Honkela, T. and Winter, J. (2003). Simulating Language Learning in Community of Agents Using Self-Organizing Maps. Helsinki University of Technology, Publications in Computer and Information Science, Report A71. (ISBN 951-22-6881-7)

Simulating Language Learning in Community of Agents Using Self-Organizing Maps

Timo Honkela and Juha Winter

Helsinki University of Technology
Neural Networks Research Centre
timo.honkela@hut.fi

December 15, 2003

Abstract

In this article, we present a model of a cognitive system, or an agent, with the following properties: it can perceive its environment, it can move in its environment, it can perform some simple actions, and it can send and receive messages. The main components of its internal structure include a working memory, a semantic memory, and a decision making mechanism. In our implemented simulation model, the agent associates linguistic expressions and visual perceptions. The main motivation for communication is to exchange information. The linguistic expressions are not symbolic but pattern-like. With the current framework and simulation tool, we wish to provide a useful model for language emergence based on the unsupervised learning paradigm among a community of communicating autonomous agents. In the future, we plan to include other aspects of cognitive modeling including more realistic multimodal information processing, anticipatory decision making, language evolution, and emotional modeling.

1 Introduction

Distributed artificial intelligence is a branch of artificial intelligence that is concerned with cooperative problem solving by a decentralized group of collaborating agents. The basic idea is to obtain emergent behaviors of agents in the system based on local decision-making of each agent. The effectiveness of the system depends on the form of interactions between agents and on the capabilities and functioning of each one. One objective of the research may be to develop effective systems for application areas such as electronic commerce, and automated negotiation and contracting (see, e.g., Sandholm 1999). Another objective is to model social and cognitive systems in order to gain insight on the nature of such distributed systems. The latter is our main point of view.

A predominant approach among computational models of language has been based on the idea that the linguistic categories and rules are predetermined and coded manually in the systems. In the traditional artificial intelligence based on symbol manipulation the conceptual framework of the agents is usually given: all agents share the elementary concepts. The emerging field of connectionist natural language processing, based on the use of artificial neural networks, may be characterized to take an opposite stand. Some basic ideas of radically connectionist approach are outlined in the following.

The fact that the expressions in written natural language appear to be inherently symbolic and discrete does not imply that symbolic descriptions of linguistic phenomena are sufficient as such especially when semantic and pragmatic issues are considered.

To be able to model the relation between continuous phenomena and discrete symbols, the building blocks of the theory must contain sufficient tools for that. It is rather obvious that reference from a symbol to the reality is often fuzzy but also linguistic categories may be fuzzy (Zadeh 1965).

The ability to understand natural language utterances may mainly be learned via examples. The categories necessary in the interpretation emerge in the self-organizing learning processes and they may basically be implicit rather than explicit. An implicit category can be used during interpretation even if it is not named.

In this article, the main focus is in modeling communities of conceptually autonomous agents following the basic principles outlined above. An agent is conceptually autonomous if it learns its representation of the environment by itself, where a concept is taken to be simply a means of specifying a relationship between language and world.

Partial autonomy means a setting in which the learning process of an agent is influenced in some way by other agents (Honkela et al. 2003). This influence can then serve as a basis for communication between agents. Thus, although each agent has an individual representation of the environment, the representations are related through the coordinating effect of communication between agents in situations where all agents have access to similar perceptions of the environment. A counter example to a conceptually autonomous agent is a traditional expert system which has been pre-programmed and therefore is given a priori its representation of the environment.

In this article, we consider both the individual and community point of view. For instance, we are interested in how we could model the vocabulary learning of an individual agent. For that purpose, we make a distinction between working memory and semantic memory. We are also interested in the society of agents: how language emerges in the community, and how intersubjective agreement on the use of words is reached.

There has recently been considerable interest on the models of language evolution (see, e.g., Christiansen and Kirby 2003). In this article, we only consider one single population of agents, not any evolution through subsequent populations. However, we share one idea that is common with some evolutionary models, namely that the agents in the population "live" in a grid world (see, e.g., Grim et al. 1999 and 2001).

The focus in models of language learning is often in learning morphology or syntax. This seems to be true both in the case of modeling individual learning (see, e.g., Rumelhart and McClelland 1986, Chen and Honavar 1999) or learning through evolution (see, e.g., Nowak et al. 2001). In comparison with these examples, the main contrasting elements in our research has been the focus on semantics (see also, e.g., Regier 1992, or Bailey 1997). Moreover, we apply unsupervised learning in general (see, e.g., Oja 2002; on unsupervised language learning see, e.g., Powers 1997) and specifically the use of the self-organizing map (see, e.g., Ritter and Kohonen 1989, Honkela and Vepsäläinen 1991, Scholtes 1991, Miikkulainen 1993, and Honkela et al. 1995). As a model of community of communicating adaptive agents, a closely related approach to ours is presented by Grim et al. (1999, 2001).

As a unique feature in our research, considering models with similar overall aims and framework, is the idea that the vocabulary does not need to be represented as discrete symbols. We represent the linguistic expressions as pattern vectors. This approach is compatible with the idea of processing and generating spoken rather than written language.

In general, our research is strongly inspired by the cognitive science research and philosophical argumentation that emphasizes the need to take embodiment and contextuality into account within cognitive modeling. Among the large number of relevant researchers one can mention, for instance, Clark (1997), Cole and Engeström (1991), Hendriks-Jansen (1996), Hörmann (1986), Hutchins (1995), Lakoff and Johnson (1999), Maturana and Varela (1980), Pfeifer and Scheier (1999), Rorty (1979), Sharkey and Ziemke (2001), Steels (1996), Steels and Kaplan (2001), Tomasello (1999), Varela et al. (1991), Von Foerster (1981), and Ziemke (2003).

2 Neural Model of Situated Cognition and Conceptual Autonomy

We will consider a model of a cognitive system based on an artificial neural network. The system, an agent, has the following properties: it can perceive its environment, it can move in its environment, it can perform some simple actions that will be discussed later, and it can send and receive messages. The main components of its internal structure include a working memory, a semantic memory, and a decision making mechanism. In the current version of our model, we focus on the semantic memory and therefore many known details of human memory system are not taken into account, e.g., the episodic memory (for more details, see, e.g., Cowan 1998, or Miyake and Shah 1999). Moreover, in our current model, we do not take explicitly into account the phenomenon of categorical perception and the feedback from semantic memory to the perceptual processes.

In our model, the main emphasis is to model the conceptual autonomy of an agent, i.e., it learns the representation of the environment by itself. In our model, the environment consists of a grid of places each of which is empty or contains an agent, a potentially edible object, or an obstacle. The agents can perceive other subjects and objects in its environment. The perception is a high-dimensional vector that indicates only indirectly the category into which each of them belongs. These vectorial patterns are discussed in more detail later in this article. The overall architecture of an agent is presented in Fig. 1.

2.1 Self-Organizing Map as an adaptive semantic memory

The self-organizing map (SOM) (Kohonen 1982, 2001) is a widely used artificial neural network model. In the SOM, the learning process is unsupervised: no a priori classifications for the input examples are needed. The learning process is based on similarity comparisons in a continuous space. The result is a system that associates similar inputs close to each other in the two-dimensional grid called the map.

During the past two decades, a constructive approach to learning and knowledge has become dominant in educational psychology. Learning is viewed as an active, constructive process rather than a passive, reproductive process (Glaser and Bassok 1989).

The theories of knowledge have traditionally been based on predicate logic and related methodologies and frameworks. The basic ontological assumption is that the world consists of objects, events and relationships. The language and the conceptual structures are then supposed to reflect rather straightforwardly this structure. Learning has been seen as a means to memorize the mapping from the epistemological domain (to put it simply: words) into the ontological domain (objects, events and relationships). This view has been dominant at least partly because of the consistent formalization of the theory through the use of symbolic logic. Moreover, the use of the von Neumann computer as the model or metaphor of human learning and memory has had similar effects and has strengthened the idea of the memory as a storage of separate compartments which are accessed and processed separately and which are used in storing and retrieving information more or less as such. (Honkela et al., 2000)

Next we will consider some details of the self-organizing map algorithm that are necessary for considering the details of the semantic memory of the SOMAgent model. Let's assume that some sample data sets would be mapped onto an array that will called the map. The set of input samples is described by a real vector

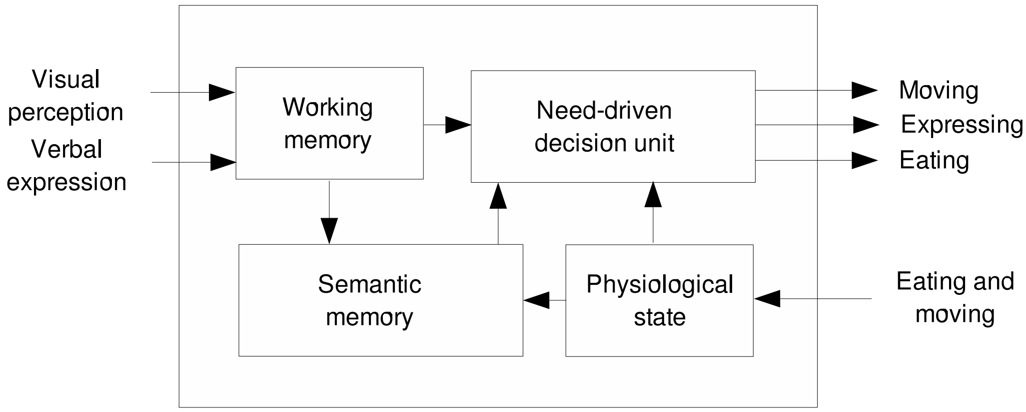

Figure 1: The basic SOMAgent architecture. The agent receives two kinds of perceptual inputs: visual images and linguistic expressions. There are three kinds of potential actions: the agent can either eat, move or utter an expression. The perceptions are primarily stored in the working memory. The semantic memory associates perceptual information and information considering its physiological state. Sudden changes in the physiological state are related to eating actions and the quality of the eaten object determines the direction of the change. The physiological state is also influenced by moving: gradually the agent loses energy. The physiological state serves as the basic motivational factor for the agent. If the energy level is low the agent prefers eating and high energy level makes the exploration of the environment to become more probable action. Communication between the agents is motivated by the exchange of information related to the edibility of the food items. It is assumed that the visual perceptual characteristics of the food items correlate strongly with their level of edibility. The agents do not have explicit information on their position in the environment and therefore their navigation is based on the perceived landmarks in the environment. The environment of an agent consists of other agents, a number of obstacles and food items. Each obstacle has unique visual characteristics which helps the agents in their navigation.

$\mathbf{x}(t) \in R^n$ where t is the discrete-time coordinate. Each unit in the map contains a model vector $\mathbf{m}_i(t) \in R^n$, which has the same number of elements as the input vector $\mathbf{x}(t)$. The self-organizing map algorithm is a regression process. Thereby, the initial values of the components of the model vector, $\mathbf{m}_i(t)$, may even be selected at random.

Any input item is thought to be mapped into the location, the $\mathbf{m}_i(t)$ of which matches best with $\mathbf{x}(t)$ in some metric. The self-organizing algorithm creates the ordered mapping as a repetition of the following basic tasks:

1. An input vector $\mathbf{x}(t)$ is compared with all the model vectors $\mathbf{m}_i(t)$. The best-matching unit on the map, i.e., the unit where the model vector is most similar to the input vector in some metric (e.g. Euclidean) is identified.
2. The model vectors of the winner and a number of its neighboring units in the array are changed towards the input vector according to the learning principle specified below.

The basic idea in the SOM is that, for each sample input vector $\mathbf{x}(t)$, the winner and the units in its neighborhood are changed closer to $\mathbf{x}(t)$ in the input data space. During the learning process, individual changes may be contradictory, but the net outcome in the process is that ordered values for the $\mathbf{m}_i(t)$ emerge over the array. If the number of available input samples is restricted, the samples must be presented reiteratively to the SOM algorithm[1].

Adaptation of the model vectors in the learning process may take place according to the following equations:

$$\begin{aligned} &\mathbf{m}_i(t+1) = \mathbf{m}_i(t) + \alpha(t)[\mathbf{x}(t) - \mathbf{m}_i(t)] \qquad \text{for each } i \in N_c(t), \\ &\mathbf{m}_i(t+1) = \mathbf{m}_i(t) \qquad \text{otherwise,} \end{aligned} \tag{1}$$

where t is the discrete-time index of the variables, the factor $\alpha(t) \in [0,1]$ is a scalar that defines the relative size of the learning step, and $N_c(t)$ specifies the neighborhood around the winner in the map array.

At the beginning of the learning process the radius of the neighborhood is fairly large, but it is made to shrink during learning. This ensures that the global order

[1] In the agent model, this can be achieved, if necessary, using the working memory as an intermediate memory

is obtained already at the beginning, whereas towards the end, as the radius gets smaller, the local corrections of the model vectors in the map will be more specific. The factor $\alpha(t)$ also decreases during learning[2].

There is neurophysiological evidence to support the idea that the self-organizing map captures some of the fundamental processing principles of the brain, especially of the experimentally found ordered maps in the cortex (consider, e.g., Caramazza et al. 1994, Kohonen 1993). Some earlier artificial-neural-network models of self-organization have been presented, e.g., by Amari (1967), von der Malsburg (1973), and Carpenter and Grossberg (1987). The self-organizing map appears to model the formation of brain maps most efficiently. Based on information theoretic considerations, Bishop et al. (1996) have presented a variant of the SOM, called Generative Topographic Mapping, GTM.

The self-organizing map can be considered as an adaptive semantic memory model. It is dynamic, associative and consists of elements that can be called adaptive prototypes. For instance, MacWhinney (1998) and Gärdenfors (2000) have presented how the self-organizing map can be used in modeling language learning and formation of conceptual spaces.

In the self-organizing map, inputs are not stored as such but comparison is made between the input and the collection of prototypes. The closest prototype of the input is adapted towards the input, or to say with other words, the prototype is made more similar to the input to some degree. The same operation is also conducted for the neighboring prototypes, which gives rise to the topographical order on the map. Thus, the adaptation process in the self-organizing map algorithm is based on the principle that what already exists in the system also influences the learning result.

Honkela and Vepsäläinen (1991) used self-organizing map in modeling impreciseness of language and how the imprecise mapping between language and world can be learned. One basic idea was to consider the meaning of a word as a distribution, potentially in a multidimensional space, not as a point or node in a network of symbols (consider, e.g., Ritter and Kohonen 1989, and Churchland 1989).

The self-organizing map is a good example of a “non-trivial machine” (Von Foester

[2]In data analysis and visualization applications, the neighborhood radius and the learning factor can approach zero. In the agent model, the plasticity of the semantic memory needs to be kept in such a level that the agent can learn in novel situations.

1984). It is not behavioristic model, but on the contrary, the internal state of the system influences the behavior and it is changing during the “life” of the map. One of the reasons why the self-organizing map and other artificial neural network models may appear behavioristic is that their internal state is difficult to grasp and analyze. For instance, multilayer perceptrons typically are used as black-box systems: only the input-output behavior is considered in applications. However, as cognitive models these systems should be considered only in the continuous learning mode: the relationship between the input and the output is constantly subject to change and, thus, the system is non-trivial. As a system based on unsupervised learning paradigm, the self-organizing map is even less trivial than multilayer perceptrons and other networks based on supervised learning while there no classification framework predetermined by the designer of the system.

Some early ideas that have been implemented and taken much further in this article were presented in (Honkela 1993). Closely related early research on natural language processing based on the self-organizing map includes (Ritter and Kohonen 1989, Scholtes 1991, Finch and Chater 1992, Miikkulainen 1993, Honkela et al., 1995, and Miikkulainen 1997).

2.2 Communication and association of language and perceptions

Gärdenfors (2000) distinguishes three cognitive levels of representation: the most abstract level is the symbolic level, on which the information is represented in terms of symbols that can be manipulated without taking into account their meaning. The least abstract level is the subconceptual representation. In the mediating level of the conceptual representation the concepts are explicitly modeled. For instance, de Sa (1994) has studied how the self-organizing map can be used in creating cross-modal associations.

In our model, the agent associates linguistic expressions and visual perceptions. The main motivation for the agents to communicate is to exchange information on the edibility of the food items. In our experiment it is assumed that the level of edibility correlates strongly with the visual characteristics of the food item. Thus, the agents can collect useful indirect experience. The original situation is such that the expressions that the agents use are random: each agent has, in principle, a language of its own. However, originally the information exchange is contextual, i.e., two communicating agents can both perceive the same item. This is the basis for symbol formation, symbol grounding and transformation from subjective individual languages into intersubjective language shared by the community (consider also Steels 1996, Steels and Kaplan 2001). However, even after convergence into common language, the agents have a certain level of subjectivity, i.e., the reference relation between language and world is not identical between any two agents but is generally close enough in order useful communication to take place. This individuality of interpretation is a natural phenomenon when the multidimensional and continuous nature of both linguistic expressions and perceptions is taken into account. The individuality of interpretation is considered a problem when meaning and understanding are studied in the framework of symbolic representations and within model theoretical approach. By emphasizing the pattern nature of language and world, we avoid the idea that some relativism would be a problem as it is in the traditional epistemological research in which language is considered as a collection of propositions and the world consists of a collection of distinct objects and their relationships. The basic idea of imitation may be considered to be central for language learning in social contexts (see, e.g., Boer 2001, Breidegard and Balkenius 2003, and Meltzoff and Decety 2003). Through imitation a gradual process of convergence between individuals is achieved.

2.3 Continuous learning

Batch learning and on-line or continuous learning are two training paradigms for artificial neural networks. In batch learning, the training is carried out with an entire data set simultaneously. In on-line learning the network is updated after the presentation of each training sample. On-line learning is a natural approach in non-stationary environments and tasks.

One central difficulty in on-line training is the sensitivity of most training methods to the choice of training parameters. Wrong choice of parameters may lead to the slowdown of training. It may also influence the system's ability to converge successfully. The Bayesian approach has been applied quite successfully within the framework of batch learning. Extensions to the on-line case, where explicit information on past examples is not stored, have been more limited (Saad 1999).

In the current version of our simulation system we apply a rather straightforward approach for using the self-organizing map for continuous learning. More grounded model have been presented, for instance, by Hung and Wermter (2003). However, our approach suffices rather well for the purposes of the overall scheme.

3 Experiments

The experimental setting is as follows: there is a group of autonomous agents, for instance five, living in an environment. The environment is a two-dimensional grid, typically of size 12 by 12 squares (see Fig. 2). In addition to agents, the environment contains objects, namely food and obstacles. Each agent and object has a unique appearance which is determined by its pattern vector. There are roughly two types of food: non-poisonous and poisonous. The amount of nutrition gained or lost through eating a specific food object varies greatly based on the object's internal characteristics. Basically, no two objects are alike.

3.1 Behavioral modes

There are two basic behavioral modes or strategies that the agents follow. The first one is survival which simply means that the agents try to ensure their survival by finding and eating food. The second mode is exploration which emphasizes scouting unfamiliar areas of the environment. Each agent chooses individually between these modes based on its current status attributes which will explained later in more detail. The basic actions are moving, eating and messaging.

Agents have two basic attributes: condition and motivation. They are both numeric quantities. Condition declines due to moving and improves through eating non-poisonous food. Motivation determines the willingness to take risks, e.g., eating unfamiliar food. Initially, each agent has a high motivation but eating poisonous food causes the motivation to decline. Decision making is a process based on the environment and each agent's individual experiences and attributes.

3.2 Memory

Each agent has an individual memory consisting of two parts: working memory and semantic memory. Although these two memory types are closely interconnected, they have different implementations. Working memory is dynamic and accurate in nature, whereas semantic memory is adaptive and approximative. All perceptions and experiences (e.g. food related) are stored in memory.

The working memory holds only the most recent experiences, typically 7. It is based on a simple FIFO scheme, i.e. new experiences force older ones out in a pipeline-like fashion. The contents are primarily used in messaging but also in the adaptation process of the semantic memory.

The semantic memory, in contrast, is based on the self-organizing map. The typical size of the map is 40×40 in our experiments The map continuously adapts to new experiences.

Each memory element consists of a pattern vector, an experience vector, a word vector and a message indicator. The pattern vector describes the looks and physical characteristics of an object. The experience vector contains experience information related to the object, e.g. whether the food was poisonous or not. The word vector is a language "word" used in messaging to denote the object. It is not a symbol but an array of floating-point numbers like any other vector in the simulation. Finally, the message indicator simply tells whether or not the experience was received from another agent in a message.

3.3 Communication

Agents share their most recent experiences (i.e. those in working memory) by sending messages to other nearby agents. Collective learning and collaboration is thus possible, and redoing the same mistakes can be avoided. Communication happens throughout the simulation and it affects both exploration and food seeking. When exploring the environment, agents tend to avoid scouting areas that have already been scouted by other agents. This is possible because obstacle perceptions are stored in memory and shared just like food experiences. The more familiar the obstacle, the less likely it is that an agent approaches it. Information on the edibility of food is also gathered and shared collectively.

At the beginning of the simulation, each agent has a random "vocabulary" i.e. set of word vectors that it uses to denote different kinds of objects. This vocabulary has been integrated as a part of the agent's semantic memory. As the simulation proceeds and messages are exchanged between agents, these vocabularies slowly converge to form a common "language". In the later phases, only a word vector is sufficient to denote a specific object, so no pattern vectors are needed in messaging.

4 Conclusions and discussion

In this article, the main focus has been in modeling communities of conceptually autonomous agents. We have developed a simulation environment in which we model agents that learn their representation of the environment, and the relationship between language and their environment.

Future steps include near-term tasks such as studying the convergence properties when different settings are varied as well as wider developments. Some potential long-term directions are outlined in the following.

- Crossmodal and multimodal information processing can be studied much more in detail. The vision component is, for the moment, very simplistic and can be taken further, e.g., by following the approach outlined by Kopp (2003). In visual perception, detecting invariances is an important task. An extension of the self-organizing map, the ASSOM model (Kohonen 1997) appears to be one well-suited method for this purpose.

  The aspects of spoken language are important from the basic approach and there is a clear motivation for studying the use of speech in communication and in language learning (Linell 1982). Interesting recent results include (Boer 2001, and Breidegard and Balkenius 2003). It is also to be remembered that the self-organizing map was originally developed in the context of speech recognition research (Kohonen 1988).

- The current agent model does not include any notion of anticipation. Kaipainen (1994) has used SOMs in modeling anticipation. In addition to the basic SOM, Kaipainen uses a a list of lateral connections that record the transition probabilities from one map node to another. The time dynamics of a process are characterized by the trajectory path on the map. This aspect has been important already in the first application area of the SOMs, namely speech recognition, and more recently in process monitoring. In Kaipainen's experiments the most natural model was open to to the input

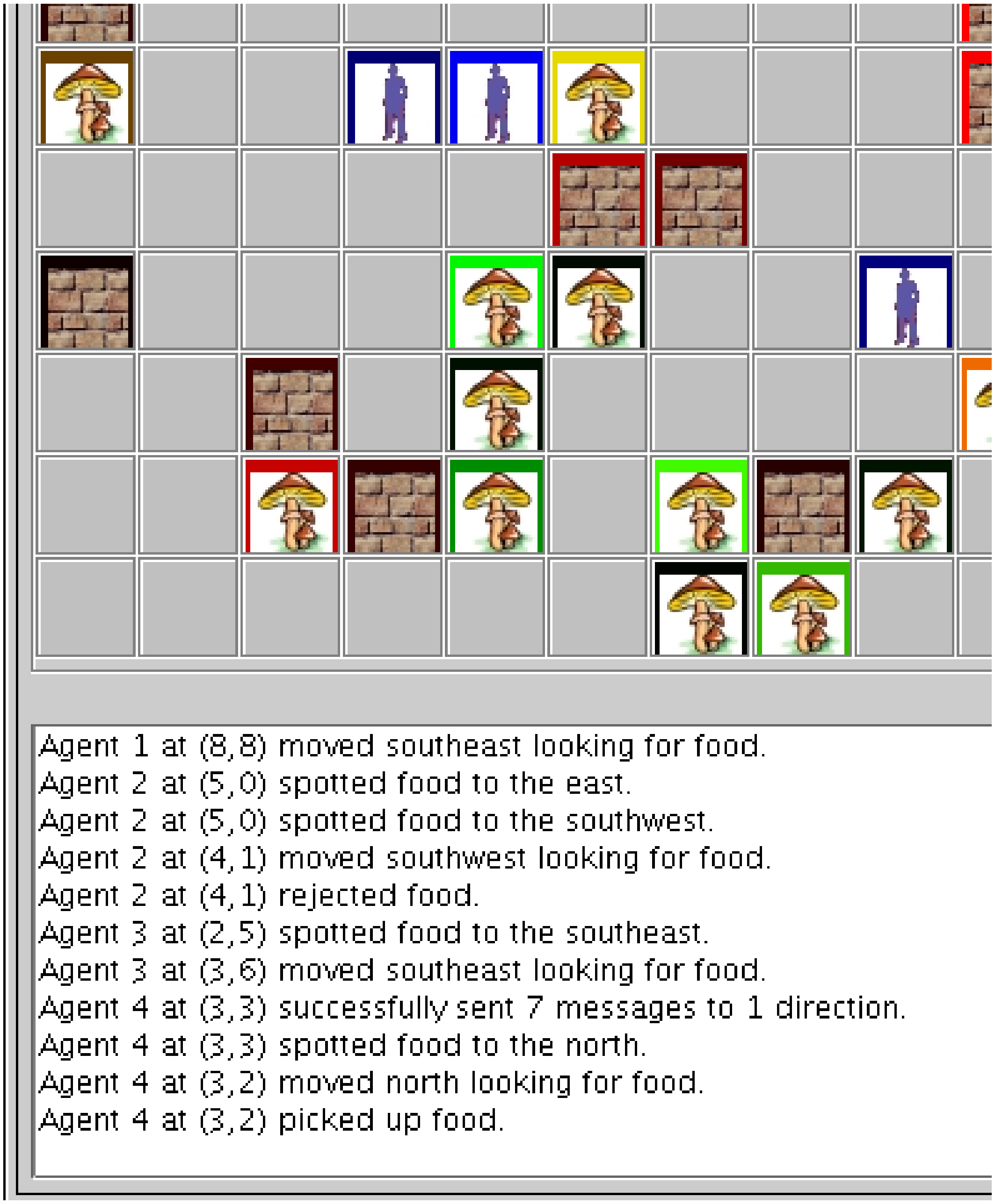

Figure 2: The SOMAgent simulation interface consists of a 2-dimensional grid. Each location is either empty or contains an agent, an obstacle or a food instance. The visual appearance is meant to give an overall view, not to reflect accurately the nature of the objects.

having at the same time an internal schematic drive, anticipation, which intentionally actualized situations rather than just recognizing them as given. Also Honkela (1997) considers models for anticipation based on the self-organizing map.

- In this article, the evolutionary aspects, for instance, development of communication based on simulating subsequent generations in a population was not considered. This is an active area of research (see, e.g., Hurford 2002 or Dowman 2003). It is quite obvious that our model could be extended to include evolutionary modeling.

- The current system can rather straightforwardly be extended to include models of emotions and feelings (cf., e.g., Sloman and Croucher 1981, Picard 1997, Hyvärinen and Honkela 1999, Lisetti and Gmytrasiewicz 2002, Damasio 2003). It has become obvious that emotions have a central role in, e.g., decision making.

- In addition to the constructive knowledge building that is modeled in our current system using the self-organizing map, an agent needs to keep track of invariances in the processes in the environment. Here it is also relevant to keep in mind that, as Von Foerster (1981) among others points out, the reality exists but objects and events as discrete entities are constructions of our minds. The conditional probabilities are related to processes in the environments. The number of data needed for accurate estimation of large number of combinations of variables and contextual features is so large that human beings tend to use simplified heuristics. These heuristics have been researched in detail by Gigerenzer et al. (2000).

- Realistic simulations of the social and cultural level are seemingly difficult to build due to the complexity of the overall system. The richness of human culture makes it difficult as a phenomenon to model Moreover, already the world knowledge of a single human being is so vast that it is difficult to approach it successfully. However, useful development may be possible by taking into account the aspects presented, e.g., by Vygotsky (1978, 1986), Cole and Engeström (1991), and Lindblom and Ziemke (2003).

- In general, we wish to be able to develop a framework that would be useful for multidisciplinary research. This framework could be used in linking together aspects of, for instance, cognitive science, general linguistics, cognitive linguistics, psycholinguistics, sociolinguistics, theoretical philosophy, and artificial intelligence in relation to research on language learning, communication and collaboration.

Some of the aspects presented above are most naturally studied within robotics research (consider, e.g., Billard and Dautenhahn 1998, Brooks 1991, Brooks et al. 1998, Pfeifer and Scheier 1999, Ziemke 2003). However, we believe that also computer simulations can still be useful in the future in studying some specific aspects of cognitive processes. Moreover, many aspects of our research can be linked with practical application areas in computer science including development of large software systems. For instance, in the future the systems need to be able to cope (semi)autonomously within the contexts they are used without having been explicitly designed and programmed for their specific tasks. Moreover, with increasing complexity of software, it will be useful if the modules of a large system are able to communicate with each other based on emerging compatibility, based on automatic meaning negotiations, rather than being designed and programmed to be compatible. This feature alone would be of high practical value for software and telecommunications industry.

Intentionality or goal-directedness is one important aspect that has only briefly touched upon in this article. However, it has been pointed out that intentionality is one central notion to be taken into account when concept of meaning is discussed (Zlatev 2002).

In summary, we have developed a simulation environment for studying the emergence of language among autonomous agents. We foresee that these kinds of experiments will be useful in increasing understanding on phenomena that have been for long studied in philosophy, general linguistics and cognitive science.

Acknowledgements

The first author is grateful for a number of discussions related to the theme of the paper during the past years. It would be impossible to thank all the relevant persons, nor to detail their influence. However, being aware of the incompleteness of the list, also in order to appreciate science as a social activity, warm regards for their intellectual contribution are directed, in alphabetical order, to Christian Balkenius, Petra Björne, Catherine Bounsaythip, Melissa

Bowerman, Angelo Cangelosi, Lauri Carlson, Mathias Creutz, Peter Gärdenfors, Ulrike Hahn, James Hampton, Jana Holsánová, James Hurford, Kevin I. Hynnä, Aapo Hyvärinen, Heikki Hyötyniemi, Mauri Kaipainen, Jonna Kangasoja, Fred Karlsson, Samuel Kaski, Jim Kirby, Tarja Knuuttila, Teuvo Kohonen, Lars Kopp, Kimmo Koskenniemi, Mikko Kurimo, Krista Lagus, Christine Lisetti, Kirsti Lonka, Sini Maury, Risto Miikkulainen, Erkki Oja, Ville Pulkki, Friedeman Pulvermüller, Antti Raike, Jarno Raukko, Tuomas Sandholm, Linda B. Smith, Jaakko Särelä, Naftali Tishby, Ville Tuulos, Harri Valpola, Ricardo Vigário, Sakari Virkki, Stella Vosniadou, Lotfi Zadeh, and Jordan Zlatev. This report would not have emerged without the influence of all of these people. However, the report's contents are the responsibility of the authors.

References

Amari, S.-I. (1967). A theory of adaptive pattern classifiers. *IEEC*, vol. 16, pp. 299-307.

Bailey, D. (1997). *When push comes to shove: A computational model of the role of motor control in the acquisition of action verbs.* Ph.D. dissertation, Computer science division, EECS Department, University of California at Berkeley.

Billard, A. and Dautenhahn, K. (1998). Grounding communication in autonomous robots: An experimental study. *Robotics and Autonomous Systems*, vol. 24, pp. 71-79.

Bishop, C.M., Svensén, M. and Williams, C.K.I. (1996). *GTM: A Principled Alternative to the Self-Organizing Map*. M.C. Mozer, M.I. Jordan and T. Petche (eds.), *Advances in Neural Information Processing Systems*, pp. 354-360.

Boer, B. de (2001). *The Origins of Vowel Systems*. Oxford University Press.

Brazdil, P. and Muggleton, S. (1991). Learning to Relate Terms in a Multiple Agent Environment. *Proceedings of Machine Learning - EWSL-91*, Springer, Berlin, pp. 424-439.

Breidegard, B., and Balkenius, C. (2003). Speech development by imitation. C. G. Prince, L. Berthouze, H. Kozima, D. Bullock, G. Stojanov, and C. Balkenius (eds.), *Proceedings of the Second International Workshop on Epigenetic Robotics:*

Modeling Cognitive Development in Robotic Systems, Lund University Cognitive Studies, pp. 57-64.

Brighton, H., Kirby, S., and Smith, K. (2003). Situated cognition and the role of multi-agent models in explaining language structure. D. Kudenko, E. Alonso, and D. Kazakov (eds.), *Adaptive Agents and Multi-Agent Systems: Adaptation and Multi-Agent Learning*, Springer, pp. 88-109.

Brooks, R. A. (1991). Intelligence Without Representation. *Artificial Intelligence Journal*, vol. 47, pp. 139-159.

Brooks, R., Breazeal, C., Marjanovic, M., Scassellati, B. and Williamson, M. (1998). The Cog project: Building a humanoid robot. C. Nehaniv (ed.), *Computation for metaphors, analogy, and agents*. Springer, New York, pp. 52-87.

Caramazza, A., Hillis, A., Leek, E. C., and Miozzo, M. (1994). The organization of lexical knowledge in the brain: Evidence from category- and modality-specific deficits. *Mapping the mind: Domain specificity in cognition and culture*, Cambridge University Press, Cambridge, pp. 68-84.

Carpenter, G. and Grossberg, S. (1987). A massively parallel architecture for a self-organizing neural pattern recognition machine. *Computer Vision, Graphics, and Image Processing*, vol. 37, pp. 54-115.

Chen, C-H., and Honavar, V. (1999). A Neural Network Architecture for Syntax Analysis. *IEEE Transactions on Neural Networks*, vol. 10, pp. 91-114.

Christiansen, M. and Kirby, S. (eds.) (2003). *Language Evolution*. Oxford University Press.

Churchland, P.M. (1989). *A Neurocomputational Perspective: The Nature of Mind and the Structure of Science*. MIT Press, Cambridge, Massachusetts.

Clark, A. (1997). *Being there - putting brain, body and world together again*. MIT Press, Cambridge, MA.

Cole, M. and Engeström, Y. (1991). A Cultural-Historical Approach to Distributed Cognition. G. Salomon, (ed.), *Distributed Cognition*. Cambrigde University Press, pp. 1-47.

Cowan, N. (1998). *Attention and Memory: An Integrated Framework*. Oxford University Press.

Damasio, A. (2003). *Looking for Spinoza: Joy, Sorrow, and the Feeling Brain*. Harcourt.

de Sa, V.R. (1994). *Unsupervised Classification Learning from Cross-Modal Environmental Structure*. PhD thesis, University of Rochester, Department of Computer Science, Rochester, New York.

Dowman, M. (2003). Explaining Color Term Typology as the Product of Cultural Evolution using a Bayesian Multi-agent Model. *Proceedings of Cognitive Science Conference,* (CD-ROM), Boston, MA.

Finch, S. and Chater, N. (1992). Unsupervised methods for finding linguistic categories. Aleksander, I. and Taylor, J., (eds.), *Artificial Neural Networks*, vol. 2, North-Holland, pp. 1365-1368.

Gärdenfors, P. (1999). *Conceptual Spaces*. MIT Press.

Gigerenzer, G. (2000). *Adaptive thinking: Rationality in the real world*. Oxford University Press, New York.

Glaser, R., and Bassok, M. (1989). Learning theory and the study of instruction. *Annual Review of Psychology*, vol. 40, pp. 631-666.

Grim, P., Kokalis, T., Tafti, A. and Kilb, N. (1999). Evolution of communication with a spatialized genetic algorithm. *Evolution of Communication*, vol. 3, no. 2.

Grim, P., Denis, P.S., and Kokalis, T. (2001). *Learning to communicate: The emergence of signaling in spatialized arrays of neural nets.* Technical report, Department of Philosophy, SUNY at Stony Brooks.

Hendriks-Jansen, H. (1996). *Catching ourselves in the act - situated activity, interactive emergence, evolution, and human thought*. MIT Press, Cambridge, MA.

Honkela, T. and Vepsäläinen, A.M. (1991). Interpreting imprecise expressions: experiments with Kohonen's Self-Organizing Maps and associative memory. T. Kohonen, K. Mäkisara, O. Simula and J. Kangas (eds.), *Proceedings of ICANN-91, International Conference on Artificial Neural Networks,* North-Holland, vol. I, pp. 897-902.

Honkela, T. (1993). Neural nets that discuss: a general model of communication based on self-organizing maps. Stan Gielen and Bert Kappen (eds.), *Proceedings of ICANN'93, International Conference on Artificial Neural Networks,* Springer-

Verlag, London, pp. 408-411.

Honkela, T., Pulkki, V. and Kohonen, T. (1995). Contextual Relations of Words in Grimm Tales Analyzed by Self-Organizing Map. Proceedings of International Conference on Artificial Neural Networks, ICANN-95, F. Fogelman-Soulié and P. Gallinari (eds.), EC2 et Cie, Paris, pp. 3-7.

Honkela, T. (1997). Learning to Understand - General Aspects of Using Self-Organizing Maps in Natural Language Processing. *Computing Anticipatory Systems*, D. Dubois (ed.), American Institute of Physics, Woodbury, New York, pp. 563-576.

Honkela, T., Leinonen, T., Lonka, K., and Raike, A. (2000). Self-Organizing Maps and Constructive Learning. *Proceedings of ICEUT'2000, International Conference on Educational Uses of Communication and Information Technologies,* Beijing, China, August 21-25, pp. 339-343.

Honkela, T., Hynnä, K.I., and Knuuttila, T. (2003). Framework for Modeling Partial Conceptual Autonomy of Adaptive and Communicating Agents. *Proceedings of Cognitive Science Conference,* (CD-ROM), Boston, MA.

Hörmann, H. (1986). *Meaning and Context*. Plenum Press, New York.

Hung, C. and Wermter, S. (2003). A Dynamic Adaptive Self-Organising Hybrid Model for Text Clustering. *Proceedings of the Third IEEE International Conference on Data Mining*, pp. 75-82.

Hurford, J. (2002). The Language Mosaic and its Evolution. Language Evolution, M. Christiansen and S. Kirby (eds.), Oxford University Press, pp. 38-57.

Hutchins, E. (1995). *Cognition in the wild*. MIT Press, Cambridge, MA:

Hyvärinen, A. and Honkela, T. (1999). Emotional Disorders in Autonomous Agents? *Proceedings of European Conference on Artificial Life*, pp. 350-354.

Kaipainen, M. (1994). *Dynamics of Musical Knowledge Ecology - Knowing-What and Knowing-How in the World of Sounds*. PhD thesis, University of Helsinki, Helsinki, Finland, Acta Musicologica Fennica 19.

Kohonen, T. (1982). Self-organizing formation of topologically correct feature maps. *Biological Cybernetics*, vol. 43, no. 1, pp. 59-69.

Kohonen, T. (1988). The ’neural’ phonetic typewriter. *Computer*, vol. 21, no. 3, pp. 11-22.

Kohonen, T. (1993). Physiolocigal interpretation of the self-organizing map algorithm. *Neural Networks*, vol. 6. no. 7, pp. 895-905.

Kohonen, T. (2001). *Self-Organizing Maps*, 3rd Ed. Springer.

Kohonen, T., Kaski, S., and Lappalainen, H. (1997). Self-organized formation of various invariant-feature filters in the adaptive-subspace SOM. *Neural Computation*, vol. 9, pp. 1321-1344.

Kopp. L. (2003). *Natural Vision for Artificial Systems, Active Vision and Thought*. PhD Thesis, Lund University Cognitive Science.

Knuuttila, T., Kangasoja, J. and Honkela, T. (2003). Representation and Activity. Abstract for a Symposium in Cognitive Science Conference, *Proceedings of Cognitive Science Conference*, (CD-ROM), Boston, MA.

Lakoff, G. and Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenges to western thought.* Basic Books, New York.

Lindblom, J. and Ziemke, T. (2003). Social Situatedness of Natural and Artificial Intelligence: Vygotsky and Beyond. *Adaptive Behavior*, vol. 11, no. 2, pp. 79-96.

Linell, P. (1982). *The Written Language Bias in Linguistics*. Linköping.

Lisetti, C. L. and Gmytrasiewicz, P. (2002). Can a Rational Agent Afford to be Affectless? A Formal Approach. *Applied Artificial Intelligence, An International Journal*, vol. 16, pp. 577-609.

MacLennan, B. (1995). Continuous formal systems: A unifying model in language and cognition. *Proceedings of the IEEE Workshop on Architectures for Semiotic Modeling and Situation Analysis in Large Complex Systems*, August 27-29, Monterey, CA.

MacWhinney, B. (1998). Models of the emergence of language. *Annual Review of Psychology*, vol. 49, pp. 199-227.

Maturana, H. R. and Varela, F. J. (1980). *Autopoiesis and cognition: The realization of the living*. Reidel, Dordrecht.

Meltzoff, A.N. and Decety, J. (2003). What imitation tells us about social cognition: a rapproachement between developmental psychology and cognitive neuroscience.

Miikkulainen, R. (1993). *Subsymbolic Natural Language Processing: An Integrated Model of Scripts, Lexicon, and Memory*. MIT Press, Cambridge, MA.

Miikkulainen, R. (1997). Self-organizing feature map model of the lexicon. *Brain and Language*, vol. 59, pp. 334-366.

Miyake, A. and Shah, P. (eds.) (1999). *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*. Cambridge University Press, New York.

Nowak, M. A., Komarova, N. L., and Niyogi, P. (2001). Evolution of universal grammar. Science, vol. 291 ,pp. 114-118.

Oja, E. (2002). Unsupervised learning in neural computation. *Theoretical Computer Science*, vol. 287, pp. 187-207.

Pfeifer, R. and Scheier, C. (1999). *Understanding intelligence*. MIT Press, Cambridge, MA.

Picard, R. W. (1997). *Affective Computing*. M.I.T. Press, Cambridge, MA.

Powers, D.M.W. (1997). Unsupervised learning of linguistic structure: an empirical evaluation. International Journal of Corpus Linguistics, vol. 2m no. 1, pp. 91-131.

Regier, T. (1992). *The Acquisition of Lexical Semantics for Spatial Terms: A Connectionist Model of Perceptual Categorization*. University of California at Berkeley, TR-92-062, Berkeley, CA.

Ritter, H. and Kohonen, T.: Self-Organizing Semantic Maps. *Biological Cybernetics*, 61, 1989, pp. 241-254.

Rorty, R. (1979). *Philosophy and the Mirror of Nature*. Princeton University Press, Princeton.

Rumelhart, D.E. and McClelland, J.L. (1986). On learning the past tenses of english verbs. D.E. Rumelhart and J.L. McClelland, (eds.), Parallel Distributed Processing, volume 2, MIT Press, pp. 216-271.

Saad, D. (1999). On-line Learning in Neural Networks. Cambridge University

Press.

Sandholm, T. 1999. Distributed Rational Decision Making. Weiss, G. (ed.), *Multiagent Systems: A Modern Introduction to Distributed Artificial Intelligence*, MIT Press, pp. 201-258.

Scholtes, J.C. (1991). Learning simple semantics by self-organization. *Worknotes of the AAAI Spring Symposium Series on Machine Learning of Natural Language and Ontology*, Palo Alto, CA, March 26-29, American Association for Artificial Intelligence.

Sharkey, N. E. and Ziemke, T. (2001). Mechanistic versus phenomenal embodiment: Can robot embodiment lead to strong AI? *Cognitive Systems Research*, vol. 2, no. 4, pp. 251-262.

Sloman, A. and Croucher, M. (1981). Why robots will have emotions. *Proceedings of the Seventh International Joint Conference on Artificial Intelligence (IJCAI'81)*, Vancouver, B.C., Morgan-Kaufmann, San Mateo, CA., pp. 197-202.

Smolensky, P. (1994). Grammar-based connectionist approaches to language. Cognitive Science.

Steels, L. (1996). Synthesising the origins of language and meaning using coevolution, self-organisation and level formation. J. Hurford, C. Knight and M. Studdert-Kennedy (eds.), *Evolution of Human Language*, Edinburgh Univ. Press, Edinburgh.

Steels, L. and Kaplan, F. (2001). AIBO's first words. The social learning of language and meaning. Gouzoules, H. (ed), *Evolution of Communication*, vol. 4, nr. 1, John Benjamins Publishing Company, Amsterdam.

Tomasello, M. (1999). *The Cultural Origin of Human Cognition*. Harvard University Press, Cambridge, MA.

Varela, F., Thompson, E. and Rosch, E. (1991). *The Embodied Mind: Cognitive Science and Human Experience*. MIT Press, Cambridge, MA.

von der Malsburg, C. (1973). Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik*, vol. 14, pp. 85-100.

Von Foerster, H. (1981). Notes on an epistemology for living things. *Observing Systems*. Intersystems Publications, pp. 257-271. Originally published in 1972 as BCL Report No 9.3., Biological Computer Laboratory, University of Illinois, Urbana.

Von Foerster, H. (1984). Principles of Self-Organization in a Socio-Managerial Context. H. Ulrich und C. J. B. Probst (eds.), *Self-Organization and Management of Social Systems*, Springer, Berlin, pp. 2-24.

Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press, Cambridge, Massachusetts. (Originally published in 1934).

Vygotsky, L.S. (1986). *Thought and Language*. The MIT Press, Cambridge, Massachusetts. (Originally published in 1934).

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, vol. 8. pp. 338-353.

Ziemke (2003). On the Role of Robot Simulations in Embodied Cognitive Science. AISB Journal, vol. 1, no. 4, pp. 389-399.

Zlatev, J. (2002). Meaning = Life (+ Culture) An outline of a unified biocultural theory of meaning. *Evolution of Communication Journal*, vol. 4, no 2.