

Language and Cognition Integration through Modeling Field Theory: Category Formation for Symbol Grounding

Vadim Tikhanoff ⁽¹⁾, José F. Fontanari ⁽²⁾, Angelo Cangelosi ⁽¹⁾,
Leonid I. Perlovsky ⁽³⁾

(1) Adaptive Behaviour & Cognition, University of Plymouth, Plymouth PL4 8AA, UK

(2) Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, SP, Brazil

(3) Air Force Research Laboratory, Hanscom Air Force Base, MA 01731, USA

vadim.tikhanoff@plymouth.ac.uk; fontanari@ifsc.usp.br; acangelosi@plymouth.ac.uk;

leonid.perlovsky@hanscom.af.mil

Neural Modeling Field Theory is based on the principle of associating lower-level signals (e.g., inputs, bottom-up signals) with higher-level concept-models (e.g. internal representations, categories/concepts, top-down signals) avoiding the combinatorial complexity inherent to such a task. In this paper we present an extension of the Modeling Field Theory neural network for the classification of objects. Simulations show that (i) the system is able to dynamically adapt when an additional feature is introduced during learning, (ii) that this algorithm can be applied to the classification of action patterns in the context of cognitive robotics and (iii) that it is able to classify multi-feature objects from complex stimulus set. The use of Modeling Field Theory for studying the integration of language and cognition in robots is discussed.

Introduction

Grounding language in categorical representations

A growing amount of research on interactive intelligent systems and cognitive robotics is focusing on the close integration of language and other cognitive capabilities [1,3,13]. One of the most important aspects in language and cognition integration is the grounding of language in perception and action. This is based on the principle that cognitive agents and robots learn to name entities, individuals and states in the external (and internal) world whilst they interact with their environment and build sensorimotor representations of it. For example, the strict relationship between language and action has been demonstrated in various empirical and theoretical studies, such as psycholinguistic experiments [10], neuroscientific studies [16] and language evolution theories [17]. This link has also been demonstrated in computational models of language [5,21].

Approaches based on language and cognition integration are based on the principle of grounding symbols (e.g. words) in internal meaning representations. These are

normally based on categorical representations [11]. Much research has been dedicated on modeling the acquisition of categorical representation for the grounding of symbols and language. For example, Steels [19,20] has studied the emergence of shared languages in group of autonomous cognitive robotics that learn categories of objects. He uses discrimination tree techniques to represent the formation of categories of geometric shapes and colors. Cangelosi and collaborators have studied the emergence of language in multi-agent systems performing navigation and foraging tasks [2], and object manipulation tasks [6,12]. They use neural networks that acquire, through evolutionary learning, categorical representations of the objects in the world that they have to recognize and name.

Modeling Field Theory

Current grounded agent and robotic approaches have their own limitations. For example, one important issue is the scaling up of the agents' lexicon. Present models can typically deal with a few tens of words (e.g. [20]) and with a limited set of syntactic categories (e.g. nouns and verbs in [2]). This is mostly due to the use of computational intelligent techniques, the performance of which is considerably degraded by the combinatorial complexity (CC) of this problem. The issue of scaling up and combinatorial complexity in cognitive systems has been recently addressed by Perlovsky [14]. In linguistic systems, CC refers to the hierarchical combinations of bottom-up perceptual and linguistic signals and top-down internal concept-models of objects, scenes and other complex meanings. Perlovsky proposed the neural Modeling Field Theory (MFT) as a new method for overcoming the exponential growth of combinatorial complexity in the computational intelligent techniques traditionally used in cognitive systems design. Perlovsky [15] has suggested the use of MFT specifically to model linguistic abilities. By using concept-models with multiple sensorimotor modalities, a MFT system can integrate language-specific signals with other internal cognitive representations.

Modeling Field Theory is based on the principle of associating lower-level signals (e.g., inputs, bottom-up signals) with higher-level concept-models (e.g. internal representations, categories/concepts, top-down signals) avoiding the combinatorial complexity inherent to such a task. This is achieved by using measures of similarity between concept-models and input signals together with a new type of logic, so-called dynamic logic. MFT may be viewed as an unsupervised learning algorithm whereby a series of concept-models adapt to the features of the input stimuli via gradual adjustment dependent on the fuzzy similarity measures.

A MFT neural architecture was described in [14]. It combines neural architecture with models of objects. For feature-based object classification considered here, each input neuron $i = 1, \dots, N$ encodes feature values O_i (potentially a vector of several features); each neuron i may contain a signal from a real object or from irrelevant context, clutter, or noise. We term the set $O_i, i = 1, \dots, N$ an input neural field: it is a set of bottom-up input signals. Top-down, or priming signal-fields to these neurons are generated by models, $M_k(S_k)$ where we enumerate models by index $k = 1, \dots, M$. Each model is characterized by its parameters S_k , which may also be a vector of

several features. In this contribution we will consider the simplest possible case, in which parameters model represent feature values of object, $M_k(S_k) = S_k$. Interaction between bottom-up and top-down signals is determined by neural weights associating signals and models as follows. We introduce an arbitrary similarity measure $l(i|k)$ between bottom-up signals O_i and top-down signals S_k [see equation (2)], and define the neural weights by

$$f(k|i) = l(i|k) / \sum_{k'} l(i|k'). \quad (1)$$

These weights are functions of the model parameters S_k , which in turn are dynamically adjusted so as to maximize the overall similarity between object and models. This formulation sets MFT apart from many other neural networks.

Recently, MFT has been applied to the problem of categorization and symbol grounding in language evolution models. Fontanari and Perlovsky [7] use MFT as an alternative categorization and meaning creation method to that of discrimination trees used by Steels [19]. They consider a simple world composed of few objects characterized by real-valued features. Whilst in Steels's work each object is defined by 9 features (e.g. vertical position, horizontal, R, G and B color component values), here each object consists of a real-valued number that identifies only one feature (sensor). The task of the MFT learning algorithm is to find the concept-models that best match these values. Systematic simulations with various numbers of objects, concept-models and object/model ratios, show that the algorithm can easily learn the appropriate categorical model. This MFT model has been recently extended to study the dynamic generation of concept-models to match the correct number of distinct objects in a complex environment [8]. They use the Akaike Information Criterion to gradually add concept-models until the system settles to the correct number of concepts, which corresponds to the original number of distinct objects defined by the experimenter. This method has been applied to complex classification tasks with high degree of variance and overlap between categories. Fontanari and Perlovsky [9] have also used MFT in simulations on the emergence of communication. Meanings are created through MFT categorization, and word-meaning associations are learned using two variants of the overter procedure [18], in which the agents may, or may not, receive feedback about the success of the communication episodes. They show that optimal communication success is guaranteed in the supervised scheme, provided the size of the repertoire of signals is sufficiently large, though only a few signals are actually used in the final lexicon.

MFT for categorization of multi-dimensional object feature representations

The above studies have demonstrated the feasibility of using MFT to model symbol grounding and fuzzy similarity-based category learning. However, the model has been applied to a very simplified definition of objects, each consisting of one feature. Simulations have also been applied to a limited number of categories (concept-models). In more realistic contexts, perceptual representations of objects

consist of multiple features or complex models for each sensor, or result from the integration of different sensors. For example, in the context of interactive intelligent systems able to integrate language and cognition, their visual input would consist of objects with a high number of dimensions or complex models. These could be low-level vision features (e.g. individual pixel intensities), or some intermediate image processing features (e.g. edges and regions), or higher-level object features (color, shape, size etc.). In the context of action perception and imitation, a robot would have to integrate various input features from the posture of the teacher robot to identify the action or complex models (e.g. [6]). The same need for multiple-feature objects applies to audio stimuli related to language/speech. In addition, the interactive robot would have to deal with hundreds, or thousands, categories, and with high degrees of overlap between categories.

To address the issue of multi-feature representation of objects and that of the scaling up of the model we have extended the MFT algorithm to work with multiple-feature objects. We consider both the cases in which all features are present from the start, and the case in which the features are dynamically added during learning. For didactic purposes, first we will carry out simulations on very simple data sets, and then on data related to the problem of action recognition in interactive robots. Finally, we will present some results on the scale up of the model, using hundred of objects.

The Model

We consider the problem of categorizing N objects $i=1,\dots,N$, each of which characterized by d features $e=1,\dots,d$. These features are represented by real numbers $O_{ie} \in (0,1)$ - the input signals - as described before. Accordingly, we assume that there are M d -dimensional concept-models $k=1,\dots,M$ described by real-valued fields S_{ke} , with $e=1,\dots,d$ as before, that should match the object features O_{ie} . Since each feature represents a different property of the object as, for instance, color, smell, texture, height, etc. and each concept-model component is associated to a sensor sensitive to only one of those properties, we must, of course, seek for matches between the same component of objects and concept-models. Hence it is natural to define the following partial similarity measure between object i and concept k

$$l(i|k) = \prod_{e=1}^d (2\pi\sigma_{ke}^2)^{-1/2} \exp\left[-(S_{ke} - O_{ie})^2 / 2\sigma_{ke}^2\right] \quad (2)$$

where, at this stage, the fuzziness σ_{ke} is a parameter given *a priori*. The goal is to find an assignment between models and objects such that the global similarity

$$L = \sum_i \log \sum_k l(i|k) \quad (3)$$

is maximized. This maximization can be achieved using the MFT mechanism of concept formation which is based on the following dynamics for the modeling field components

$$dS_{ke}/dt = \sum_i f(k|i) [\partial \log l(i|k) / \partial S_{ke}], \quad (4)$$

which, using the similarity (1), becomes

$$dS_{ke}/dt = - \sum_i f(k|i) (S_{ke} - O_{ie}) / \sigma_{ke}^2. \quad (5)$$

Here the fuzzy association variables $f(k|i)$ are the neural weights defined in equation (1) and give a measure of the correspondence between object i and concept k relative to all other concepts k' . These fuzzy associations are responsible for the coupling of the equations for the different modeling fields and, even more importantly for our purposes, for the coupling of the distinct components of a same field. In this sense, the categorization of multi-dimensional objects is not a straightforward extension of the one-dimensional case because new dimensions should be associated with the appropriate models. This nontrivial interplay between the field components will become clearer in the discussion of the simulation results.

It can be shown that the dynamics (4) always converges to a (possibly local) maximum of the similarity L [14], but by properly adjusting the fuzziness σ_{ke} the global maximum often can be attained. A salient feature of dynamic logic is a match between parameter uncertainty and fuzziness of similarity. In what follows we decrease the fuzziness during the time evolution of the modeling fields according to the following prescription

$$\sigma_{ke}^2(t) = \sigma_a^2 \exp(-\alpha t) + \sigma_b^2 \quad (6)$$

with $\alpha = 5 \times 10^{-4}$, $\sigma_a = 1$ and $\sigma_b = 0.03$. Unless stated otherwise, these are the parameters we will use in the forthcoming analysis.

Simulations

In this section we will report results from three simulations. The first will use very simple data sets that necessitate the use of two features to correctly classify the input objects. We will demonstrate the gradual formation of appropriate concept-models through the dynamic introduction of features. In the second simulation we will demonstrate the application of the multi-feature MFT on data related to the classification of actions from interactive robotics study. Finally, in the third simulation we will consider the scaling up of the MFT to complex data sets.

To facilitate the presentation of the results, we will interpret both the object feature values and the modeling fields as d -dimensional vectors and follow the time evolution of the corresponding vector length

$$S_k = \sqrt{\sum_{e=1}^d (S_{ke})^2} / d, \quad (7)$$

which should then match the object length $O_i = \sqrt{\sum_{e=1}^d (O_{ie})^2} / d$.

Simulation I: Incremental addition of feature

Consider the case in which we have the 5 objects, initially with only one-feature information. For instance, we can consider color information only on Red, the first of the 3 RGB feature values, as used in Steels’s [19] discrimination-tree implementation. The objects have the following R feature values: $O_1 = [0.1]$, $O_2 = [0.2]$, $O_3 = [0.3]$, $O_4 = [0.5]$, $O_5 = [0.5]$.

A first look at the data indicates that these 5 input stimuli belong to four color categories (concept-models) with Red values respectively 0.1, 0.2, 0.3 and 0.5. As a matter of fact, the application of the MFT algorithm to the above mono-dimensional input objects reveal the formation of 4 model fields, even when we start with the condition in which 5 fields are randomly initialized (Fig. 1).

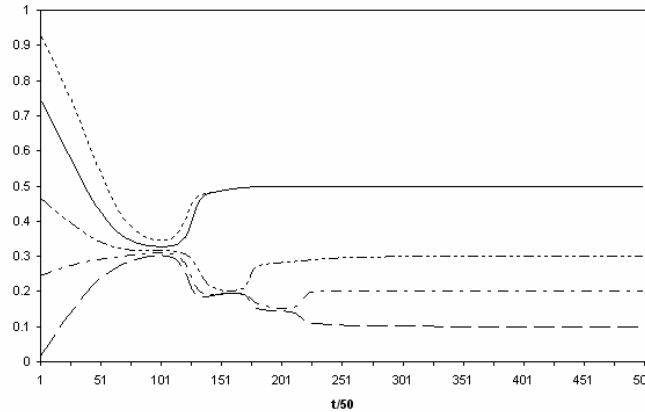


Fig. 1 – Time evolution of the fields with only the first feature being used as input. Only 4 models are found, with two initial random fields converging towards the same .5 Red concept-model value.

Let us now consider the case in which we add information from the second color sensor, Green. The object input data will now look like these: $O_1 = [0.1, 0.4]$, $O_2 = [0.2, 0.5]$, $O_3 = [0.3, 0.2]$, $O_4 = [0.5, 0.3]$, $O_5 = [0.5, 0.1]$.

The same MFT algorithm is applied with 5 initial random fields. For the first 12500 training cycles (half of the previous training time), only the first feature is utilized. At timestep 12500, both features are considered when computing the fuzzy similarities. From timestep 12500, the dynamics of the σ_2 fuzziness value is initialized, following equation (7), whilst σ_1 continues¹ its decrease pattern started at timestep 0. Results in Fig. 2 show that the model is now able to correctly identify 5 different fields, one per combined RG color type.

¹ We have also experimented with the alternative method of re-initializing both σ_g values, as in equation (7), whenever a new feature is added. This method produces similar results.

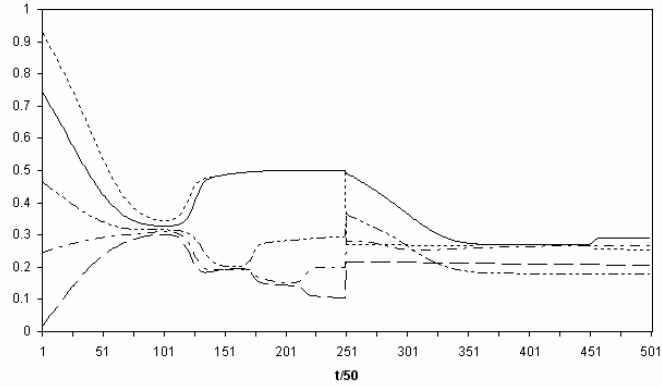


Fig. 2 – Time evolution of the fields when the second feature is added at timestep 12500. The dynamic fuzziness reduction for σ_2 starts at the moment the 2nd feature is introduced, and is independent from σ_1 . Note the restructuring of 4 fields initially found up to timestep 12500, and the further discovery of the model. The fields values in the first 12500 cycles is the actual mono-dimensional field value, whilst from timestep 12500 the equation in (7) is used to plot the combined fields' value.

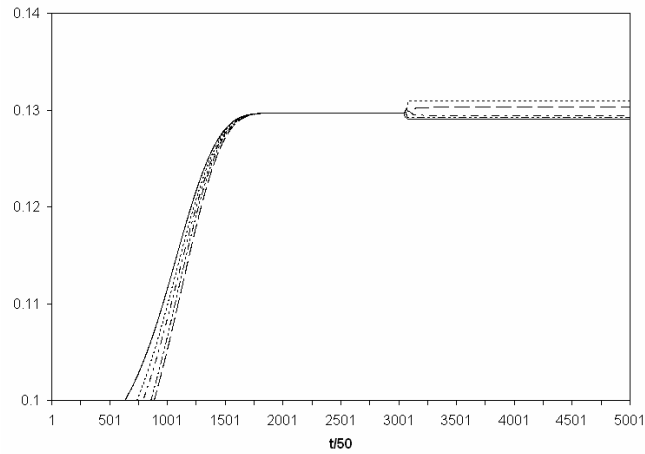


Fig. 3 – Evolution of fields in the robot posture classification task. The value of the field corresponds to equation (7). Although the five fields look very close, in reality the individual field values match very well the 42 parameters of the original positions.

Simulation II: Categorization of robotic actions

In the introduction we have proposed the use of MFT for modeling the integration of language and cognition in cognitive robotic studies. This is a domain where the input to the cognitive agent (e.g. visual and auditory input) typically consists of multi-

dimensional data such as images of objects/robots and speech signals. Here we apply the multi-dimensional MFT algorithm to the data on the classification of the posture of robots, as in an imitation task. We use data from a cognitive robotic model of symbol grounding [4,6]. We have collected data on the posture of robots using 42 features. This consist of the 7 main data (X, Y, Z, and rotations of joints 1, 2, 3, and 4) for each of the 6 segments of the robot's arms (right shoulder, right upperarm, right elbow, left shoulder, left upperarm, left elbow). As training set we consider 5 postures: resting position with both arms open, left arm in front, right arm in front, both arms in front, and both arms down. In this simulation, all 42 features are present from timestep 0. Fig. 3 reports the evolution of fields and the successful identification of the 5 postures.

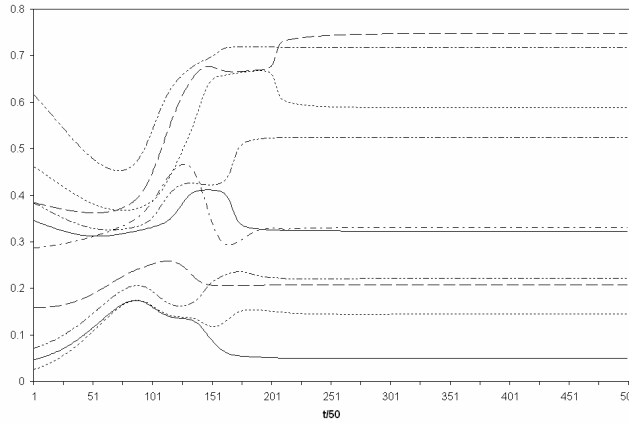


Fig. 4 – Evolution of fields in the case with 1000 input objects and 10 prototypes.

Simulation III: Scaling up with complex stimuli sets

Finally, we have tested the scaling-up of the multi-dimensional MFT algorithm with a complex categorization data set. The training environment is composed of 1000 objects belonging to the following 10 2-feature object prototypes: [0.1, 0.8], [0.2, 1.0], [0.3, 0.1], [0.4, 0.5], [0.5, 0.2], [0.6, 0.3], [0.7, 0.4], [0.8, 0.9], [0.9, 0.6] and [1.0, 0.7]. For each prototype, we generated 100 objects using a Gaussian distribution with standard deviation of 0.05. During training, we used 10 initial random fields.

Fig. 4 reports the time evolution of the 10 concept-models fields. The analysis of results also shows the successful identification of the 10 prototype models and the matching between the 100 stimuli generated by each object and the final values of the fields.

Discussion and Conclusion

In this paper we have presented an extension of the MFT algorithm for the classification of objects. In particular we have focused on the introduction of multi-dimensional features for the representation of objects. The various simulations showed that (i) the system is able to dynamically adapt when an additional feature is introduced during learning, (ii) that this algorithm can be applied to the classification of action patterns in the context of cognitive robotics and (iii) that it is able to classify multi-feature objects from complex stimulus set.

Our main interest in the adaptation of MFT to multi-dimensional objects is for its use in the integration of cognitive and linguistic abilities in cognitive robotics. MFT permits the easy integration of low-level models and objects to form higher-order concepts. This is the case of language, which is characterized by the hierarchical organization of underlying cognitive models. For example, the acquisition of the concept of “word” in a robot consists in the creation of a higher-order model that combines a semantic representation of an object model (e.g. prototype) and the phonetic representation of its lexical entry [15]. The grounding of language into categorical representation constitutes a cognitively-plausible approach to the symbol grounding problem [11]. In addition, MFT permits us to deal with the problem of combinatorial complexity, typical of models dealing with symbolic and linguistic representation. Current cognitive robotics model of language typically deal with few tens or hundred of words (e.g. [6,19]). With the integration of MFT and robotics experiments we hope to deal satisfactory with the combinatorial complexity problem.

Ongoing research is investigating the use of MFT for the acquisition of language in cognitive robotics. In particular we are currently looking at the use of multi-dimensional MFT to study the emergence of shared languages in a population of robots. Agents first develop an ability to categorize objects and actions by building concept-models of objects prototypes. Subsequently, they start to learn a lexicon to describe these objects/actions through a process of cultural learning. This is based on the acquisition of a higher-order MFT.

Acknowledgements

Effort sponsored by the Air Force Office of Scientific Research, Air Force Material Command, USAF, under grants number FA8655-05-1-3060 and FA8655-05-1-3031. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purpose notwithstanding any copyright notation thereon.

References

- [1] Barsalou, L. (1999), Perceptual symbol systems. *Behavioral and Brain Sciences*, 22, 577-609.
- [2] Cangelosi A. (2001). Evolution of communication and language using signals, symbols and words. *IEEE Transactions on Evolutionary Computation*. 5(2), 93-101

- [3] Cangelosi A., Bugmann G. & Borisyuk R. (Eds.) (2005). *Modeling Language, Cognition and Action: Proceedings of the 9th Neural Computation and Psychology Workshop*. Singapore: World Scientific.
- [4] Cangelosi A., Hourdakis E. & Tikhanoff V. (2006). Language acquisition and symbol grounding transfer with neural networks and cognitive robots. *Proceedings of IJCNN2006: 2006 International Joint Conference on Neural Networks*. Vancouver, July 2006.
- [5] Cangelosi, A., & Parisi, D. (2004). The processing of verbs and nouns in neural networks: Insights from synthetic brain imaging. *Brain and Language*, 89(2), 401-408.
- [6] Cangelosi A, Riga T (2006). An Embodied Model for Sensorimotor Grounding and Grounding Transfer: Experiments with Epigenetic Robots, *Cognitive Science*, 30(4), 1-17.
- [7] Fontanari J.F., Perlovsky L.I. (2005). Meaning creation and modeling field theory. In C. Thompson & H. Hexmoor (Eds.), *IEEE KIMAS2005: International Conference on Integration of Knowledge Intensive Multi-Agent Systems*. IEEE Press, pp. 405-410.
- [8] Fontanari J.F., Perlovsky L.I. (2006a). Categorization and symbol grounding in a complex environment. *Proceedings of IJCNN2006: 2006 International Joint Conference on Neural Networks*. Vancouver, July 2006.
- [9] Fontanari J.F., Perlovsky L.I. (2006b). Meaning creation and communication in a community of agents. *Proceedings of IJCNN2006: 2006 International Joint Conference on Neural Networks*. Vancouver, July 2006.
- [10] Glenberg A., & Kaschak, M. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, 9(3), 558-565.
- [11] Harnad, S. (1990). The symbol grounding problem. *Physica D*, 42, 335-346.
- [12] Marocco, D., Cangelosi, A., & Nolfi, S. (2003). The emergence of communication in evolutionary robots. *Philosophical Transactions of the Royal Society of London – A* 361, 2397-2421.
- [13] Pecher, D., & Zwaan, R.A., (Eds.). (2005). *Grounding cognition: The role of perception and action in memory, language, and thinking*. Cambridge: Cambridge University Press.
- [14] Perlovsky L. *Neural Networks and Intellect: Using Model-Based Concepts*. Oxford University Press, New York, 2001.
- [15] Perlovsky L., "Integrating language and cognition," *IEEE Connections*, vol. 2, pp. 8-13, 2004
- [16] Pulvermuller F. (2003) *The neuroscience of language. On brain circuits of words and serial order*. Cambridge: Cambridge University Press.
- [17] Rizzolatti, G., & Arbib, M. (1998). Language within our grasp. *Trends in Neuroscience*, 21: 188-194.
- [18] Smith A. D. M. (2003). Semantic generalization and the inference of meaning," In: W. Banzhaf, T. Christaller, P. Dittrich, J. T. Kim, J. Ziegler (Eds.), *Proceedings of the 7th European Conference on Artificial Life*, Lecture Notes in Artificial Intelligence, vol. 2801, pp. 499-506.
- [19] Steels L (1999) *The talking heads experiment (Volume I. Words and meanings)*. Antwerpen: Laboratorium.
- [20] Steels, L. (2003) Evolving grounded communication for robots. *Trends in Cognitive Sciences*, 7(7):308-312.
- [21] Wermter, S., Elshaw, M., and Farrand, S., 2003, A modular approach to self-organization of robot control based on language instruction. *Connection Science*, 15(2-3): 73-94.