Journal of Economic Methodology (in print)

SIMULATING PROCESSES OF CONCEPT FORMATION AND COMMUNICATION

Timo HONKELA, Ville KÖNÖNEN,
Tiina LINDH-KNUUTILA and Mari-Sanna PAUKKERI

Adaptive Informatics Research Center
Laboratory of Computer and Information Science
Helsinki University of Technology, P.O.Box 5400
FI-02015 TKK, Espoo, Finland
http://www.cis.hut.fi/research/compcogsys/

Abstract

We propose a theoretical framework for modeling communication between agents that have different conceptual models of their current context. We describe how the emergence of subjective models of the world can be simulated and what the role of language and communication in that process is. We consider, in particular, the role of unsupervised learning in the formation of agents' conceptual models, the relative subjectivity of these models, and the communication and learning processes that lead into intersubjective sharing of concepts. We also discuss some implications of the subjectivity of conceptual learning in the area of economics.

1 Introduction

The main focus of our article is on linguistic communication as a basis for interactions within a community of human or computerized agents. We are interested both in the philosophical and practical aspects, also aiming at certain formalization of the overall framework. We do not, however, consider the language as an autonomous formal system but recognize its neuro-cognitive (Feldman 2006), embodied (Clark 1997, Lakoff and Johnson 1999, Maturana and Varela 1980, Varela, et al. 1991) and socio-cultural nature (Hutchins 1995, Moore and Carling 1988, Rorty 1979, Vygotsky 1978, Vygotsky 1986). We emphasize the statistical and probabilistic nature of semantics. Our approach has some connections with the research on truthlikeness (Przelecki 1976, Kieseppä 1996, Niiniluoto 1987, Zamora Bonilla 2000). However, we tend to emphasize the dynamics of concept formation through learning processes (Honkela 2000) and the probabilistic aspects based on considering the continuum from raw perceptions to symbolic representations (Gärdenfors 2000).

We have earlier conducted research on simulating processes of language learning and emergence (Honkela 2000, Honkela, et al. 2003, Raitio, et al. 2004) and learning and communication in multiagent systems (Könönen 2004a, Könönen 2004b, Könönen and Oja 2004). In this paper, we discuss how the emergence of subjective models of the world can be simulated using different approaches in learning and what the role of communication and language is. We consider, in particular, the role of unsupervised learning in the formation of agents' conceptual models, the original subjectivity of these models, and the communication and learning processes that lead into intersubjective sharing of concepts.

1.1 Multiagent systems

An intelligent agent usually has a purpose or a goal, probably set by the designer of the agent, and the agent tries to act rationally to satisfy the goal. For making rational decisions, the agent has to model its environment. In the following, we consider some specific issues related to the modeling.

In the easiest case, the environment in which the agent is located is static, i.e. all properties of the environment remain constant for the whole lifetime of the agent. However, the situation is often not so simple. The properties of the environment may vary with time, i.e. the environment is called *non-stationary*. The non-stationarity may be due to the environment itself or to the limited resources of the agent. For example, in many real problem instances, the learning agent is not capable of sensing the real state of the environment and thus some relevant properties of the environment remain hidden.

Pfeifer and Scheier stress that the behavior of an agent is always the result of system-environment interaction (Pfeifer and Scheier 1999). It cannot be explained on the basis of internal mechanisms only. They illustrate that the complexity that we as observers attribute to a particular behavior does not always indicate accurately the complexity of the underlying mechanisms. Experiments with very simple robots that merely react to stimuli in their environment have shown that rather complex behavior can emerge.

Another example of non-stationarity are multiagent systems. In these systems, properties of the environment for each individual agent depend on the actions of all agents located in the real environment. Thus for acting rationally, the agents should also model the other agents in the system.

1.2 Language learning and game theory

The focus on modeling of language learning may be on learning morphology (Rumelhart and McClelland 1986, Daugherty and Seidenberg 1994, Creutz and Lagus 2007), syntax (Chen and Honavar 1999), or semantics (Regier 1996, Bailey 1997).The language learning may take place within one agent or in a community of agents through evolution (Nowak, et al. 2001). The game theoretical approach in linguistics and in particular in pragmatics is studied in depth in (A. Benz and van Rooij 2005). Benz and his colleagues discuss various aspects of natural language that are subject to replication, variation and selection, on various timescales that range from minutes to millennia. They focus on cultural (as opposed to biological) evolution on short time scales. They also state the important basic fact: Natural languages are not passed on via biological but via cultural transmission (Benz et al. 2005). First language acquisition is a qualitatively different mode of replication in which iterated learning is the principal mode. Nowak et al. (Nowak, et al. 2002) consider a language learner that grows up in a community of speakers of some language and might acquire another language with a certain probability. This means that those languages will spread in a population that are likely targets of acquisition for children that are exposed to other languages, and are likely to be acquired faithfully themselves. In this approach, language change is considered as a Markov process rather than as evolution through natural selection. In Section 3.1, we discuss issues related to language games in more detail.

1.3 Grounding

In this article, we assume that semantics is considered within a broad, potentially multimodal context (Hörmann 1986) and the symbol grounding problem is taken into account (Harnad 1990).

1.4 Learning paradigms

Different machine learning techniques can be utilized for adding adaptivity to agent-based systems. There are basically three major learning paradigms in machine learning: *supervised learning*, *unsupervised learning* and *reinforcement learning*. In supervised learning, there exists a teacher having knowledge of the environment, in the form of input-output pairs, and the learning system for which the environment is unknown. The teacher provides samples from the environment by giving correct outputs to inputs and the goal of the learning system is to learn to emulate the teacher and to generalize the samples to unseen data. In unsupervised learning, contrary to supervised learning, there exists no external teacher and therefore no correct outputs are provided. Reinforcement learning is located between supervised and unsupervised learning: correct answers are not directly provided to the learning system but it learns features of the environment by continuously interacting with it. The learning system takes actions in the environment and receives reward signals from the environment corresponding to these action selections.

Unsupervised learning (Kohonen 2001, Oja 2002, Powers 1997) appears to be an important component for any realistic language learning system. As the most commonly used neural network model within the unsupervised learning paradigm, the self-organizing map has proven to be a useful in modeling the emergence on conceptual systems (Ritter and Kohonen 1989, Miikkulainen 1993, Honkela, et al. 1995).

In this paper, we propose a theoretical framework that is suitable for modeling communication between agents and for interagent conceptual modeling. We begin with the basic definitions related to the framework in Section 2. Then we proceed to communication models that utilize the proposed framework in Section 3. In Section 4, we discuss utilization of machine learning in the framework and in Section 5 we consider some practical implications of conceptual learning in economics. Finally, in Section 6 we conclude the paper.

2 Basic theoretical framework

In this section we introduce the basic definitions and notation used in our communication model for two agents. The key concept is the agent's internal view of its context, the concept space. The concept space is spanned by a number of features. Here we can use the terminology coined by Gärdenfors calling each feature f_i a quality dimension (Gärdenfors 2000). Dimensionalities of the concept spaces can be different for each agent. Therefore, we denote the features used by agent 1 f_i^1, $i = 1 \ldots N$. Similarly, agent 2 uses features f_i^2, $i = 1 \ldots M$. Thus, the concept space of agent 1 is an N-dimensional metric space C^1, and for agent 2, C^2. In the framework, there exist two distance measures, namely ω and λ. ω gives a distance between two points inside the concept space of the agent, i.e. $\omega : C^i \times C^i \rightarrow$ R, $i = 1, 2$. λ gives a distance between two points in the concept spaces of the different agents, i.e. $\lambda : C^i \times C^j \rightarrow$ R, $i \neq j$.

The symbol space S^1 of agent 1 is its vocabulary that consists of discrete symbols. Similarly, the vocabulary of agent 2 consists of symbols S^2. An agent i has an individual mapping function ξ^i that maps the symbol $s^i \in S^i$ to C^i.

An agent i expresses each symbol $s^i \in S^i$ as a signal d in the signal space D. We assume that the signal space D is multidimensional, continuous and shared between the agents. However, each agent i has an individual mapping function ϕ^i from its vocabulary to the signal space, i.e. $\phi^i : S^i \rightarrow D$ and an inverse mapping ϕ^{-i} from the signal space to the symbol space. A schematic view of the framework can be seen in Fig. 1.

Add figure 1 here.

3 Communication between agents

We categorize communication in two classes: the single-agent communication model and the two-agent communication model. In both cases the context is similar to the one discussed above and the models differ only in how the other agent is modeled. We proceed with a short definition and introduction to language games.

3.1 Language games

The notion of language games was originally introduced by Ludwig Wittgenstein. To him, every occasion of language use is a language game (Wittgenstein 1953). Carlson mentions that the strength of Wittgenstein's ideas lies in the following points (Carlson 1983). The language games connect a calculus to a form of life. Language games are what establish the link between language, conceived as a calculus, to the reality interpreted, described and transformed by it. The second point mentioned by Carlson is that the game comparison brings us the goal-directed character of language use. Another influential early work is the study of conventions by David K. Lewis. He used concepts of game theory to analyze the nature of social conventions (Lewis 1969). Lewis also considered linguistic communication as a system governed by similar conventions. He explained communication as conventions that are truthful with respect to a particular assignment of truth conditions to sentences or other units of communication (Lewis 1969). Lewis also presupposes that speakers are conceptually autonomous which is also the point of view taken in this article.

Linguistic conventions are reflected, among other things, in word usage patterns. The conventions are spread in a population through iterated learning processes as well as thought explicit and implicit decisions. There are various ways how a new word can enter a language. Once a word is part of a language, it is gradually adapted to this language. Benz et al., who consider this in detail, point out the process of establishing a new word is predominantly driven by mature language users (Benz et al. 2005). Somebody introduces the new word, and people start imitating it. The probability of whether the new word catches on depends on various conditions related to the role of the word in the conceptual space and relevant social context.

Steels has adopted language games to the environment of communicating agents, where a coherent lexicon spontaneously emerges in the community and where the lexicon adapts to cope with new meanings (Steels 1996). Gärdenfors and Warglien consider the evolution of semantics in a game-theoretical evolutionary framework (Gärdenfors and Warglien 2006). Vogt and Coumans described three different types of language games: observational game, guessing game and selfish game. In the observational game two agents, the speaker and the hearer, both know the topic of the game (e.g. a color) (Vogt and Coumans 2003). The speaker utters a word denoting the topic (e.g. the name of the color). The game succeeds if the hearer knows the uttered word in the right meaning, and fails otherwise. In the guessing game, only the speaker knows the topic of the game. The speaker utters a word denoting the topic and the hearer must guess, which topic (of a finite number of topics) the speaker means. The agent playing the selfish game has to infer the meanings of words from their co-occurrences indifferent contexts or situations.

The observational game is more commonly referred to as the naming game (Steels and Vogt 1997). The naming game has many modifications, including e.g. analogical naming games (Kaplan 1998), multiple word naming games (Van Looveren 1999), advertising games (Avesani and Agostini 2003), and query-answering games (Agostini and Avesani 2004).

Smith introduces the discrimination game algorithm, which is produced by Steels (Smith 2001). In the discrimination game, the topic of the game is tried to distinguish from the other topics in the context. Discrimination games can be used together with guessing games in their multi-agent simulations: each agent is able to perceive, categorize its perceptions (discrimination game) and then lexicalize the resulting categories to the other agents (guessing game) (Belpaeme 2001, Belpaeme and Bleys 2005).

From the viewpoint of spoken language, a more general language game method is used by de Boer. His game is called imitation game and the utterances in the game consist of phoneme sounds instead of written letters (deBoer 1997). While playing the game, the first agent utters a word and the other tries to imitate it.

Jäger uses evolutionary game theory as a mathematical framework to model the consequences of interaction between individuals in a population (Jäger 2006). He considers conditions under which the interaction between individuals can be modeled as a strategic game, where utility can be identified with fitness. Jäger builds on Gärdenfors' theory of conceptual spaces (Gärdenfors 2000) that suggests that simple adjectives usually denote natural properties, where a natural property is a convex region of a conceptual space. Jäger shows that under some natural assumptions about the utility function, convexity of meanings falls out as a consequence of evolutionary stability (Jäger 2006).

Next, we present our formalization of the concept formation and communication processes.

3.2 Single-agent model

In the single-agent communication model, the sender is communicating with the receiver but it has only its own conceptual space available. The sender, agent 1, selects a symbol that corresponds best to the current context by the means of its own conceptual model. Therefore this model can be seen as an optimal decision problem.

Let us denote the symbol agent 1 selects and communicates to agent 2 as s^* and the features agent 1 observes corresponding to the current context as f^1.Then agent 1 selects the symbol that corresponds the current observations best by the means of some distance measure ω. A suitable choice for ω in that case is for example Euclidean distance. Formally this can be represented as in Eq. (1).

$$s^* = \arg\min_{s \in S^1} \omega\left(f^1, \xi^1(s)\right) \tag{1}$$

After symbol selection process, agent 1 communicates the symbol s* to agent 2, i.e.:

$$d = \phi^1(s^*) \tag{2}$$

When agent 2 observes d, it maps it to some $s^2 \in S^2$ by using the function ϕ^{-2}. Then it maps the symbol to some point in its conceptual space by using ξ^2. If this point is very near of its own observation f^2, we can say that the communication process has succeeded. Mathematically this is as follows:

$$\left\| \xi^2\left(\phi^{-2}(d)\right) - f^2 \right\| \leq \varepsilon\,, \tag{3}$$

where ε is a small constant and $\| \cdot \|$ is some suitable norm in C^2.

3.3 Two-agent model

The two-agent model relates to the single-agent model very closely but now the sender has some estimate of the receiver's conceptual space available. This model can be learned from communication samples or it can be known a priori. The symbol selection process is formally given in Eq. (4). In this equation, $\tilde{\xi}^2$ is the model of the receiver. Note that also the vocabulary of the receiver can be unknown and should be estimated by $\tilde{S}^2$.

$$s^* = \arg\min_{s \in \tilde{S}^2} \lambda\left(f^1, \tilde{\xi}^2(s)\right) \tag{4}$$

As the dimensionalities of the conceptual spaces C^1 and C^2 can be different, we should use some special method to calculate the distance between points in these two spaces. Dynamic programming and neural networks may provide solutions to this task. Note that in the special case where the dimensionality of the concept spaces is the same for both agents and the dimensions are the same, Euclidean distance may be used for this purpose. This kind of situation should be considered highly artificial, though. A neural network solution for the general case when agents have different representations for the same phenomenon has been considered for a supervised learning task (Laakso and Cottrell 2000) as well as for unsupervised learning (Raitio et al. 2004).

Other parts of the communication process remain the same as in the single-agent case. As agent 1 utilizes the explicit model of agent 2, the symbol selection problem can be seen as a game theoretical problem. Due to its sequential nature, it has a relationship with Stackelberg style games where players have different roles, namely the roles of the leader (corresponding to the sender) and the follower (corresponding to the receiver).

The messaging behavior of the agents can be called a language game. The aim of a language game is to create common language in an agent community, and thus make it possible for the agents to understand each other. Two agents participating in a language game both see the same context around them and if they are able to agree with the symbol corresponding to the context, we can say that the language game was successful.

4 Learning of conceptual models

In this section, we discuss the position of machine learning in our framework for agent communication. We start the section with a short introduction to machine learning and then we proceed to the applications of unsupervised learning in conceptual modeling. We also provide pointers to relevant research work on conceptual modeling.

4.1 Unsupervised learning of conceptual systems

In the following, we consider the unsupervised learning of conceptual systems. We first study the modeling an individual agent that learns to create a conceptual space of its own and learns to associate words and expressions with the conceptual space. After considering one individual agent, we consider a multiagent system in which a shared conceptual system is formed in a self-organized manner.

4.1.1 Single-agent learning

As an example related to the vocabulary problem, two persons may have different conceptual or terminological ‘density’ of the topic under consideration. A layman, for instance, is likely to describe a phenomenon in general terms whereas an expert uses more specific terms. In our framework, symbols generate clusters in the conceptual spaces of the agents.

If some agents speak the ‘same language’, many of the symbols in their vocabularies are the same. However, for the sake of realistic theory formation, we do not assume that the vocabularies of any two agents are exactly the same. In our formal framework, the intersection of the sets S^1 and S^2 may contain only a small number of symbols or be even empty. The basic approach how autonomous agents could learn to communicate and form an internal model of the environment applying the self-organizing map algorithm was introduced, in a simple form, in (Honkela 1993).

4.1.2 Community of agents

Later we extended the framework that would enable modeling the degree of conceptual autonomy of natural and artificial agents (Honkela et al. 2003). The basic claim was that the aspects related to learning and communication necessitate adaptive agents that are partially autonomous. We demonstrated how the partial conceptual autonomy can be obtained through a self-organization process. The input for the agents consists of perceptions of the environment, expressions communicated by other agents as well as the recognized identities of other agents (Honkela et al. 2003). This information is used for computing the mappings ξ^1 and $\tilde{\xi}$ in our communication model.

When language games were included in the simulation model, it resulted in a simple language emerging in a population of communicating autonomous agents (Lindh-Knuutila, et al. 2006). In this population, each agent was able to create their own associations between the conceptual level and the emerged words, although each agent had a slightly different conceptual representation of the world. The learning paradigm for the conceptual learning was fundamentally unsupervised, but the language learning tested has so far been supervised, i.e. the communicating agents are provided feedback of the outcome of the game as well as the 'right answer'. Reinforcement learning and unsupervised learning models for language games remain to be implemented.

5 Practical implications

Here we will discuss the potential practical implications of the rather theoretical and abstract framework that has been presented earlier in this paper. The implications and applications are considered, in particular, in the area of economics. There is a lot of research related to the applications of game theory in economics (Bierman and Fernandez 1993). Moreover, very many applications of neural networks and statistical machine learning exist (consider, e.g., Deboeck and Kohonen 1998). This article focuses on a theme that brings together aspects of both game theory and learning systems such as neural networks. The connecting theme is the conceptual systems of the interacting agents in a community.

It has been commonplace to consider uncertainty in decision making (Gollier 2001, Parsons 2001): an agent estimates probabilities of different outcomes and their values. In the following, we discuss two specific topics related to the certain uncertainty in the use of conceptual systems.

5.1 Meaning negotiations

The traditional notion of uncertainty in decision making does not cover the uncertainties caused by differences in conceptual systems of individual agents within a community. We claim that in all transactions including symbolic/linguistic communication the differences in the underlying conceptual systems play an important role. For instance, serious efforts have been made to harmonize or to standardize the classification systems used by business agents, e.g., using Semantic Web technologies (Sintek and Decker 2003). However, even if the standardization is conducted, there can not be any true guarantee that all the participating agents would share the meaning of all the expressions used in the business transactions in various contexts.

Add figure 2 here.

We can consider an example illustrated by Fig. 2: a buyer agent expresses the wish for finding an item belonging to a specific category, represented as an area on the diagram. It may very well be that the selling agent understands the query differently and therefore the items considered as candidates by the selling agent differs from the buyers intentions. Within traditional epistemological theory, this situation could be considered through set theory: by a symbol the agents refer to different sets of items. However, as many features (or quality dimensions) are continuous, it appears to be more natural to consider the situation within a continuous multidimensional space. Moreover, the exact features of an item may not be known but they need to be considered as a probability distribution.

One implication is that in business transactions there should be means for checking what is a meant by some expressions via an access to a broader context (cf. symbol grounding). Moreover, rather than relying solely on a standardized conceptual system, one could introduce mechanisms of meaning negotiation. Before two business agents get into negotiation about, for instance, the price of some commodity, they should first check if they agree on what they refer to by the expressions that are used in the negotiation. This concern is valid both for human and computerized systems, even though humans are usually capable of conducting meaning negotiations even when they are not aware of it.

The need for meaning negotiations becomes even more obvious when one considers qualities like being ‘interesting’ or ‘beautiful’. The concepts of interestingness and beauty are highly subjective and the criteria used may depend on very different underlying features. These kinds of concepts are, however, the basis of a vast number of daily business related transactions. For instance, when a person is considering a gift for her spouse, she is usually, at least implicitly, considering the taste of her spouse, for which she has a mental model formed over time.

5.2 Costs associated with harmonization of conceptual systems

Here we discuss the costs of creating and learning a shared conceptual system. The harmonization of conceptual systems, such as the creation of ontologies for business transactions, has obvious benefits when, for instance, the interoperability of related information systems is considered. It appears ideal that all systems within some domain would use similar terminologies and shared ontologies. However, this approach can be claimed to be idealistic because the continuous change through innovations and other activities and the underlying learning processes within the human community lead into the situation carefully considered in the earlier sections of this paper. All the agents have a conceptual system of their own, at least to some degree. Therefore, the harmonization of the conceptual systems should be considered only as a relative goal. One may aim for a larger degree of sharing of the conceptual system as before. A central theme is then to assess the associated benefits and costs. Here we do not try to provide any means to estimate the benefit of well working harmonized conceptual system implemented, e.g., as an ontology within the Semantic Web framework.

The costs stem from two main sources: the development of a shared conceptual system and the use of it. The development of an ontology typically consists of defining the concepts and the relationships between the concepts. The typical stages of an ontology building process are the following (Fernández-López and Gómez-Pérez 2002): (1) domain analysis resulting into the requirements specification, (2) conceptualization resulting into the conceptual model, (3) implementation that leads into the specification of the conceptual model in the selected representation language, and (4) the ontology population i.e. the generation of instances and their alignment to the model that results into the instantiated ontology. Ontology maintenance includes getting familiar with and modifying the ontology and ontology reuse involves costs for the discovery and reuse of existing ontologies in order to generate a new ontology (Fernández-López and Gómez-Pérez 2002). Simperl with colleagues present a cost estimation model for ontology engineering. They estimate the person months associated to building, maintaining and reusing ontologies calculated as the product of the size of the ontology, expressed in thousands of ontological primitives, and the cost drivers with some effort multipliers that are typically expressed in months or even years (Simperl, et al. 2006). This emphasizes the fact that manual ontology engineering is a costly process.

The estimation of costs related to the use of ontologies is rather difficult. There are many kinds of uses of ontologies that require a higher or a lower degree of familiarity of conceptual structures of the domain. A widely cited claim from expertise research is the 10-year rule, first proposed in relation to expertise development among chess players (Simon and Chase 1973), and later generalized to other domains (Bloom 1985). The essential content of the rule is that anyone seeking to perform at world-class level in any significant domain must engage in sustained, deliberate practice in the activity for a period of at least ten years. This figure serves only as an upper bound of an estimate for a person to learn to master the conceptual content of a complex domain.

In summary, we wish to emphasize that for an organization, the costs related to ontologies are not limited to the development and maintenance of them but also include the use of ontologies by even a large group of people. Each person may require months or even years to familiarize herself with an ontology. This may be a natural part of the task at hand, but it also may be an additional cost for a person who already masters another conceptual system within the same domain. Therefore, it may not be sufficient that one develops methods for automated acquisition of ontologies. We suggest that methods and tools supporting meaning negotiations are a cost-effective means for dealing with knowledge-intensive task contexts in organizations.

6 Discussion

Language does not need to be viewed plainly as a means for labeling the world but as an instrument by which the society and the individuals within it construct a model of the world. The world is continuous and changing. Thus, the language is a medium of abstraction rather than a tool for creation and mediation of an accurate ‘picture’ of the world. The point of view chosen is always subject to some criteria of relevance or usefulness. This is not only true for the individual expressions that are used in communication but also concerns the creation or emergence of conceptual systems. It makes sense to make such distinctions in a language that are useful in one way another.

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)

Realistic simulations of the socio-economical and cultural levels 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, Vygotsky 1986, Salomon 1991, Lindblom and Ziemke 2003, Hakkarainen, et al. 2004). For instance, Vygotsky has stated that ‘… the world of experience must be greatly simplified and generalized before it can be translated into symbols. Only in this way does communication become possible, for the individual's experience resides only in his own consciousness and is, strictly speaking, not communicable.’ (Vygotsky 1986) Later, he continues: ‘The relation of thought to word is not a thing but a process, a continual movement back and forth from thought to word and from word to thought. In that process the relation of thought to word undergoes changes which themselves may be regarded as development in the functional sense.’ This means in practice that conceptualization is a complex process that takes place in a socio-cultural context, i.e., within a community of interacting individuals whose activities result into various kinds of cultural artifacts such as written texts.

In this paper we proposed a theoretical framework for communication between agents that have different concept spaces of their context. In addition, we discussed the position of machine learning, in particular unsupervised learning, in the framework. We also proposed a possible application of our framework in the area of economics. In the current state of our research, we have only considered supervised and unsupervised learning with our framework. In the future, we will study the possibility to employ reinforcement learning with language games.

Acknowledgments

We wish to thank Prof. Jesús Zamora Bonilla who kindly invited us to give a presentation in the Economics and Language workshop in Madrid in June 2006. This article is an extension of the paper presented in the workshop. We owe a debt of gratitude to the anonymous reviewers who provided thoughtful advice on how to improve the manuscript; any remaining errors are our own. We want to thank Maarten Peeters for his help in finalizing the paper.

This work was supported by the Academy of Finland through the Finnish Centre of Excellence Programme.

References

A. Agostini and P. Avesani (2004). `On the discovery of the semantic context of queries by game-playing'. In *Sixth International Conference on Flexible Query Answering Systems (FQAS-04)*, vol. 3055 of LNAI, pp. 203-216, Berlin Heidelberg. Springer-Verlag.

P. Avesani and A. Agostini (2003). `A Peer-to-Peer Advertising Game'. In *Proceedings of the First International Conference on Service Oriented Computing (ICSOC-03)*, pp. 28-42. Springer-Verlag.

D. R. Bailey (1997). *When Push Comes to Shove: A Computational Model of the Role of Motor Control in the Acquisition of Action Verbs*. Ph.D. thesis, Berkeley, CA.

T. Belpaeme (2001). `Reaching coherent color categories through communication'. In B. K. et al. (ed.), *Proceedings of the 13th Belgium-Netherlands Conference on Artificial Intelligence (BNAIC'01)*, pp. 41-48, Amsterdam, The Netherlands.

T. Belpaeme and J. Bleys (2005). `Colourful language and colour categories'. In A. Cangelosi and C. L. Nehaniv (eds), *Proceedings of the Second International Symposium on the Emergence and Evolution of Linguistic Communication*, vol. 1, pp. 1-7, Hatfield, UK. The University of Hertfordshire.

A. Benz , G. Jäger and R. van Rooij (2005). *Game Theory and Pragmatics*, chap. An Introduction to Game Theory for Linguists. Palgrave Studies in Pragmatics, Language and Cognition. Palgrave MacMillan.

H. Bierman and L. Fernandez (1993). *Game Theory with Economic Applications*. Addison-Wesley.

B. Bloom (1985). *Developing talent in young people*. Ballantine, New York.

L. Carlson (1983). *Dialogue Games: An Approach to Discourse Analysis*. Reidel, Dordrecht.

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

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

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

M. Creutz and K. Lagus (2007). `Unsupervised Models for Morpheme Segmentation and Morphology Learning'. *ACM Transactions on Speech and Language Processing.*

K. Daugherty and M. Seidenberg (1994). `Beyond Rules and Exceptions: A Connectionist Approach to Inflectional Morphology'. In S. Lima, R. Corrigan, and G. Iverson (eds), *The Reality of Linguistic Rules*, pp. 358-388. John Benjamins, Amsterdam.

B. de Boer (1997). `Generating vowel systems in a population of agents'. In *Fourth European conference on Artificial Life*, pp. 503-510, Cambridge, MA. MIT Press.

G. Deboeck and T. Kohonen (1998). *Visual Explorations in Finance with Self-Organizing Maps*. Springer-Verlag.

J. A. Feldman (2006). *From Molecule to Metaphor: A Neural Theory of Language*. MIT Press, Cambridge, MA.

M. Fernández-López and A. Gómez-Pérez (2002). `Overview and analysis of methodologies for building ontologies'. *Knowledge Engineering Review* 17:129-156.

P. Gärdenfors (2000). *Conceptual Spaces*. MIT Press.

P. Gärdenfors and M. Warglien (2006). `Cooperation, Conceptual Spaces and the Evolution of Semantics'. In *Symbol Grounding and Beyond: Proceedings of the Third International Workshop on the Emergence and Evolution of Linguistic Communication*, pp. 16-30. Springer.

C. Gollier (2001). *The Economics of Risk and Time*. MIT Press.

K. Hakkarainen, T. Palonen, S. Paavola and E. Lehtinen (2004). *Communities of networked expertise: Professional and educational perspectives*. Elsevier.

S. Harnad (1990). `The symbol grounding problem'. *Physica D* 42:335-346.

R. Hilpinen (1976). *Formal Methods in the Methodology of the Empirical Sciences*, chap. Approximate truth and truthlikeness. Reidel, Dordrecht.

T. Honkela (1993). `Neural nets that discuss: a general model of communication based on self-organizing maps'. In S. Gielen and B. Kappen (eds), *Proceedings of ICANN'93, International Conference on Artificial Neural Networks*, pp. 408-411, Amsterdam, the Netherlands. Springer-Verlag, London.

T. Honkela (2000). *Hybrid Neural Systems*, chap. Self-Organizing Maps in Symbol Processing, pp. 348-362. Springer.

T. Honkela, V. Pulkki and T. Kohonen (1995). `Contextual relations of words in Grimm tales analyzed by self-organizing map'. In *Proceedings of ICANN-95, International Conference on Artificial Neural Networks*, vol. 2, pp. 3-7, Paris, France. EC2 et Cie.

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

T. Honkela, K. I. Hynnä and T. Knuuttila (2003). `Framework for Modeling Partial Conceptual Autonomy of Adaptive and Communicating Agents'. In *Proceedings of Cognitive Science Conference*, Boston, MA.

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

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

G. Jäger (2006). `Convex meanings and evolutionary stability'. In The Evolution of Language. *Proceedings of the 6th International Conference (EVOLANG6)*, pp. 139-144. World Scientific.

F. Kaplan (1998). `A New Approach to Class Formation in Multi-Agent Simulations of Language Evolution'. In Y. Demazeau (ed.), *Proceedings of the Third International Conference on Multi Agent Systems (ICMAS98)*, pp. 158-165, Los Alamitos, CA. IEEE Computer Society.

I. A. Kieseppä (1996). *Truthlikeness for Multidimensional, Quantitative Cognitive Problems*. Kluwer, Dordrecht.

T. Kohonen (2001). *Self-Organizing Maps*, vol. 30 of *Springer Series in Information Sciences*. Springer, Berlin, Heidelberg.

V. J. Könönen (2004a). `Asymmetric multiagent reinforcement learning'. *Web Intelligence and Agent Systems: An International Journal (WIAS)* 2(2):105-121.

V. J. Könönen (2004b). `Policy Gradient Method for Team Markov Games'. In *Proceedings of the Fifth International Conference on Intelligent Data Engineering and Automated Learning (IDEAL-2004)*, pp. 733-739, Exeter, UK.

V. J. Könönen and E. Oja (2004). `Asymmetric Multiagent Reinforcement Learning in Pricing Applications'. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN-2004)*, vol. 2, pp. 1097-1102, Budapest, Hungary.

A. Laakso and G. Cottrell (2000). `Content and cluster analysis: assessing representational similarity in neural systems'. *Philosophical psychology* 13(1):47-76.

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

D. K. Lewis (1969). *Convention: A Philosophical Study*. Harvard University Press.

J. Lindblom and T. Ziemke (2003). `Social Situatedness of Natural and Artificial Intelligence: Vygotsky and Beyond'. *Adaptive Behavior* 11(2):79-96.

T. Lindh-Knuutila, T. Honkela and K. Lagus (2006). `Simulating Meaning Negotiation Using Observational Language Games'. In P. V. et al. (ed.), *Symbol Grounding and Beyond: Proceedings of the Third International Workshop on the Emergence and Evolution of Linguistic Communication*, pp. 168-179, Rome, Italy. Springer, Berlin/Heidelberg.

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

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

T. Moore and C. Carling (1988). *The Limitations of Language*. Macmillan Press, Houndmills.

I. Niiniluoto (1987). *Truthlikeness*. Reidel, Dordrecht.

M. A. Nowak, N. L. Komarova and P. Niyogi (2001). `Evolution of universal grammar'. *Science* 291:114-118.

M. A. Nowak, N. L. Komarova and P. Niyogi (2002). `Computational and evolutionary aspects of language'. *Nature* 417:611-617.

E. Oja (2002). `Unsupervised learning in neural computation'. *Theoretical Computer Science* 287:187-207.

S. Parsons (2001). *Qualitative approaches to reasoning under uncertainty*. MIT Press, Cambridge, MA.

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

D. M. Powers (1997). `Unsupervised learning of linguistic structure: an empirical evaluation'. *International Journal of Corpus Linguistics* 2:91-131.

J. Raitio, R. Vigário, J. Särelä and T. Honkela (2004). `Assessing similarity of emergent representations based on unsupervised learning'. In *Proceedings of IJCNN 2004, International Joint Conference on Neural Networks*, vol. 1, Budapest, Hungary.

T. Regier (1996). *The Human Semantic Potential*. MIT Press, Cambridge, MA.

H. Ritter and T. Kohonen (1989). `Self-organizing semantic maps'. *Biological Cybernetics* 61(4):241-254.

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

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

H. A. Simon and W. G. Chase (1973). `Skill in chess'. *American Scientist* 61:391-403.

E. P. B. Simperl, C. Tempich and Y. Sure (2006). `A Cost Estimation Model for Ontology Engineering'. In *International Semantic Web Conference 2006*, pp. 625-639.

M. Sintek and S. Decker (2003). `Using TRIPLE for business agents on the Semantic Web'. *Electronic Commerce Research and Applications* 2(4):315-322.

A. D. M. Smith (2001). `Establishing Communication Systems without Explicit Meaning Transmission'. In J. Kelemen and P. Sosík (eds), *ECAL01, Lectures Notes in Computer Science*, pp. 381-390, Prague. Springer.

L. Steels (1996). `Emergent adaptive lexicons'. In P. Maes, M. J. Mataric, J. A. Meyer, J. Pollack, and S. W. Wilson (eds), *Proceedings of the Simulation of Adaptive Behavior Conference, SAB96*, pp. 562-567, Cambridge, MA. MIT Press.

L. Steels and P. Vogt (1997). `Grounding adaptive language games in robotic agents'. In I. Harvey and P. Husbands (eds), *ECAL97*, Cambridge, MA. MIT Press.

J. Van Looveren (1999). `Multiple-Word Naming Games'. In E. Postma and M. Gyssens (eds), *Proceedings of the 11th Belgium-Netherlands Conference on Artificial Intelligence (BNAIC 99)*, Maastricht, the Netherlands.

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

P. Vogt and H. Coumans (2003). `Investigating social interaction strategies for bootstrapping lexicon development'. *Journal of Artificial Societies and Social Simulation* 6(1).

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

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

L. Wittgenstein (1953). *Philosophical Investigations*. Blackwell Publishers, Oxford.

J. Zamora Bonilla (2000). `Truthlikeness, Rationality and Scientific Method'. *Synthese* 122:321-335.

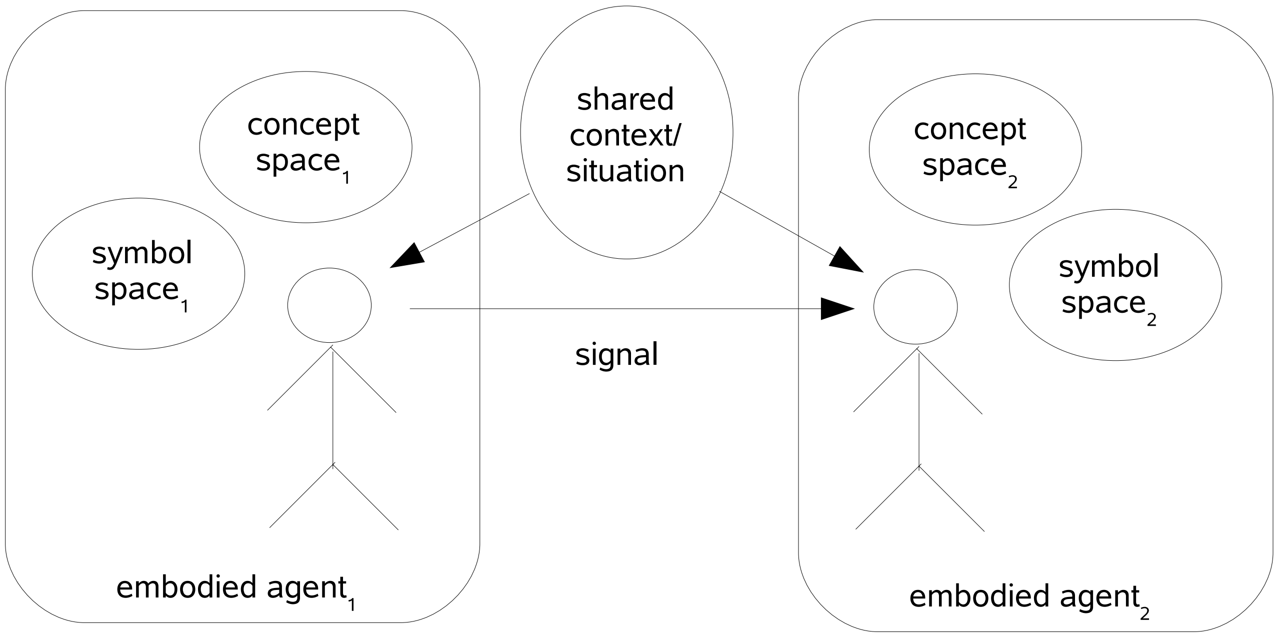

Figure 1: Overview of the framework. Two agents, the sender and the receiver, are observing the same context and they have their own internal conceptual models. The sender sends a message to the receiver that interprets it. The communication has succeeded if both agents agree with the concept.

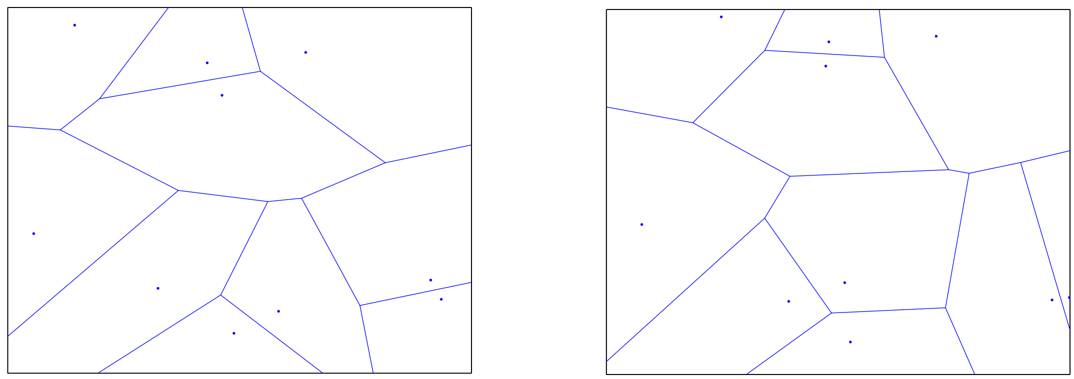

Figure 2: Two agents with differing conceptual systems.