (2005), I account for the possibility of intersubjective interpetation of actions with reference to an innate body schema. The body schema is a system of sensory-motor capacities, encompassing all of the non-conscious aspects of motor control--including sub-cortical, pre-motor, and motor processes in the brain, as well as the information systems required for these processes to function properly. The body schema acts as an organizational framework not just for the actions of the agent herself, but for the agent's perceptions of the actions of others. Thus, for instance, are infants capable of imitating facial gestures at birth (Meltzoff & Moore, 1977). They don't have to learn to see, much less learn to interpret what they see in terms of their own motor possibilities; the motoric equivalent of a visually perceived facial gesture is already a part of their experience.

(3) The multiple modes theory of epistemic openness (Anderson 2005) holds that embodied agents are more epistemically porous than is generally pictured, open to the world via multiple channels, each operating for particular purposes and according to its own logic. Among the most important of these modes of epistemic openness is our physical intervention in the world, which is not (in all its aspects) theory-laden in the manner of visual perception, and which therefore can serve to ground our knowledge in a way that vision alone cannot. This mode of epistemic openness is what allows one to ground normativity in the success or failure of action, grounding representational ocntent in contact with the external (physical and social) world.

(4) The massive redeployment hypothesis (Anderson 2006; in press) is an account of the functional structure of the brain that emphasizes the fact that many brain areas are used to support multiple cognitive functions. For instance, there is evidence for the use of brain areas primarily associated with motor control in functions as diverse as language understanding and working memory. When this fact is combined with the idea that motor control is a matter of guiding sensory-motor feedback loops, then this suggest we should treat motor control in terms of affordance processing. Since affordances, the perceived availability of objects for certain kinds of interaction, aren't just motor programs, but interpretations of the environment, this opens the possibility that the motor control system is also, already, a primitive meaning processor. This would offer one explanation for how it is even possible to leverage motor control to support and constrain higher-order processes like language understanding. This is not just of great theoretical interest but has significant practical implications; understanding this phenomenon can help us to design a system that could be used both for motor-control, and to ground higher-order representations. Such a breakthrough could lead to the development of new, more effective robotic control systems, with better integration between reasoning, perceiving, and acting.

Anderson, M.L. (in press). The massive redeployment hypothesis and the functional topography of the brain. Philosophical Psychology.

Anderson, M.L. (2006). Evidence for massive redeployment of brain areas in cognitive functions. Proceedings of the 28th Annual Meeting of the Cognitive Science Society.

Anderson, M.L. (2005) Cognitive science and epistemic openness. Phenomenology and the Cognitive Sciences 4(4).

Anderson, M.L. and Rosenberg, G. (in press). Content and action: The guidance theory of representation. In: D. Smith (ed) Evolutionary Biology and the Central Problems of Cognitive Science, a special issue of Journal of Mind and Behavior.

Gallagher, S. (2005). How the Body Shapes the Mind. Oxford: Oxford University Press.

Meltzoff, A. & Moore, M.K. (1977). Imitation of facial and manual gestures by human neonates. Science, 198, 75.78.

Rosenberg, G. and Anderson, M.L. (2004). A brief introduction to the guidance theory of representation. Proceedings of the 26th Annual Conference of the Cognitive Science Society.

The memetic evolution of colour words and colour categories

Tony Belpaeme
University of Plymouth
tony.belpaeme@plymouth.ac.uk

The influence of Chomskian doctrine has slowly but surely made place for a more culturalist view of language and cognition. This is obvious in the renewed attention for linguistic relativism. Ever since Sapir and Whorf in the 1950s suggested that different languages carve up the world in different ways and thus propose that language has an immediate impact on how its users experience and reason about the world (Sapir, 1921; Whorf, 1956), the Sapir-Whorf hypothesis has been attacked and ridiculed. Pinker (1994) for example popularised this by taking on strong linguistic relativism at a time when no scholar longer underwrote it. This depreciatory interpretation of linguistic relativism has in recent years been countered by a growing number of experimental evidence for weak linguistic relativism. Language *does* have an influence on perception and cognition, albeit a subtle one. This has recently been demonstrated for, among others, time (Boroditsky, 2001), space (Gumperz and Levinson, 1996; McDonough et al., 2000), shapes (Lucy and Gaskins, 2001) and objects. For an overview see (Lucy, 1996) or (Boroditsky, in press).

Moreover, when limiting ourselves to categories and concepts, language seems to have a beneficial impact on learning novel categories and concepts. Xu (2002) shows how the use of linguistic labels speeds up the learning of concepts in young children. This effect is not only observed in childhood, recently (Lupyan, 2006) showed how labelling objects facilitated the learning of a conceptual distinction between objects, an effect which was not observed when object were not linguistically labelled.

When studying colour categories, ideas of linguistic relativism have always been latently present. In 1961 Gleason wrote

> There is a continuous gradation of colour from one end of the spectrum to the other. Yet an American describing it will list the hues as red, orange, yellow, green, blue, purple, or something of the kind. There is nothing inherent either in the spectrum or the human perception of it which would compel its division in this way. (Gleason, 1961, quoted in Berlin and Kay, 1969, p. 159).

But this relativist thinking was soon to be replaced by the universalist and nativist theories laid out in (Berlin and Kay, 1969). B&K have noticed that referents of colour terms were remarkably similar between different languages (an observation which has been reconfirmed in the more extensive and better executed World Color Survey; Kay et al., 2003). B&K blew all linguistic relativistic thinking about colour out of the water, although a few pockets of resistance were left (f.i. Lucy and Shweder, 1979; Davies and Corbett, 1997; Gellatly, 1995). Recently however, Gilbert et al. (2006) convincingly showed that linguistic relativism does apply to colour perception, even more, it seems to be lateralised as well: colour discrimination in the right visual field is more affected by language than colour discrimination in the left visual field.

There is a convincing base of support for language having an impact on colour perception, colour categorisation and colour cognition. But little is known and little has been suggested as to how this can be put in the larger frame of language evolution and acquisition. If colour cognition is subject to the language we use to describe colour, and if we accept that language is culturally distributed (Hutchins, 1995), then colour categories must be subject to culture as well. However, culture is sometimes seen as the epitome of arbitrariness, and although this is somewhat of a caricature, it illustrates the problem of seeing colour cognition as being subject to cultural conventions. The question that we need to find an answer to is: how can different languages have similar colour categories, if those colour categories are under influence of language, which an inherently arbitrary cultural convention?

I will argue how constraints on colour perception and colour categorisation can drive colour categories to take up positions in perceptual space that are shared between different languages. To illustrate this, a computational model has been devised that demonstrates the principle of cultural evolution of perceptual categories (Belpaeme and Bleys, 2005; Steels and Belpaeme, 2005). The crux of the model is that language users coordinate their perceptual categories through dyadic linguistic interactions. The need for successful communication drives the categories to become similar between language users, and the constraints laid down by perception (both neurophysiological and psychological) makes that categories will generally crystallise in positions that are similar between different languages. This memetic process of cultural transmission of colour categories does not happen in a void, but is constrained and steered by the body and the environment.

References

Belpaeme, Tony and Bleys, Joris (2005) Explaining universal colour categories through a constrained acquisition process. *Adaptive Behavior.* 13(4):293-310.

Berlin, B. and Kay, P. (1969). Basic Color Terms: Their Universality and Evolution. University of California Press, Berkeley, CA.

Boroditsky, L. (2001) Does language shape thought? English and Mandarin speakers' conceptions of time. *Cognitive Psychology*, 43(1), 1-22.

Boroditsky, L. (in press). Linguistic Relativity. To Appear in the Encyclopedia of Cognitive Science. MacMillan Press.

Davies, I. R. and Corbett, G. (1997). A cross-cultural study of colour grouping: Evidence for weak linguistic relativity. British Journal of Psychology, 88:493–517.

Gellatly, A. (1995). Colourful Whorfian ideas: Linguistic and cultural influences on the perception and cognition of colour, and on the investigation of them. Mind and Language, 10(3):199–225.

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

Hutchins, E. (1995) Cognition in the wild. The MIT Press, 1995.

Kay, P., Berlin, B., Maffi, L., and Merrifield, W. R. (2003). The World Color Survey. Center for the Study of Language and Information, Stanford.

Lucy, J. A. and Shweder, R. A. (1979). Whorf and his critics: Linguistic and nonlinguistic influences on color memory. American Anthropologist, 81:581–615.Lucy, J. A. (1996). The scope of linguistic relativity: an analysis and review of empirical research. In Gumperz, J. J. and Levinson, S. C., editors, Rethinking linguistic relativity, Studies in the Social and Cultural Foundations of Language 17. Cambridge University Press, Cambridge.

Lucy, J. & Gaskins, S. (2001) Grammatical categories and the development of classification preferences: a comparative approach. In Bowerman, M. & Levinson, S. (eds.) Language acquisition and conceptual development. Cambridge University Press, 257-283.

Lupyan, G. (2006) Labels facilitate learning of novel categories. In Cangelosi, A., Smith, A.D.M. and Smith, K. (eds.) The evolution of language: proceedings of the 6th conference on the evolution of language. World Scientific, Singapore. 190-197.

Gleason, H. (1961). An Introduction to Descriptive Linguistics. Holt, Rinehart and Winston, New York.

Gumperz, J. J. and Levinson, S. C. (1996). Rethinking Linguistic Relativity. Studies in the Social and Cultural Foundations of Language 17. Cambridge University Press, Cambridge.

McDonough, L., Choi, S. and Mandler, J. (2000) Development of language-specific categorization of spatial relations from prelinguistic to linguistic stage: a preliminary study. In Finding the Words Conference, Stanford University, Stanford, California.

Pinker, S. (1994). The language instinct: How the mind creates language. W. Morrow, New York.

Sapir, E. (1921). Language: An introduction to the study of speech. Harcourt, Brace and Co., New York.

Steels, L. and Belpaeme, T. (2005). Coordinating perceptually grounded categories through language. A case study for colour. Behavioral and Brain Sciences, 24(8):469–529.

Whorf, B. L. (1956). Language, Thought and Reality: selected writings of Benjamin Lee Whorf. The MIT Press, Cambridge, MA. Edited by Carrol, J.B.

Xu, F. (2002). The role of language in acquiring object kind concepts in infancy. Cognition, 85:223–250.

The grounding and sharing of symbols

Angelo Cangelosi
University of Plymouth

The double function of language, as a social/communicative means, and as an individual/ cognitive capability, derives from its fundamental property that allows us to internally re-represent the world we live in. This is possible through the mechanism of symbol grounding, i.e., the ability to associate entities and states in the external and internal world with internal categorical representations. The symbol grounding mechanism, as language, has both an individual and a social component. The individual component, called the "Physical Symbol Grounding", refers to the ability of each individual to create an intrinsic link between world entities and internal categorical representations. The social component, called "Social Symbol Grounding", refers to the collective negotiation for the selection of shared symbols (words) and their grounded meanings. In the talk we will discuss these two aspects of symbol grounding in relation to distributed cognition, using examples from cognitive modeling research on grounded agents and robots.

Main reference:

Cangelosi A. (2006). The grounding and sharing of symbols. Pragmatics and Cognition, 14(2), 275-285

Semiotic Symbols and the Missing Theory of Thinking

Robert Clowes
COGS Centre for Research in Cognitive Science
University of Sussex
robertc@sussex.ac.uk

Paul Vogt's (2002) *The Physical Symbol Grounding Problem* is perhaps the most sustained attempt to show that the Adaptive Language Game (ALG) framework (Steels 1999) can provide the role of showing not only how symbols are grounded but also why they should still be regarded as a central concept to an embodied cognitive science. Vogt's approach rests on solving the symbol grounding problem within the framework of embodied cognitive science. He argues that symbolic structures can be used within the paradigm of embodied cognitive science by adopting an alternative definition of a symbol: the semiotic symbol (cf. Deacon 1997). Versions of such a definition has found favour in much recent modelling work (Cangelosi, Greco and Harnad 2000; Kaplan 2000).

But showing that symbols can be grounded does not amount to showing that they can play a cognitive role. This requires showing that semiotic symbols can also play a role in thinking. While some recent computational work demonstrates some cognitive effects based on what has been called *symbolic theft* (Cangelosi, Greco et al. 2000) it does not yet amount to a theory of symbolic thinking. For this, I argue, a theory of *internalisation* must first be given. I.e. a theory of how symbols can take up properly cognitive roles that allow the reshaping and reconstruction of an agent. The semiotic symbol system approach as it is currently generally expressed is, I claim, embarrassed by a missing theory of thinking.

In experiments reported elsewhere (Clowes and Morse 2005) a proof of concept model is offered as to how symbols come to play a role over developmental time in cognitive economy. The model presents one way in which a symbol system can be appropriated and put to novel use by a cognitive agent. It demonstrates, perhaps surprisingly, that even an incredibly minimal simulated agent can appropriate and make use of some of the potential intelligence encapsulated in the symbol and in so doing reshape its own cognitive architecture. I argue this model provides some strength for the Vygotskian theory that sees the establishment of properly human thinking as the internalisation of social tools.

This model also allows us an intuitive basis on which to construct a theory of symbolically mediated thinking that can attempt to do justice some of the unique features of human thought. In particular, it allows us to give an account of the role that the internalisation of symbols play in the constitution of an inner world. Moreover, by basing this account upon a theory of symbol internalisation, rather than just presupposing internal symbols, we do not risk sliding back into any version of GOFAI.

Cangelosi, A., A. Greco, et al. (2000). "From Robotic Toil to Symbolic Theft: Grounding Transfer from Entry-Level to Higher-Level Categories." Cognitive Science **12**(2): 143 - 162.

Clowes, R. W. and A. Morse (2005). Scaffolding Cognition with Words. Proceedings of the 5th International Workshop on Epigenetic Robotics. L. Berthouze, F. Kaplan, H. Kozimaet al. Nara, Japan, Lund University Cognitive Studies, 123. Lund: LUCS.

Deacon, T. W. (1997). The Symbolic Species: The Co-Evolution of Language and the human brain, The Penguin Press, Penguin Book Ltd.

Kaplan, F. (2000). Semiotic schemata: Selection units for linguistic cultural evolution. Proceedings of Artificial Life 7, MA, MIT Press.

Steels, L. (1999). The Talking Heads Experiment: Volume I. Words and Meanings. Antwerpen, Laboratorium.

Vogt, P. (2002). "The physical symbol grounding problem." Cognitive Systems Research Journal **3**(3): 429-457.

Grounding external symbols in humans and other agents

Stephen J. Cowley,
University of Hertfordshire & University of KwaZulu-Natal

A distributed view of language throws new light on *human symbol grounding*. How infants learn to talk is traced to mutual gearing. The model can thus be used as a benchmark in re-examining computer simulations of linguistic agency and thus symbol grounding. This is because, if language is a meshwork, dyads become the means whereby infants connect body-based signals with a culture's virtual resources. While dyads initially use affect, their developing routines enable a baby to become skilled in assessing and managing caregivers. As an attachment or relationship grows, infants make real-time use of *extended* symbols or utterance-activity. Talk emerges as they discover effective ways of participating in what the dyad does. Infants gain from attending to how adults enact beliefs, affect and attitudes: without understanding, they slowly become quasi-linguistic agents. By the end of their first year, infants often show practical knowledge of *when* to act as instructed and, indeed, produce syllables heard as 'more', 'milk' or 'car'. Using good old folk psychology, adults formulate the belief that babies 'know' words: they take a *language stance*. Later, of course, children also adopt the stance. It serves them not only predict what others want but also in developing the many behavioural strategies that are made possible by selves.

To take a language stance is to hear (and think) with the verbal patterns that allow culture to permeate behaviour. Symbolic theft (Cangelosi et al., 2002) thus becomes possible as infant agents begin to hear in words. Far from needing inner verbal symbols, the child links phonetic patterns to virtual patterns caregivers use consistently in managing joint activity. The parties depend on *gearing* to both objects and real-time actions (Cowley, in press b). As infants become skilled in indexing norms that link circumstances to a caregiver's dynamics (Cowley, 2005), their agency changes. Reviewing evidence of these transformations, I highlight simple tricks. First, infant biases drive a dyad to develop *norm-based* routines (Cowley, 2003). Second, caregiver affect, beliefs and desires serve to bring activity under *dual control* (Cowley et al., 2004). Third, infant behaviour becomes *analysis amenable* as its functions begin to co-vary with what adults call 'words' (Spurrett & Cowley, 2004; Cowley, 2004). Skills in timing enable infants to use their voices to control adults and, eventually, in taking a language stance. Once this occurs, children too use historically derived virtual patterns in seeking to interpret and control events. Later developmental stages use the (false) belief that children's selves can reveal facts about words and norms.

While physically grounded, no linguistic representations are needed to ground learning to talk. Rather, social learning enables agents to develop behaviour that sustains practical *belief* in symbols. Social routines exploit a virtual system which, I suggest, is powerful enough to sustain – not only verbal aspects of language –but also those which permit compositionality and productivity. Initially, however, infants learn to discern what is of interest to others. Alongside internal symbol grounding, they sensitise to *how* caregivers regard the world-perceived. As the perspectives mesh, routines develop in ways that allow the dyad's co-action to align with conventional vocalizations. A baby's capacity to hear external symbols is thus based in *other-mediated* interaction. Whereas body-world relations are the basis for learning physical categories, a child's sensitivity to virtual counterparts depends on how adults enact cultural norms (Cowley and MacDorman, in press). In grounding external symbols, infant decision-making makes increasing use of cultural norms. This arises as the baby is sensitized to what can be gained by actively anticipating what adults want and expect. Affective signals are used to organize routines around physical objects that enable a child to increase its grasp of how adult vocalizations manifest reasons.

In conclusion, I pursue how the gearing model can be applied to computer simulations and, perhaps, robots (Cowley, in press a). Agents, it is suggested, can be designed such that they can mimic how infants use the dynamics of utterance-activity to discover the power of virtual symbols.

References

Cangelosi, A. Greco, A. & Harnad, S. (2002). Symbol grounding and the symbolic theft hypothesis. In A. Cangelosi & D. Parisi (eds) *Simulating the evolution of language*. Springer: Berlin, 191-210.

Cowley, S.J. (2003). Distributed cognition at three months: mother-infant dyads in kwaZulu Natal. *Alternation*, 10.2: 229-257.

Cowley, S.J. (2004). Contextualizing bodies: how human responsiveness constrains distributed cognition. *Language Sciences*, 26/6, 565-591.

Cowley, S.J. (2005). Languaging: How humans and bonobos lock on to human modes of life. *International Journal of Computational Cognition*, 3/1: 44-55.

Cowley, S.J. (in press a). Distributed language, biomechanics and the origins of talk. To appear in Lyon C., Nehaniv C.L., Cangelosi A. (Eds.) (in preparation*). Emergence and Evolution of Linguistic Communication*. Springer.
Early version available at: http://www.tech.plym.ac.uk/SoCCE/ESG2006/

Cowley, S. J. (in press b). The Cradle of Language: making sense of bodily connections. To appear in D. Moyal-Sharrock (ed.) *Perspicuous Presentations*.

Cowley, S.J., Moodley, S. & Fiori-Cowley, A. (2004). Grounding signs of culture: primary intersubjectivity in social semiosis. *Mind, Culture and Activity*, 11/2: 109-132.

Cowley, S.J. & MacDorman, K. (in press). What baboons, babies and Tetris players tell us about interaction: a biosocial view of norm-based social learning. To appear in *Connection Science*.

Spurrett, D. & Cowley, S.J. (2004) How to do things without words. *Language Sciences*, 26/5: 443- 466.

Wittgenstein's Error: All Language is Public, But Not Necessarily Social

Stevan Harnad
Université du Québec à Montréal

Abstract: Wittgenstein argued that an individual cannot invent a "private language." The reason is the problem of error. To name things there has to be a right and a wrong of the matter: if I am the sole arbiter of what is called an "X", who/what determines whether it is really an X that I'm calling an X on every occasion? (The best example is subjective categories: a mood that I call M every time it recurs: how can I know it's the same mood?) So for symbol grounding there has to be external error-correction, in the form of feedback from the objective consequences of naming things correctly or incorrectly. Wittgenstein thought the feedback had to come from a community of language-users, agreeing on shared rules of a language "game" (i.e., agreeing on what they would call what). But that was an error. Although there is no real motivation for a lifelong hermit to invent a solo language, nothing would prevent him from naming or even describing things based solely on feedback from the objective consequences of misnaming (calling "toxic" mushrooms "edible," to use a familiar example). The "social" dimension of naming has more to do with why we bother to invent a language at all: to communicate truths (share categories) to one another, for mutual benefit. Social categories (kin, enemy, ally, alliance) are no different from other "external" categories. And our subjective categories are only grounded inasmuch as misnaming them has external correlates and consequences (am I really hungry? tired? depressed?).

One-Class Lifelong Learning Approach to Grounding

L. Seabra Lopes and A. Chauhan

Transverse Activity on Intelligent Robotics,
IEETA/DETI, Universidade de Aveiro
Aveiro 3810-193, Portugal
lsl@det.ua.pt and aneesh.chauhan@ieeta.pt

Abstarct

This paper presents a novel approach to tackle the problem of symbol grounding in robotic systems. The focus is on making a visually guided robot agent aware of its surroundings, by learning the names of the objects that can be found in its environment. Humans (Instructor) are used to help the robot agent (Student) ground the words used to refer to those objects. A lifelong learning server, based on one-class learning, was developed. This server is incremental and evolves with the introduction of any new word (class) to the robot, relying on Instructor supervision. The architecture and implementation of the server are discussed in detail. Results obtained following a pre-defined teaching protocol, indicate that the server is capable of incrementally evolving by correcting class descriptions, based on instructor feedback to classification results. The experimental results also suggest that the learning capacity of the system is limited, although the overall performance may still be improved, in particular through the use of complementary object features.

Introduction

Intelligence, as we know, has no concrete definition, but most people will agree that it includes the ability to learn, communicate, remember and inferring from the learned information to successfully use it in previously unknown situations. Researches for long have suggested theoretical models and more recently have developed artificial systems that exhibit similar characteristics, but their success has been very limited. Amongst the reasons cited, the most likely perhaps has been the inability of these systems to interact with their users, which led to the development of the field of Human-Robot Interaction. In recent years, many new methodologies were proposed to integrate and enhance learning and interaction between artificial systems and their surroundings, including their users, leading to encouraging results. Different approaches emerged - Socially Embedded Learning Systems (Asoh et al. 1997), Cognitive Developmental Robotics (CDR) (Chatila 2004; Kelleher, Costello and van Genabith 2005; Wang and Seabra Lopes 2004a), Social Robotics (Breazeal and Scassellati 2000), Robotic Assistants (Graf, Hans and Schraft 2004) etc. The similarity of these approaches lies in their objective of developing user-friendly robotic systems, which are intelligent, flexible, adaptable and able to interact with the humans at language level.

A user-friendly robot must be prepared to adapt to the user rather than requiring the user to adapt to the robot. This includes using/understanding the communication modalities of the user. Spoken-language is probably the most powerful communication modality. It can reduce the problem of assigning a task to the robot to a simple sentence, and it can also play a major role in teaching the robot new facts and behaviors. There is, therefore, a trend to develop robots with spoken language capabilities (Asada et al. 2001; Chatila 2004; Kelleher, Costello and van Genabith 2005; Seabra Lopes 2002; see several reports in Seabra Lopes and Connell 2001a; Seabra Lopes et al. 2005; Weng 2002) Robots are limited by their sensors and, therefore with present state of the art, they won't be able to learn completely to use a natural language. They will be limited to talking about the tasks they are supposed to perform but still, grounding language in the robot's sensors is essential.

Language communication raises the grounding problem (Harnad 1990; Roy, 2004; Seabra Lopes and Connell 2001a and 2001b), i.e. defining symbol meanings based on the agent's perception of the world. Several theoretical and/or experimental works concerned with grounding in robotics have been reported (Billard and Dautenhahn 1999; Billard, Dautenhahn and Hayes 1998;Roy2004, 2005ab; Steels 2001 and 2002b). When the human user wants to talk to the robot about a task to be performed in the physical world, the symbols (words) used in communication must be grounded in the robot's own sensors.

Language grounding (of words, symbols, gestures, sentences *etc.*) is highly dependent on the techniques and methods being used for learning. To close the learning loop in robotics, the most successful approaches (Asada et al. 2001; Asoh et al. 1997; Billard and Dautenhahn 1999; Billard, Dautenhahn and Hayes 1998; Roy 2005ab; Sloman 2005; Steels 2001 and 2002b) have used the paradigm of introducing a human (Instructor) to help the robot

(Student) acquaint with its surroundings. This paper discusses the implementation of a similar approach which will be referred to as *Instructor-Student model*.

Learning new concepts and behaviors with human guidance must be supported by appropriate machine learning algorithms. A learning system in a robot should support long-term (lifelong) learning and adaptation, as is common in animals and, particularly, in humans. For that purpose, the learning system should satisfy several basic requirements (Seabra Lopes and Wang 2002), namely:

- Support opportunistic on-online learning,
- Allow learning to be incremental,
- Support supervised learning, in order to include the human instructor in the learning process,
- Allow concurrent learning, i.e. the ability to handle multiple learning problems at the same time, and
- Include meta-learning capabilities, i.e. the ability to learn which learning parameters are more promising for different problems, ensuring each problem is handled efficiently.

Previous work at Universidade de Aveiro produced a learning tool (LLL – Life-Long Learning server: Seabra Lopes and Wang 2002) that satisfies the above requirements. LLL is based on backpropagation neural networks. In this paper, a new lifelong learning tool, OCLL, based on the One-Class Learning paradigm (Japkowicz 1999; Wang and Seabra Lopes 2004a and 2004b; Tax 2001), is described. The choice of learning algorithm used is based on human ways to approach learning problems. Humans often learn from positive examples only (negative examples being everything else). To implement such single class learning, OCLL specifically uses *Support Vector data description* (SVDD) (Tax 2001).

The rest of the paper is organized as follows. Next Section describes the proposed Instructor-Student model and its main components. Then the paper covers the implementation of the OCLL server. Later, the experiments and the results obtained are discussed. Finally, the last two sections respectively discuss the related research and list the conclusions with the proposed future extension of this work.

System Architecture

The initial research scenario for this project consists of an artificial agent (the student) with an ability to learn the names of real-world object classes, given concrete instances of these classes (using visual sensor feedback) and appropriate teaching actions from a human instructor. The result is the association of class names to their learned sensor-based descriptions. The whole system thus comprises two main components (see Fig. 1), namely the Student (artificial agent) and its World (including the Instructor).

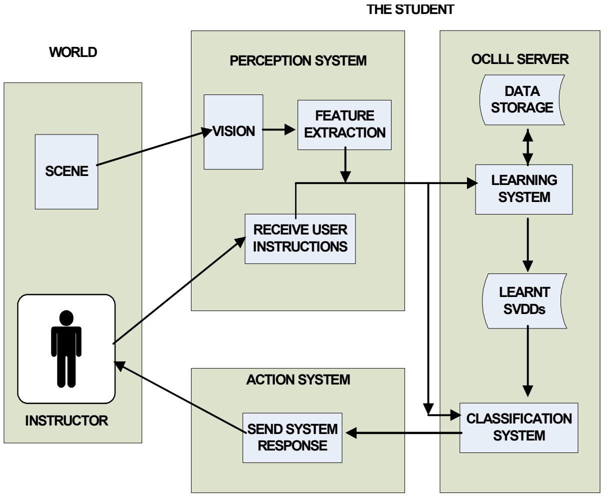


Fig. 1 – IS Model Architecture

The agent architecture itself consists of: a perception system, an internal inference system (OCLL server – for learning and classification) and a limited action system. At present the action system abilities are limited to reporting the classification results back to the Instructor, but in the near future, the action capabilities of the agent will be much extended through the introduction of a robotic arm. Nevertheless, since the current agent perceives and acts upon the physical world, it will also be referred to as robot.

Instructor and the World

The world includes the instructor, a visually observable area and real-world objects (e.g. pen, stapler, mobile, mouse, etc.) that the instructor may whish to teach. The instructor is typically not visible to the robot. The instructor has the key role of communicating with the robot. At present, a linux process has been designed to act as the user interface for communicating with the agent. Using this interface, an instructor can select any object from the robot's visible scene (objects that the instructor him/herself placed there) and perform the following basic actions:

- Teach the object's class name for learning, or
- Ask the class name, which the robot will determine based on previously learned classification knowledge;
- If the class returned in the previous case is wrong, the instructor can send a correction.

The instructor must comply with a basic requirement of the used learning algorithms, namely that a sufficient number of examples must be provided in order for learning to start (the current minimum is 10 examples, since 10-fold cross-validation is performed).

The Student

The student robot currently is a computer with an attached camera (IEEE1394 compliant *Unibrain Fire-i digital camera* is being used). The computer runs the visual perception and learning/classification software as well as the communication interface for the instructor.

The main tasks of the perception system are threefold (shown in Fig. 1). As soon as the perception system receives an instruction (user sends object for either learning or classification), the vision system starts pre-processing (using the functions available in openCV[1]) the whole world image to extract the object selected by the user. Once the user points the mouse on the desired object in the image, an edge-based counterpart of the whole image is generated using *canny algorithm* for edge detection. From this edges image and taking into account the user-pointed position, the boundary of the object is extracted using a region growing algorithm. The boundary image contains all pixels located at the boundary edges of the object. Fig. 2 shows the stated stages of pre-processing to extract the boundary image of the object class *Stapler.* At this point, the instructor can check if the extracted boundary image adequately represents the object and, based on that, decide to use it for learning or classification.

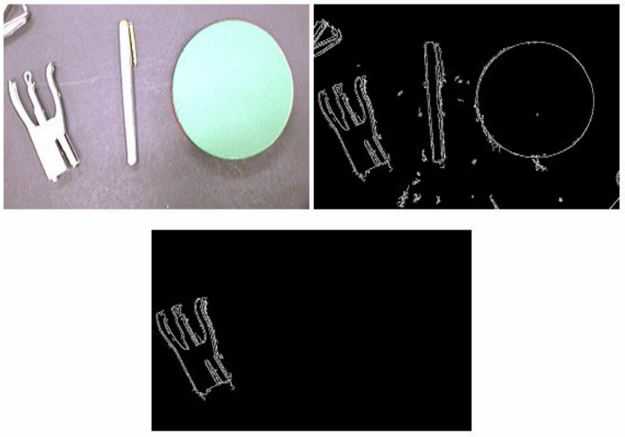


Fig. 2 – Image pre-processing stages in extracting the boundary of the object *Stapler* from the original image. (*Top Left – The Original Scene*; *Top Right – The edges image*; *Bottom: The Boundary image of Stapler*)

Objects should be described to the learning algorithm in terms of a small set of informative features. A small number of features will enable a shorter running time for the learning algorithm. Information content of the features will determine the learning performance. For visual object recognition, it is important that features capture the object's shape and size independently of its position and orientation in the scene. Size, translation and rotation invariance cannot be achieved through widely applied technologies like edge histograms.

To address these requirements, a feature extraction strategy was devised that captures the variation of the distance of boundary pixels to the center of the object. For this purpose, the smallest circle enclosing the object is divided into 36 sections of 10º. Each section i contains a number of boundary pixels with angle θ_i, such that $10\times(i\text{-}1) \leq \theta_i < 10\times i$. The average distance of these pixels to the center of the circle, a radius R_i, is computed. Based on the R_i values, the following features are then computed:

- Radius average, R - the average of all R_i.
- Radius standard deviation, S – again computed over all R_i.
- Normalized radii, r_i – this is a vector containing the normalization of all R_i values with respect to the average radius R, but rotated in order to make it orientation-invariant. It is computed in two steps:
  - First, the normalized values are computed as $r_i = R_i/R$.
  - Then, all values are rotated in the vector in such a way that highest values are at the center, according to a mobile average measure. Specifically, a given section i will be at the center if the average of all values r_j, with $j = i\text{-}4, \ldots, i\text{+}4$, is the highest.
- Normalized radius average, r - the average of all r_i
- Normalized radius standard deviation, s – again computed over all r_i.
- Block averages, B_k – the normalized radius values are divided into six blocks; for each block k, where k=1, .., 6, B_k is defined as the average of all r_i values, for $i = (k\text{-}1) \times 6\text{+}1, \ldots, k\times 6$.

This feature extraction strategy provides 46 features to the learning algorithm. The first 2 features (R and S) provide size information. The remaining 44 normalized features faithfully capture the boundary of a segmented object, invariant of its size, translation and rotation. Fig. 3 shows a scene with three objects (a stapler, a pen and a ball). Fig. 4 shows the normalized radius vectors for the three objects.

The communication between student robot and human instructor is supported by the perception and action systems (respectively for instructor input and robot feedback). At present, the communication capabilities of the robot are limited to reading the teaching options (teach, ask, correct) in a menu-based interface and displaying classification results. In future, simple spoken language communication will be supported.

[1] http://www.intel.com/technology/computing/opencv/index.htm

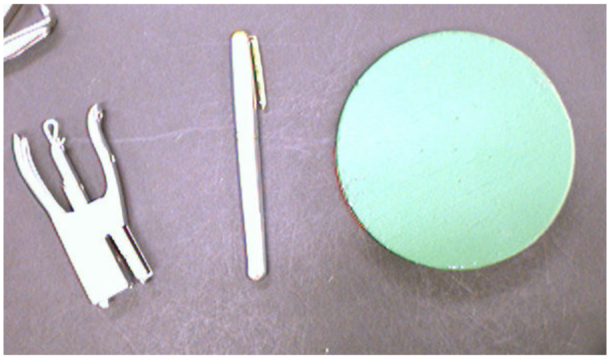


Fig. 3 – Scene with three objects (S*tapler, Pen and Ball*)

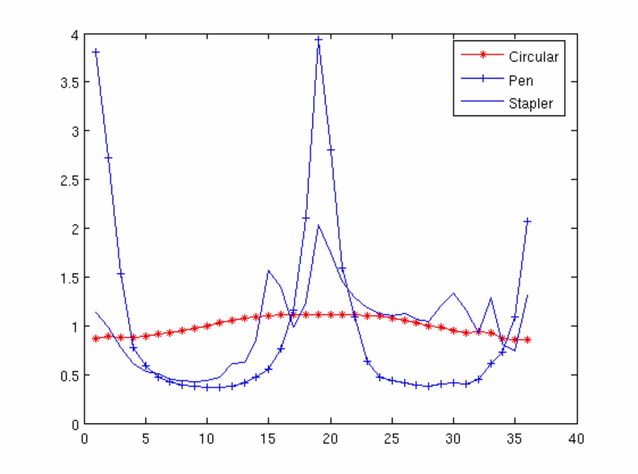


Fig. 4 – Normalized and rotated radius feature vectors for the three objects in Fig. 2

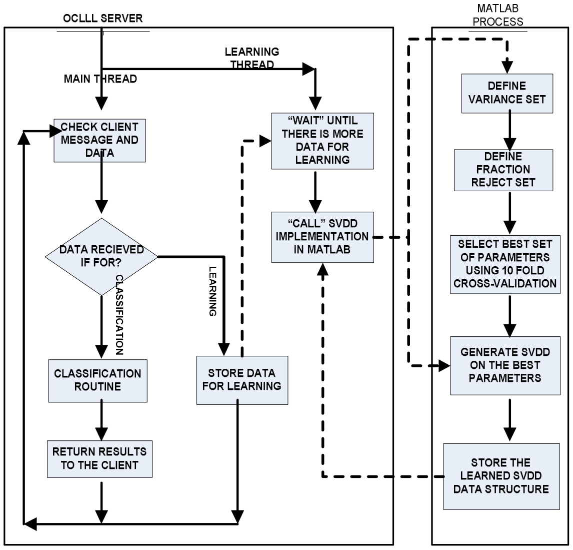
Learning and classification capabilities are provided to the agent using a client-server approach. A new learning server, OCLL, performs concurrent incremental on-line learning as well as classification as requested by the user. OCLL itself has been implemented as a separate process and it's overall importance in this work has led to devote the next Section for presentation of its design and functionality.

One-Class Lifelong Learning Server

As mentioned in first Section, the design of LLL (Seabra Lopes and Wang 2002) is the basis for the proposed learning system, OCLL (One-Class Lifelong Learning).

OCLL includes two processes, namely the OCLL server process and a MATLAB-based auxiliary process, both running in Linux. The server process divides into two concurrent threads (Fig. 5).

The *main thread* supports the communication with the learning client (the agent) and also runs the classification routines, and the *learning thread*. Having separate threads for learning and classification allows the OCLL server to execute client requests for classification and save new data for learning, while the learning thread concurrently handles learning.


OCLLL SERVER
MAIN THREAD
LEARNING THREAD
CHECK CLIENT MESSAGE AND DATA
DATA RECIEVED IF FOR?
CLASSIFICATION
LEARNING
CLASSIFICATION ROUTINE
STORE DATA FOR LEARNING
RETURN RESULTS TO THE CLIENT
"WAIT" UNTIL THERE IS MORE DATA FOR LEARNING
"CALL" SVDD IMPLEMENTATION IN MATLAB
MATLAB PROCESS
DEFINE VARIANCE SET
DEFINE FRACTION REJECT SET
SELECT BEST SET OF PARAMETERS USING 10 FOLD CROSS-VALIDATION
GENERATE SVDD ON THE BEST PARAMETERS
STORE THE LEARNED SVDD DATA STRUCTURE

F
ig. 5 – Flow chart of OCLL (dashed lines indicate sequencing across different threads or processes)

Learning

The basic one-class learning algorithm used in OCLL is SVDD (Support Vector Data Description (Tax 2001)). In the normal case, SVDD is trained only with positive instances of the class. It tries to form a hypersphere around the data by finding an optimized set of support vectors. These support vectors are data points on the boundary of a hypersphere whose center is also determined through optimization. The hypersphere center is taken as representing the center of the data distribution itself. The optimization process, that determines the center and support vectors, attempts to minimize two errors:

- Empirical error – percentage of misclassified training samples.
- Structural error – defined as R^2, where R is the radius of the hypershpere, must be minimized with respect to constraints $\|x_i - a\|^2 \leq R^2$, for every training object x_i.

In the ideal case (no noise), all training objects can be included in the hypersphere and therefore the empirical error will be 0. In practical applications, however, this may result in over-fitting. Better results can be obtained with not much extra computational expense if a kernel is introduced to get a better data description (Tax 2001).

In addition, if a set of outliers (negative instances) is known, it adds to the performance of SVDD since, during optimization, an even tighter boundary around the data can be obtained. From (Tax 2001), the final error L (which includes both empirical and structural error) to be optimized is given as:

$$L = \sum_i \alpha_i K(x_i, x_j) - \sum_{i,j} \alpha_i \alpha_j K(x_i, x_j)$$

with the following constraints on Lagrange multipliers:

$$0 \le \alpha_i \le C \,,\ \forall i$$

$$\alpha_i \ge 0 \,, \sum_i \alpha_i = 1 \text{ and } a = \sum_i \alpha_i x_i$$

C gives the tradeoff between volume of the description and the errors. The kernel K maps the data into a more suitable space. Although the choice of kernel is data dependent, in most applications a Gaussian kernel will produce good results. (Tax 2001) gives a thorough explanation of the performance benefits of this kernel. It is defined as:

$$K(x_i, x_j) = \exp\left(\frac{- \| x_\mathrm{i} - x_\mathrm{j} \|^2}{SIGMA^2} \right)$$

where *SIGMA* is the variance of the kernel.

An implementation of SVDD for MATLAB is in the publicly available dd-tools toolbox[2] (Tax 2005). The magic parameters to be supplied to the SVDD algorithm are *FRACREJ* (percentage of the training objects that can be considered as outliers for better boundary description and over-fitting avoidance) and *SIGMA* (the variance of the Gaussian kernel used to map the data into a more suitable space).

OCLL performs 10-fold cross-validation for determining appropriate values for *FRACREJ* and *SIGMA*. In the current implementation of OCLL, *FRACREJ* ranges from 1% to 11% of the training data with an interval of 2% (in total 6 values of *FRACREJ*). Since object features can vary over a very wide range, *SIGMA* is divided into 10 parts on a logarithmic scale, where

$$\min \left\| x_i - x_j \right\|^2 \le SIGMA^2 \le \max \left\| x_i - x_j \right\|^2$$

For a thorough explanation on the range of values and the choice of *SIGMA* refer to (Tax 2001).

In total, 66 parameter combinations are evaluated through cross-validation. To find the best combination out of these parameters, a performance measure combining precision and recall values is used. It is defined as:

$$\frac{2 \cdot P \cdot \mathrm{R}}{P + \mathrm{R}}$$

where, $P = CTP/TP$ is precision, $R = CTP/TE$ is recall, *CTP* is the number of correct target predictions, *TP* is the number of target predictions and *TE* is the number of target examples. Comparing the average performance over 10 folds for all the learned classifiers, the best pair of *FRACREJ* and *SIGMA* is determined. The final class description is trained on those particular values.

The OCLL server is a C++ program which runs as a single process divided into two threads, as mentioned above. SVDD runs in a separate MATLAB-based process on request of the learning thread of the OCLL server. The main server thread saves any new training data in a file and informs the learning thread to process it. When there is no new data to process, the learning thread is waiting on a semaphore. When new data is received, the learning thread calls SVDD on the MATLAB process and waits until SVDD returns. The learned class description is stored in a file by the MATLAB process.

As mentioned previously, learning is incremental and supervised. Thus, when an object gets misclassified, the instructor has an option of providing the correct class, so that class descriptions can be improved.

Misclassification in this case broadly is of two types: either the object is inside the hyperspheres of several classes and the classification system chose the wrong class; or the object is outside the hyperspheres of all known classes. Given a correction from the user, OCLL will identify and retrain the class descriptions needing correction. Specifically, OCLL will add the misclassified object as outlier for retraining the classes whose hyperspheres contain the object.

Classification

Inability of MATLAB to perform multithreading limited the use of dd-tools to learning only. Therefore, classification is done in the OCLL server process itself, using the learned SVDD class descriptions. These class descriptions provide the support vectors and their respective α and *SIGMA*, for all the SVDDs. In the standard application of SVDD class descriptions, the criterion for classifying any new object z as target is:

$$\sum_i \alpha_i \exp\left(\frac{- \| z - x_i \|^2}{SIGMA^2} \right) > \frac{1}{2}(B - R)^2$$

where, $B = 1 + \sum_{i.j} \alpha_i \alpha_j K(x_i, x_j)$ and R is the radius

In OCLL, to handle multiple class candidates maximizing classification results, a more suitable criterion has been derived from the above mentioned inequality. In fact, using original SVDD classification criterion, more than one or none of the classes may get identified as being target, and is impossible to tell the real class. The following measure is therefore introduced:

$$\mathrm{NDC}(\,z\,) = \frac{\sqrt{B - 2\left(\sum_i \alpha_i \exp\left(\frac{- \| z - x_i \|^2}{SIGMA^2} \right) \right)}}{R}$$

NDC (Normalized Distance to the Center) is the distance of an object z from the center of the hypersphere given as a

[2] http://www-ict.ewi.tudelft.nl/~davidt/dd_tools.html

fraction of the radius R of the class. It captures the relative closeness of the object from the center of each class and, therefore, enables to compare its membership to different classes. Lower the value of NDC for a particular class, closer is the object to the center of that class. Of all the classes that have been learned, the one with the lowest NDC(z) will be considered as the best class candidate for object z. However, if the lowest value of NDC(z) is greater than 2.0, the object is considered to be clearly outside any of the current class descriptions and thus not belonging to any class.

Experiments and Discussion

Lifelong learning in the context of instructor-student systems requires evaluation methodologies much more complex than classical supervised learning algorithms. The following aspects should be considered:

- Evolution: Ability to modify the system to learn new concepts;
- Recovery: The system performance will mostly deteriorate at the introduction of any new concept. The time spent in system evolution until correcting/ adjusting all current concept descriptions, defines recovery. This learning is based on student mistakes and corresponding instructor feedback.
- Break Point: Inability of the system to recover and evolve, when a new concept is introduced

The learning capabilities of the student described above were evaluated taking into account these aspects. For easy comparison with other similar systems, including future versions of the described system, a precise experimental teaching protocol is proposed, as described in Fig. 6.

```
introduce $Class_0$;
$n$ = 1;
repeat
{
        introduce $Class_n$;
        $k$ = 0;
        repeat
        {
                Evaluate and Correct classifiers;
                $k$++;
        } until ( (average precision > precision threshold
                  and $k \geq n$) or
                (user sees no improvement in precision) );
        $n$++;
} until (user sees no improvement in precision).
```

Fig. 6 – Teaching protocol used for performance evaluation

For every new class introduced by the instructor, the average precision of the whole system is calculated by performing classification on all classes for which data descriptions have already been learned. Average precision is calculated over the last 3×n classification results (n being the number of classes that have already been learned). A precision of a single classification is either 1 (correct class) or 0 (wrong class). When the number of classification results since the last time a new class was introduced, k, is greater or equal to n, but less than 3×n, the average of all results is used. The criterion to indicate that the system is ready to accept a new object class is based on the precision threshold, which in these experiments was set to 0.667. However, the evaluation/correction phase continues until a local maximum is reached.

Experiments were conducted according to the protocol presented above. New object classes were introduced in the following sequence:

Pen – 5 different pens were used for teaching
Stapler – 1 object of this class
Circular - 2 circular shaped objects
Mobile – 3 objects
Key – 2 objects
Box – 2 objects
TiltedCup – 1 object
Rubber – 1 object
CoffeeCup – 1 object
StapleRemover – 1 object

Obtained results are graphically presented in Figs. 7 and 8. They respectively show the evolution of classification precision and learning efficiency against the number of question/correction iterations. Efficiency is defined as the ratio between the obtained classification precision and the precision of a random classifier.

Classification for an object of the first class (*pen*) was correct in the first attempt. This means a single iteration was enough to reach a precision of 100%. Similarly, for the second class, in the minimum number of iterations (at least n iterations for n classes, as defined above) maximum precision was obtained. On the introduction of the third class (*circular*), although starting at 100%, precision continuously dropped down to 50.3%, before it recovered to a value above the threshold. For the next classes, the pattern remained similar: at the introduction of the new class, there is a sharp initial drop in precision followed by recovery after a number of question/correction iterations.

On the introduction of the 10th class (*staple remover*), precision started at 100%, then dropped down to values between 20% and 50%, remaining like that for many iterations. As can be seen at the end of the graphs of Figs. 7 and 8, classification precision and learning efficiency seem to have stabilized. No considerable improvement in these measures could be noticed over time. Here the instructor concluded that, on the given set of features and for the above set of classes, the learning capacity of the student had reached its break point.

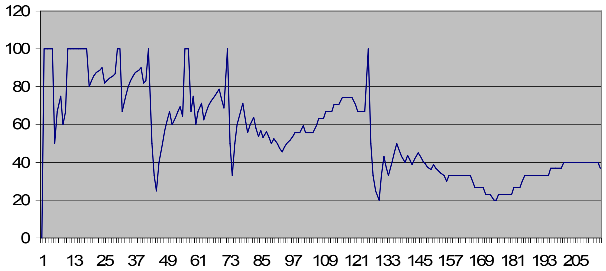


Fig. 7 – Evolution of classification precision versus number of question/correction iterations

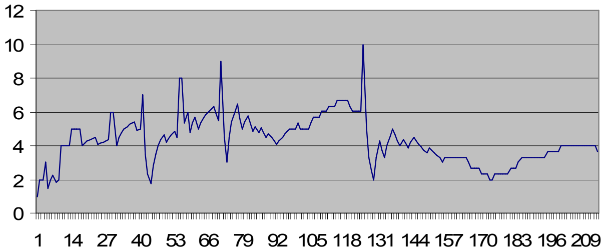


Fig. 8 – Evolution of learning efficiency versus number of question/correction iterations

It should be noted that, most of the time, learning efficiency is above 2.0, and its average is 4.3. This means that precision is clearly above the random classifier precision throughout the whole experiment. Another important observation can be made. While classification precision seems to follow a decreasing trend as the number of introduced classes increase, learning efficiency follows an increasing trend almost until the break point.

In OCLL, each correction leads to the introduction of: an extra outlier for each of the classifiers that misclassified the object (NDC<1.0) and an extra positive example for the right classier (if NDC>1). For each new class introduced, Fig. 9 shows the total number of outliers and positive examples required by the system to achieve the precision threshold (except for the last object). It can be seen from this figure that introduction of last two classes introduces considerably high number of outliers as well as positive examples. In comparison to the last 8 classes, the number of misclassifications by the system after the 9th class was introduced shows a substantial increase. In other words, it became increasingly difficult for the system to reach the precision threshold. Eventually, on the 10th object the system reached its break point. A collective analysis of Figs. 7, 8 and 9 shows an association between the number of iterations required for reaching the precision threshold, and the number of outliers and positive examples needed before the system reached the precision threshold. For the first 8 classes, the system shows fast evolution of precision and efficiency. And the number of examples and outliers that was necessary to add after introducing those classes are also relatively few. On the other hand, for the 9th and 10th classes, it took a long time to reach the precision threshold (not achieved for the 10th class) and the number of outliers and target examples introduced to the system shown a sharp increase.

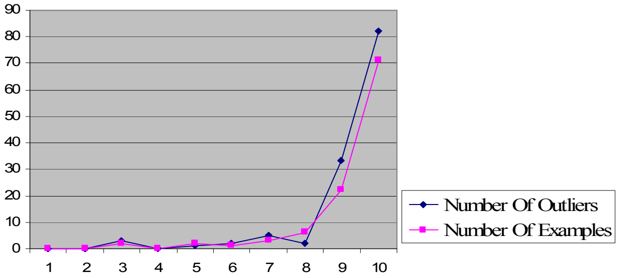


Fig. 9 – Number of outliers and positive examples added after each new class was introduced

Table I shows the number of outliers and positive examples stored for each class after the system reached the break point. As can be observed, the number of outliers in some classes (*Box* and *TiltedCup*) far outweighs the number of target examples.

Table I – Final number of target and outlier examples in each class

Object class	Target	Outliers
Pen	17	18
Stapler	24	1
Ball	30	7
Mobile	27	1
Key	22	1
Box	14	49
TiltedCup	17	40
Rubber	24	1
CoffeeCup	19	1
StapleRemover	14	9

One of the conclusions of Tax (2001) is that, for one class classifiers, the introduction of a small number of outliers in the training data results in better class descriptions. He however emphasizes, introduction of too many outliers eventually deteriorates the learning performance. For the first introduced classes, classification precision improved very fast, which supports the idea that, when using few outliers, SVDD class descriptions do describe the data better than just with positive instances. However, in the long run, the number of outliers in the training data may become higher than the number of target examples. This may also limit the number of classes that the system will be able to learn.

Related Work

One of the recent important works related to language grounding in robot/agent domains is that of Luc Steels. He introduced the concept of adaptive language games between robot agents and other agents or humans. Language games help solve the grounding problem by creating a strong context that constrains the possible meaning of words, thus making it easier for the robot to learn new words (Steels 2001). As can be imagined, the success of these language games is highly dependent on the learning approach used. More recent works have focused on learning algorithms implemented for language games. Steels (2001) discusses evolutionary language games for grounding symbols. De Jong and Steels (2003), on the other hand, develop on that idea and describe a learning algorithm for communication development and representation grounding. They have used a greedy algorithm based on Boltzmann distribution. They also provide an evaluation criterion for evolutionary symbol grounding systems, designed especially for their system. Lewin (2002) proposes a competitive learning algorithm for concept definition and formation, based on symbol grounding using Luc Steels' language games.

The line of work initiated by Steels follows a bottom-up grounding approach in which symbolic representations and their perceptual links are established by the same process. It can, therefore, be used to create an entirely new language (emergence). However, the approach has also been used to ground natural language symbols.

Roy and co-workers have also been working on natural language grounding (Roy et al., 2004; Roy, 2005ab). Most of this work has been on spoken language grounding. Roy (2005a) describes a system to maintain a mental-model of its changing environment by coupling vision to language grounding. In this paper, their focus was on spatial grounding, so as to help their visually guided robot (*Ripley*) understand its environment. Roy (2005b) discussed the requirement for grounding context dependent shift of word meanings. He also presented and implemented an action-perception system for grounding context dependent nouns.

Another major initiative is the CoSy project (*Cognitive Systems for Cognitive Assistants*). In this project, the first suggested step towards robot cognition is developing the ability for categorizing objects (Sloman 2005). Similar to the work presented in this paper, they also emphasize importance of incremental learning in the visual domain, as well as introduction of human tutors for robot learning. However, to the present moment, they have not presented an implementation of their proposed learning model.

Another aspect emphasized in this paper is the need for online incremental learning (Steels 2002ab; Weng 2002; Lewin 2002). Seabra Lopes and Wang (2002) list a set of basic characteristics for a system to be considered incremental and have developed a learning server based on that. Classical learning algorithms don't follow the human way of approaching a learning problem. This led us to explore one-class learning (Tax 2001; Japkowicz 1999) and especially SVDD (Tax 2001; Wang and Seabra Lopes, 2004ab). Tax surveys various other approaches to one-class learning and compares performance of SVDD over the rest.

Conclusion and Future Work

This paper presents and discusses an online, incremental learning module to support an instructor-student system for grounding object category names. This learning module is referred in this paper as OCLL (One-Class Lifelong Learning). Preliminary work was carried out to implement an initial version of OCLL and apply it for grounding English names of several office objects. An experimental protocol has also been proposed and used to evaluate the performance of OCLL.

From the conducted experiments, it can be concluded that the learning system has the ability to incrementally evolve to include each new class presented. Classification precision follows a pattern of sharp fall at the introduction of new objects, but quickly recovers by correcting the classifiers so that the boundaries of the class descriptions get modified to separate out all the different classes. As the number of classes increases, the training becomes more difficult and some class descriptions have to be corrected many times before the precision threshold is achieved. Although there seems to be a limit to the learning capacity of the system, there are still many ways in which it can be improved. It is therefore hard to predict how far this approach can be taken. In any case, it definitely seems to be a promising research direction.

Before improving the learning system, we intend to integrate it in a more sophisticated agent, which will include a robotic manipulator. In that new scenario, the action system of the agent is much extended. In particular, it will be able to manipulate the objects whose names were previously taught by the instructor and grounded by the student. Basic spoken language communication between the instructor and the student will also be supported.

References

Asada, M.; MacDorman, K. F.; Ishiguro, H. and Kuniyoshi, Y. 2001. Cognitive Developmental Robotics as a New Paradigm for the Design of

Humanoid Robotics. *Robotics and Autonomous Systems*, 37:185-193. Cambridge, MA: Elsevier.

Asoh, H.; Motomura, Y.; Asano, F.,; Hara, I.; Hayamizu, S.; Itou, K; Kurita, K. and Matsui, T. 1997. Socially Embedded Learning of the Office-Conversant Mobile Robot Jijo-2. In *Proceeding of IJCAI-97*, 880-885.Nagoya, Japan: IJCAI.

Billard, A. and Dautenhahn, K. 1999. Experiments in Learning by Imitation - Grounding and Use of Communication in Robotic Agents. *Adaptive Behavior* 7(3): 411-434.

Billard, A.; Dautenhahn, K. and Hayes, G. 1998. Experiments on human-robot communication with Robota, an imitative learning and communication doll robot. In *Proceedings of SAB98 Workshop Socially Situated Intelligence*, Zurich.

Breazeal, C. and Scassellati, B. 2000. Infant-like social interactions between a robot and a human caregiver. *Adaptive Behavior* 8(1):49-74.

Chatila, R. 2004. The Cognitive Robot Companion and the European 'Beyond Robotics Initiative. In *Proceedings of sixth EAJ International Symposium on Living with Robots*. Tokyo, Japan: International Symposium on Living with Robots.

de Jong, E. D. and Steels, L. 2003. A distributed Learning Algorithm for Communication Development. *Complex Systems*, 14(4).

Graf, B.; Hans, M. and Schraft, R. D. 2004. Mobile Robot Assistants – Issues for Dependable Operation in Direct Cooperation with Humans. *IEEE Robotics & Automation Magazine*, 11(2):67-77.

Harnad, S. 1990. The Symbol Grounding Problem, *Physica D*, 42:335-346.

Japkowicz, N. 1999. Are We Better-off without Counter Examples?, *Proc. First International ICSC Congress on Computational Intelligence Methods and Applications* (CIMA-99), 242-248. N.Y. ,USA: CIMA.

Kelleher, J. D.; Costello, F. and van Genabith J. A. 2005. Dynamically Updating and Interrelating Representations of Visual and Linguistic Discourse. *Artificial Intelligence Journal* 67 (1-2):62-102.

Lewin, M. 2002. *Concept Formation and Language Sharing: Combining Steels' Language Games with Simple Competitive Learning*. Master thesis, University of Sussex.

Prince, C. G. and Mislivec, E. J. 2001. Humanoid Theory Grounding, *Proc. International Conference on Humanoid Robotics*, 377-383. Tokyo, Japan.

Roy, D., K.-Y. Hsiao and N. Mavridis (2004) Mental Imagery for a Conversational Robot, *IEEE Trans. Systems, Man and Cybernetics – Part B: Cybernetics*, 34 (3), 1374-1383.

Roy, D. (2005a) Semiotic Schemas: A Framework for Grounding Language in Action and Perception, *Artificial* Intelligence, 167(1-2):170-205.

Roy, D. (2005b) Grounding Words in Perception and Action: Computational Insights. *Trends in Cognitive Sciences*, 9(8):389-396.

Seabra Lopes, L. 2002. Carl: from Situated Activity to Language-Level Interaction and Learning. In *Proceedings of IROS'02*, 890-896. Lausane: IROS.

Seabra Lopes, L. and Connell, J. H. eds. 2001a. *Semisentient Robots*. IEEE Computer Society, Special Issue of *IEEE Intelligent Systems*, 16(5).

Seabra Lopes, L. and Connell, J. H. 2001b. Semisentient Robots: Routes to Integrated Intelligence, *IEEE Intelligent Systems*. 16(5):10-14.

Seabra Lopes, L.; Teixeira, A. J. S.; Quinderé, M. and Rodrigues, M. 2005. From Robust Spoken Language Understanding to Knowledge Acquisition and Management. *Proceeding of Interspeech'2005*, 3469-3472. Lisboa, Portugal: Interspeech.

Seabra Lopes, L. and Wang, Q. H. 2002. Towards Grounded Human-Robot Communication. In *Proceeding of 11th IEEE International Workshop on Robot and Human Interactive Communication* (ROMAN'2002), 312-318. Berlin, Germany: ROMAN.

Sloman, A. 2005. CoSy: Initial Report on Simplest Scenarios and their Requirements. Information Society Technologies, University of Birmingham.

Steels, L. 2001. Language Games for Autonomous Robots. IEEE Computer Society, *IEEE Intelligent Systems*. 16(5):16-22.

Steels, L. 2002a. Grounding Symbols through Evolutionary Language Games. In Angelo Cangelosi and Domenico Parisi, *Simulating the Evolution of Language*, 211-226. London: Springer Verlag.

Steels, L. 2002b. Language Games for Emergent Semantics. *IEEE Intelligent Systems*, 17(1):83-85.

Wang, Q. and Seabra Lopes, L. 2004a. One-Class Learning for Human-Robot Interaction. *Emerging Solutions for Future Manufacturing Systems: IFIP TC 5 / WG 5.5 Sixth IFIP International Conference Information Technologies for Balanced Automation Systems in Manufacturing and Services*, Springer, 489-498.

Wang, Q. and Seabra Lopes, L. 2004b. Visual Object Recognition through One-Class Learning, *Image Analysis and Recognition: Proceedings of International Conference ICIAR 2004*, Part I, LNCS 3211, Springer, 463-469.

Weng, J. 2002. Theory for Mentally Developing Robots. In *Proc. International Conference on Development and Learning*. Cambridge, MA.

Tax, D. M. J. 2001. *One Class Classification*. PhD Thesis, Delft University of Technology.

Tax, D. M. J 2005. *DD Tools – The Data Description Toolbox for MATLAB. Version 1.4.1*, T.U. Delft.

Psychological reality of stable states: assessing stability in psycholinguistic experiments

Joanna Rączaszek – Leonardi
Faculty of Psychology
Warsaw University, Poland

During the last decade the view of language as a dynamical system has become sharper, due mainly to the modeling work pointing to dynamical forces underlying emergence (and maintenance) of symbols and their structures. However, even though most of the work concerns human communication and interaction, the input of cognitive psychologists and psycholinguists into this endeavor is rather limited. This is probably connected to the elusive nature of human information processing and difficulties with finding observables testifying to the internal states of mind. On the other hand, the information processing approach had to deal with similar problems and – even though now it is often judged harshly – over last 50 years has generated thousands of experiments, definitely increasing our knowledge of human cognition. Thus, we may hope that finding a new set of observables for the theoretical concepts we use in dynamical language description is not impossible, it is just difficult.

The present work is based on a general assumption that "language as a system of symbols" is a "linguistic" description (or a linguistic mode, see e.g., Pattee, 1987, 2001; Carriani, 2001) of a much larger, dynamical system that changes on several time-scales. This framework derives mainly from Howard Pattee's work on information in biological systems, and seems to be quite compatible with the conceptualizations of the relation between symbolic and dynamic developed (and used in simulations) by theorists working on symbol grounding problem (e.g., Harnad, 1990; Steels and Belpaeme, 2005). Applying Pattee's approach to a linguistic system allows seeing symbols of language as having a constraining role, controlling the dynamics over different time-scales, while meaning of symbols is understood as their function in the system (Pattee, 1987).

There are at least three time-scales at which the constraining role of symbols can be seen:

- the time-scale of immediate communication (milliseconds, seconds, minutes), where symbols are generated to express what speaker means, and where symbols are understood, i.e., conceptual understanding unfolds
- the time-scale of ontogeny, (months) where language shapes a culturally specific conceptual system
- the time-scale of cultural evolution (hundreds of years), where language serves as a repository of culturally important constraints

Recently much of psycholinguistic work has been done on the "middle" scale, showing how, in a developing cognitive system, differences between languages may lead to differences in development of cognitive categories (see e.g. Bowerman & Levinson, 2001, or Gentner and Goldin – Meadow, 2003).

Research pertaining to the scale of language perception and production focused mainly on single words, or even syllables in artificial situations (see e.g., Tuller, et. al., 1994). They are very valuable because they show e.g., that word understanding can be pictured as evolving to a stable state, and they emphasize the importance of the concept of stability in their explanations, but unfortunately the perceptual situations they study are rather far removed from natural situations of language processing.

The aim to participate in the workshop is (at least) three-fold:

1. To present an attempt at measuring stability changes related to the **on-line process** of contextually relevant understanding of words presented in sentences. A psycholinguistic experiment will be presented, in which general category words were embedded in biasing contexts. Variability of response times in a Cross-Modal Lexical Decision task during the contextual adaptation of category names was measured. The pattern of variability shows an interesting relationship to the pattern of response times, namely a marked increase in variability just before the decrease in response times, and then decrease in variability after a purported "settling into a stable state".
2. To discuss the usefulness of other possible stability measures of on-line, language-related processes, such as demonstration of hysteresis effects in sequentially presented stimuli, semantic variability of free associations, fidelity of repetition (or paraphrasing) or translation.
3. To learn about other possibilities of stability measures on various time-scales of language dynamics, and, last but not least, to learn in what way experimental psycholinguistics may aid in showing psychological reality of processes proposed in dynamical explanations of language.

Bibliography

Bowerman, M. & Levinson, S. (eds.) (2001). *Language Acquisition and Conceptual Development*. Cambridge: Cambridge U. Press.

Cariani, P. (2001). Symbols and dynamics in the brain *Biosystems*, 60:59-83.

Gentner, D., Goldin-Meadow, S. (eds.) (2003). Language in Mind: Advances in the Study of Language and Thought.

Harnad, S. (1990) The Symbol Grounding Problem. *Physica D* 42, 335-346

Pattee, H.H. (1987). Instabilitites and Information in Biological Self-Organization. In: F.E. Yates (ed.) Self-organizing Systems: The Emergence of Order. Plenum Press, NY, London.

Pattee, H.H. (2001). The physics of symbols: bridging the epistemic cut. *Biosystems,* 60:5-21

Steels, L., Belpaeme, T. (2005). Coordinating Perceptually Grounded Categories through Language: A Case Study for Colour. *Behavioral and Brain Sciences*, 28(4), 469-89.

Tuller, B., Case, P., Kelso, J.A.S., Ding, M. (1994). The nonlinear dynamics of categorical perception. Journal of Experimental Psychology: Human Perception and Performance, v. 20(1), 3-16.

The Fundamental Role Of Externally Mediated Interactions In Symbolic Interaction: Inverting the Symbol Grounding Problem

Norman Steinhart
University of Toronto
norman.steinhart@utoronto.ca

This paper will invert the usual approach to the Symbol Grounding Problem (SGP) and explore the theoretical and experimental evidence that externally mediated interactions play a fundamental role in the generation of symbolic processes. Some fundamental functions of symbols will be considered first: 1) They allow humans to transcend their individual egocentric perceptions and actions and find 'common ground' between people separated by different spaces and times. 2) The same word or phrase offers a range of functions and meanings and so symbolic interaction dynamically integrates the context of use to coordinate cognition.

However, I will show why some of the important theories of symbol meaning cannot provide the grounding to support the functions of symbols mentioned above. If one assumes categories are acquired innately or even prelinguistically, (Harnad) because of their egocentric origins and their rigid criteria, they cannot be used to ground new symbol meanings or use. While recent artificial agent models of language acquisition (Cangelosi) improve the grounding process by associating language with activity, the actions learned are often decontextualized, internallyreferenced movements (e.g. joint positions) and so these robots are 'too specifically grounded' to use symbols in human ways. Glenberg's proposal that cognition is adaptive because it facilitates coordination of action provides motivation for his indexical hypothesis as he has examined how language can 'index' the role of object affordances. But because he focuses on isolated bodyobject interactions his approach is still too limited to ground more abstract or metaphorical language use. Simply continuing to focus on the isolated individual body and its movements within a very sparse environments will yield limited results; this doesn't reflect the 'environment of language use' in modern humans, who are constantly interacting through artifacts, learning technical/cultural skills, and adapting with language to live an almost completely mediated life (Tomasello). To understand symbols, we must consider to what degree human activity is dependent on culturally extended functions outside the body.

External mediation involves interacting through artifacts, landmarks, skilled actions, and social customs that are 'external to' or independent from the subject and object of activity, (e.g. using cutlery to eat) and so utilizes some aspect of the environment to enhance success of interactions. Culturally mediated activity extends human physical and cognitive function to perceive and act in novel ways we cannot match with direct interactions, and can be considered hybrids of sociocultural and biological functions (Cowley). If we consider that most human perceptions and learned actions are already extended outside the body through cultural mediation, then the environment that symbols function within is 'already grounded' ! Therefore this model inverts the traditional view that internally formed categories precede language and proposes the source of the symbol grounding resides in preexisting extended biocultural activity, making language a secondary form of externally mediated interaction. Since it is unlikely that symbols from innate cognition could be useful to humans who live increasingly vicarious mediated experiences, evidence for language as a candidate to dynamically integrate and coordinate the more concrete forms of mediation, will then be explored.

Previous experimental work has examined the difference between direct and mediated interactions in perception and cognition. The incorporation of artifacts into the body schema (Iriki) extends the interface of perception and action from the usual bodyenvironment boundary to the more distal artifactenvironment interface. This provides a challenge and opportunity for the artifact user to switch from an egocentric to an allocentric or objectbased perspective that utilizes different neural pathways. (Goodale). Since the allocentric but not the direct egocentric perceptionaction system was shown to be accessible through language, this suggests that language helps reorient the observer to share the artifact user's perspective as they both consider the public, common distal bodyartifact interface as a new shared frame of reference.

The last major section of the paper will examine the current experimental approach to test an External Mediation Model of Symbol Grounding. The basic question considered was: can a significant interaction between symbolic and external, nonsymbolic forms of mediation be demonstrated? We hypothesized that there would be a significant effect between these two processes. Therefore manipulating either the linguistic or nonlinguistic factors should significantly affect the subject's responses to the other component, whereas an ungrounded or internal/embodied model of language would predict a null effect since there is a 'core' meaning of words that is predetermined, fixed, and unaffected by culture or context.

Preliminary language usage studies supported the hypothesis that extralinguistic 'grounding factors' are interacting with the symbol processing and were used to choose word pairs. For instance, the words near/far are traditionally defined in terms of distance or time amounts. However, the analysis of actual word usage revealed that a task that requires days or weeks to complete maybe described as nearly complete, while a task that is expected to take hours may be termed 'far from finished' , suggesting that these words prompt the speaker or listener to consider the potential ability and/or degree of difficulty to access some object, state, location or achievement, perhaps considering the task familiarity, object affordances, social interactions. Further work with Come/Go examined how these words prompt listeners to understand the subject's 'movement' as a change from the perspective of a person grounded in a location or situation used cognitively as an External Frame of Reference. Typical uses of Above/Below helps listeners predict the relative 'force' of each person's actions on other people by simulating the differences in affordances that emerges from being situated at various vertical positions within a gravitational field. This grounding process can be evoked to choose actions within complex social situations. (e.g. "he is above me in the company hierarchy" provides help in judging the chance of success of various activities either person undertakes)

Phrases utilizing contrasting word pairs such as near/far, come/go, up/down and above/below were then used as linguistic stimuli to interact with pictures chosen to express possible nonlinguistic aspects typically found from the analysis stage

1) If first given linguistic phrases, we tested if and how this creates a bias when people choose picture situations with different external mediating factors such as leftward or rightward movement, different vertical points of reference or methods of movement.
2) Then the order of presentation was reversed and subjects initially given picturebased situations, were tested to detect if and how this influenced language choices to communicate this situation. (to explore whether people spontaneously and systematically choose certain nonlinguistic factors that bias language choices)

Results to date have supported the hypothesis of interaction between linguistic and external factors, and as well suggests differences in the tendency of various external processes to affect word choices. Future research possibilities will be discussed, including an artifactbased adaptation of Glenberg's ACE experiments and tests of the potential of language to coordinate action between several subjects. This model of mediated symbol may be applied to the research with artificial agents, reconceptualizing actions in terms of goals mediated with objects by agents set in a rich cultural environment (like children) that could improve their quality of language learning.

In effect, this ongoing research is attempting to investigate the hypothesis that –although it seems to be confined to the head or body we can consider that symbolic interaction is grounded by the same processes of external mediation that culturally 'grounds' the brain and body to the environment. Therefore, the author proposes that significant progress in understanding human symbol–mediated cognition can be made if we invert the SGP and ask: how do we symbolize our externally mediated processes to form common perceptions, actions and goals with others?

References

Iriki et al (1996) NeuroReport v 7 p 232530
Glenberg, A. M. (1997). Behavioral and Brain Sciences, 20, 1–55.
Goodale, M. Journal of Cognitive Neuroscience 10:1, pp. 122–136
Tomasello, M. (1999) book chapter in: Ecological Approaches to Cognition:Essays in Honour of Ulrich Neisser

The Acquired Language of Thought Hypothesis: A Theory of External Symbol Grounding

Christopher Viger
Department of Philosophy
University of Western Ontario
London, Ontario
Canada N6A 3K7
cviger@uwo.ca

I develop an account of mental symbols based on David Milner's and Melvin Goodale's dual route model of vision. According to Milner and Goodale, we have two distinct pathways in the brain for processing visual information. The dorsal pathway controls real time action in the form of immediate responses to environmental stimuli, actions such as ducking or reaching. By contrast, processing in the ventral pathway results in object recognition, which makes *considered* action possible. Processing in either stream is representational, but it is the ventral stream that offers insights into the nature of symbols that figure in higherorder human cognition. Recognizing an object requires at least being able to call up relevant (to the object recognized) memories, behaviours, and inferences, suggesting a very precise interface between visual processing and processing in other subsystems of the brain. For example, recognizing something to be a dog puts us in a position to remember personal encounters with dogs, including encounters with the one recognized if we have had experience with it in the past, and disposes us to make certain inferences, such as it is a mammal, it barks, it might bite, etc. My view is that symbol learning exploits this precise interface in a way that shows how symbols are externally grounded.

I begin with a very modest notion of a symbol as something that stands in for something else. Even this simple conception faces the longstanding philosophical puzzle as to how, given materialism, anything *could* stand in for something else. How does meaning have a place in our physical world? Answering that question in my view begins with the important realization that taking mental content as somehow intrinsic and primary in order of explanation is misguided. Indeed, our starting point suggests that symbols are primarily public, conventional, and norm governed. The normativity follows from the supposition that there are correct and incorrect uses of symbols; they must be used as standins for particular kinds and any other use is in error, subject to public correction. Since anything so used is a symbol, and anything can be so used, symbols are also conventional.

The idea of symbol qua standin requires that the presence of a symbol, which can be any ordinary object, can produce behaviours and thoughts (including memories) appropriate to what the symbol stands for. Paradigmatic symbols are words of a public language, which importantly are learned; indeed, a central process in enculturation is learning the symbols, including the language, of that culture. My position, the acquired language of thought hypothesis (ALOT), is that when we learn words they are encoded in our brains at the precise interface points suggested by Milner's and Goodale's model. That is, they are encoded such that a symbol tokening puts us in a position to call up memories, behaviours, and inferences appropriate to what the symbol stands for, just as visually recognizing an object does. Mental symbols are internally encoded public symbols that stand in for something in virtue of how they are encoded, and the encoding is effected by cultural feedback.

According to ALOT, mental symbols are encoded so as to capture cultural norms of use, but also to reflect personal history through memory in virtue of which symbol tokenings are affectladen. Mental symbol tokenings are not just the processing of abstract contents; they are the activation of content that matters to us, making the contingencies of personal experience and our embeddedness in a cultural environment ineliminable aspects of our individual thought processes.

Language grounding in a complex ecological environment

Paul Vogt and Federico Divina

ILK / Language and Information Science, Tilburg University
P.O. Box 90153, 5000 LE Tilburg, The Netherlands
{*p.a.vogt,f.divina*}*@uvt.nl*

Human language is thought to have evolved from an interaction between three adaptive systems: biological evolution, individual learning and cultural evolution (Kirby & Hurford, 2002). This evolution is thought to be constrained and driven by the embodiment of humans and their situatedness in the ecology of our world. The New Ties project[1] aims at merging these aspects in a large scale simulation to evolve a cultural society of simulated agents who are situated in a complex environment (Gilbert et al., 2006). One important aspect of this simulation is to evolve language that allows the social learning of skills.

Although a lot has been achieved with computational modelling of language origins and evolution (see, e.g. Cangelosi & Parisi, 2002; Vogt, 2006, for overviews), such models necessarily have to simplify a great deal with respect to the real world, even if processed in the real world using real robots (Vogt, 2006). Of course, simplification is very useful to gain insights from simulations that only look at one particular aspect of language evolution. Such aspects vary from the evolution of sound systems (De Boer, 2001), syntax (Kirby, Smith, & Brighton, 2004), grounded lexicons (Steels, Kaplan, McIntyre, & Van Looveren, 2002; Vogt, 2002) to grounded grammars (Steels, 2005; Vogt, 2005). The problem of simplifications are that results achieved may not hold in more complex simulations. For instance, Vogt (2005) has shown that grammatical structures can emerge under completely different conditions than those reported by Kirby et al. (2004) if the meanings are perceptually grounded and acquired from scratch, and if the language is acquired using a slightly more complex learning mechanism.

The New Ties project aims to combine various aspects of language evolution models in a world that contains many agents who need to survive in a complex environment that has quite some aspects similar to our own world. Agents are to acquire behaviours that allow them to survive by combining evolutionary learning (i.e., genetic evolution), individual (reinforcement) learning and social learning. One aspect of social learning involves language learning to allow cultural evolution of language. In turn, this evolved language will be used to transfer acquired skills culturally, which thus is the second aspect of social learning involved. From the evolution of language point of view, the New Ties project will allow us to investigate many questions concerning language evolution in a realistic scenario. The sorts of questions we may ask include, for example: Under what environmental constraints will language evolve? What type of learning and interaction mechanisms are required for a language to evolve? How can the language be grounded to allow functional cooperative

[1] New Ties stands for New Emerging World models Through Individual, Evolutionary and Social learning. See http://www.new-ties.org.

(or even competitive) communication and behaviour? How can skills be grounded through social learning and language?

In this presentation we will outline this exciting project and present some details of the model with respect to language grounding and social learning. In addition, we will present some preliminary results in which we illustrate how large populations can develop shared lexicons despite many uncertainties during the interactions as to the meanings of utterances. In particular, we focus on how language can develop using a hybrid agent model involving cross-situational learning (Vogt & Smith, 2005), joint attention, feedback mechanisms and the principle of contrast (Clark, 1993).

References

Cangelosi, A., & Parisi, D. (Eds.). (2002). *Simulating the evolution of language.* London: Springer.

Clark, E. V. (1993). *The lexicon in acquisition.* Cambridge University Press.

De Boer, B. (2001). *The origins of vowel systems.* Oxford: Oxford University Press.

Gilbert, N., Besten, M. den, Bontovics, A., Craenen, B., Divina, F., Eiben, A., et al. (2006). Emerging artificial societies through learning. *Journal of Artificial Societies and Social Simulation*, *9(2).*

Kirby, S., & Hurford, J. R. (2002). The emergence of linguistic structure: An overview of the iterated learning model. In A. Cangelosi & D. Parisi (Eds.), *Simulating the evolution of language* (p. 121-148). London: Springer.

Kirby, S., Smith, K., & Brighton, H. (2004). From UG to universals: linguistic adaptation through iterated learning. *Studies in Language*, *28*(3), 587-607.

Steels, L. (2005). The emergence and evolution of linguistic structure: From lexical to grammatical communication systems. *Connection Science*, *17*(3-4), 213–230.

Steels, L., Kaplan, F., McIntyre, A., & Van Looveren, J. (2002). Crucial factors in the origins of word-meaning. In A. Wray (Ed.), *The transition to language.* Oxford, UK: Oxford University Press.

Vogt, P. (2002). The physical symbol grounding problem. *Cognitive Systems Research*, *3(3)*, 429-457.

Vogt, P. (2005). On the acquisition and evolution of compositional languages: Sparse input and the productive creativity of children. *Adaptive Behavior*, *13(4)*, 325–346.

Vogt, P. (2006). Language evolution and robotics: Issues in symbol grounding and language acquisition. In A. Loula, R. Gudwin, & J. Queiroz (Eds.), *Artificial cognition systems.* Idea Group.

Vogt, P., & Smith, A. D. M. (2005). Learning colour words is slow: a cross-situational learning account. *Behavioral and Brain Sciences*, *28*, 509–510.

Grounding Symbols in the Physics of Speech Communication

S. F. Worgan and R. I. Damper
School of Electronics and Computer Science
University of Southampton
Southampton SO17 1BJ, UK.
email {sw205r|rid}@ecs.soton.ac.uk

The symbol grounding problem (i.e., 'How can the semantic interpretation of a formal symbol system be made intrinsic to the system, rather than just parasitic on the meanings in our heads?', Harnad 1990) is crucial to cognition. Thus, it has been argued that grounding poses a challenge that cannot be neglected (Cangelosi, Greco, and Harnad 2001). We believe human communication to be the clearest, certainly best developed, example of externally grounded cognition. Despite the advantages inherent in considering speech as a grounded system, there is a danger—through simulating at too high a level of abstraction—of effectively ignoring this crucial aspect (e.g., de Boer 2000; Oudeyer 2005). But how are we to define grounding at the 'phonetic' level of speech sounds? In this paper, we argue that the emergence of speech can and should be grounded in the physics of speech communication between agents, recognising that the human's contact with the external world of sound is via their articulatory and auditory systems.

We proceed by adopting the view of speech communication offered by Lindblom and Studdert-Kennedy (1984). Specifically, we are seeking to minimise the articulatory effort of an utterance, at the same time maximising its perceptual distinctiveness to other agents. In grounding terms, the drive for perceptual distinctiveness is important in shaping the coupled production-perceptual system. The higher the perceptual distinctiveness, the clearer the meaning of the utterance. This kind of interaction has already been investigated by Kirby (2001) at the syntactic level (and so tacitly assumes the emergence of phonetic distinctiveness). Having defined the nature of phonetic grounding, we are currently implementing a system that introduces this grounding into Oudeyer's (2005) previously ungrounded investigations, Figure 1. Following Guenter and Gjaja (1996), Oudeyer's work has shown how two self-organising maps (SOMs, see Kohonen 1990)—one representing the auditory system and the other the articulatory system—can converge from producing a series of random utterances to producing a shared set of discrete speech sounds. This process is considered analogous to the emergence of early hominid speech. However, without any definition of articulatory effort or perceptual salience, this convergence process often terminates in one central point (as found by Oudeyer and confirmed by us). We propose to overcome this problem, and hopefully produce more realistic utterances, by defining a *contour space* within each SOM, i.e., an objective function which embodies measures of both effort and distinctiveness. Therefore, as well as converging to a shared language (shared between agents, that is), each SOM will attempt to optimise itself within its contour space.

This definition of contour spaces—as embodying the effort of the utterance within the articulatory system and the perceptual distinctiveness within the auditory system—provides a direct grounding to the sensory-motor process of each individual agent. The articulatory effort is measured by the muscle energy expenditure (Umberger, Karin, and Philip 2003) of an artificial vocal tract (Maeda 1982), which forms the means whereby the agent acts upon its environment, i.e., its motor process. The perceptual contour space is dictated by the human peripheral auditory system, modelled on the work of Pont and Damper (1991)—the sensors of the agent. Although this system is grounded within its environment, it does not yet form (or manipulate) any explicit symbols. However, distinct and grounded attractors do emerge during the lifetime of the agent(s), and these we count as 'symbols'.

We are still grounding the external world via these attractors, but rather than connecting an imperfect, arbitrary abstraction (as when a cat in the environment is miraculously labelled CAT in one bound), we are connecting a more complete representation of the distal object, built on the physics

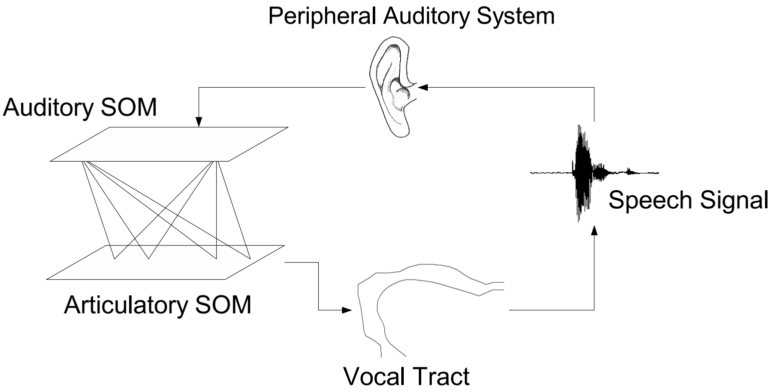


Figure 1: An agent producing and listening to its own utterances.

of the situation. Through the definition of attractors we have both a clear shared abstraction, its centre point, and a basin of attraction capturing the ambiguity and differences present in the real world. We feel that this view, based on emergence of attractors in articulatory-auditory spaces, can answer some of the current criticisms of the symbol grounding paradigm (Lakoff 1993), just because the attractors capture the ambiguities and 'shades of grey' that challenge more traditional grounded implementations (Davidsson 1993). This has precedence in other grounded implementations (e.g., Harnad 1993; Damper and Harnad 2000) that take the form of grounded, connectionist (neural network) models. These have been successful in displaying various aspects of human cognition. But, by considering grounding at the phonetic level, we have developed a new framework in which this interplay between symbolic grounding and connectionist systems can be further explored.

References

Cangelosi, A., A. Greco, and S. Harnad (2001). Symbol grounding and the symbolic theft hypothesis. In A. Cangelosi and D. Parisi (Eds.), *Simulating the Evolution of Language*, pp. 191–210. London: Springer-Verlag.

Damper, R. I. and S. R. Harnad (2000). Neural network models of categorical perception. *Perception and Psychophysics 62*(4), 843–867.

Davidsson, P. (1993). Toward a general solution to the symbol grounding problem: combining machine learning and computer vision. In *Fall Symposium Series, Machine Learning in Computer Vision: What, Why and How?*, Raleigh, NC, pp. 157–161.

de Boer, B. (2000). Self-organization in vowel systems. *Journal of Phonetics 28*(4), 441–465.

Guenter, F. H. and M. N. Gjaja (1996). The perceptual magnet effect as an emergent property of neural map formation. *Journal of the Acoustical Society of America 100*(2), 1111–1121.

Harnad, S. (1990). The symbol grounding problem. *Physica D 42*, 335–346.

Harnad, S. (1993). Grounding symbols in the analog world with neural nets. *Think 2*(1), 12–78.

Kirby, S. (2001). Spontaneous evolution of linguistic structure – an iterated learning model of the emergence of regularity and irregularity. *IEEE Transactions on Evolutionary Computation 5*(2), 102–110.

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE 78*(9), 1464–1480.

Lakoff, G. (1993). Grounded concepts without symbols. In *Proceedings of the Fifteenth Annual Meeting of the Cognitive Society*, Boulder, CO, pp. 161–164.

Lindblom, B.; MacNeilage, P. and M. Studdert-Kennedy (1984). Self-organizing processes and the explanation of phonological universals. In B. Butterworth, B.; Comrie and O. Dahl (Eds.), *Explanations for Language Universals*, pp. 181–203. Berlin: Mouton.

Maeda, S. (1982). A digital simulation method of the vocal-tract system. *Speech Communication 1*(3 – 4), 199–229.

Oudeyer, P.-Y. (2005). The self-organization of speech sounds. *Journal of Theoretical Biology 233*(3), 435–449.

Pont, M. J. and R. I. Damper (1991). A computational model of afferent neural activity from the cochlea to the dorsal acoustic stria. *Journal of the Acoustical Society of America 89*(3), 1213–1228.

Umberger, B. R., G. M. G. Karin, and E. M. Philip (2003). A model of human muscle energy expenditure. *Computational Methods of Biomechanical and Biomedical Engineering 6*(2), 99 – 111.

Author index