Living with Dynamic Concepts in Dynamic Environments
Intelligent Agents that Adapt Themselves

Matthias Rehm

Multimedia Concepts and their Applications
Institute of Computer Science, University of Augsburg
Eichleitnerstr. 30, 86159 Augsburg, Germany
rehm@multimedia-werkstatt.org

Abstract

In this paper a dynamic perspective on concepts for intelligent agents is propagated. The discussion on the nature of concepts can be simplified to a dichotomy between objective/static and subjective/dynamic approaches. In the objective case, concepts are the same for all individuals. Under a dynamic perspective, concepts depend on different factors like the learning process, the environment, i.e. the situational setting. It is indispensable for an agent to create individual concepts that adhere to restrictions imposed by the environment and the society it is living in. It is shown that changes in the environment lead to changes in existing concepts and to establishing new ones with only a small irritation in the use of the old ones.

1 Introduction

If intelligent agents interact with a complex environment, than this environment will be dynamic as well. Handling dynamic environments poses the challenge of a suitable internal representation to such an agent. In this article, a system is presented which makes use of mechanisms of multi-modal concept formation that take dynamic cognitive structurings into account. The agents sensory equipment consists of a visual sensor supplying the agent with a 3D image of its surrounding and a natural language interface, which is used by the user to supply the agent with descriptions of the current scene. Dynamic concepts can be found on different levels in the system LOCATOR. i.) In a single agent, concepts vary over time, becoming more and more specific and consistent with the history of experience of the agent. ii.) Because different agents have different histories of experience, their individual concepts may differ to a certain degree, even if they were confronted with the same type of input. iii.) One crucial aspect of the agent's environment is the language used in its peer group. Frames of spatial reference ([5]; [13]) are chosen as the domain for the simulations in LOCATOR. Cross-linguistic findings in this domain show a great variance between different languages. This dynamics in the language use can be found between agents of different language groups ([10]). Currently, German and Marquesan[1] are examined. iv.) The most interesting form of dynamic concepts is described in more detail in this paper: adapting existing concepts and creating new ones due to radical changes in the environment.

2 Related Work

Sun ([16]) proposes an approach to the question of concept formation that is in line with the work described in this paper. Concepts are formed in the context of the agent's interactions with the world. The agent starts with minimum built-in structures that develop according to the specific experiences of a single agent. To implement this kind of concept formation, Sun developed a hybrid computational model (CLARION) consisting of different levels that handle learning of comportment on the one hand and rule learning on the other hand, both in a bottom-up fashion.
A more theoretical approach is described by Madole and Oakes ([7]) who propagate a task- and context-oriented process of concept formation. In a number of experiments, Schyns and Rodet ([12]) show that subjects create the necessary discriminative features to categorize new input on the basis of their previous knowledge and context-specific experience with the encountered objects.
Regier ([9]) proposes an abstract connectionist model that combines some kind of perceptual and conceptual input to learn the concepts of spatial prepositions in different languages like Russian, English, or Mixtec. While learning

---

[1]Marquesan is a Polynesian language. Speakers employ a fixed directed reference axis tai — uta (sea — inland) and an orthogonal undirected axis ko (across). For further details on Marquesan see [2].
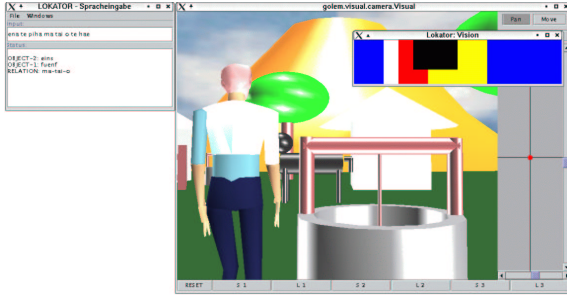
Figure 1. The Simulation system: An Agent explores its 3D-environment. The small window (Lokator: Vision) inside the main gives an impression of the agent's visual input (depth information missing). On the left, the NL interface is shown, where the user provides linguistic input to the agent.

takes place, necessary features are created and modified. Problematic is the use of the conceptual or "linguistic" input. The output nodes of the model are labeled with prepositions of the specific language under consideration. This interpretation of the output nodes as linguistic symbols is very artificial and not inherent to the model itself.

Cangelosi and Harnad ([3]) focus on the role of language during the acquisition of new perceptual categories. In their simulations they use two types of agents (*foragers* and *thieves*) that inhabit a 2D environment where a number of mushrooms can be found. Some mushrooms supply the agents with energy, some of them are poisonous. The agents have to learn the features of the mushrooms. After having established basic categories by being confronted with the actual mushrooms and their effects, the foragers learn additional categories in the same way. The thieves learn new categories solely on the basis of linguistic communication. The categories do not change during the simulations and are identifiable by fixed sets of features. Thus, although the agents form multi-modal concepts, these concepts are not dynamic.

3 LOCATOR — A model of situated concept formation

The simulation system LOCATOR[2] implements a model of situated concept formation. Autonomous anthropomorphic agents explore their complex 3D-environment (see Fig. 1). The idea of autonomous agents follows Franklin and Graesser's (1996) definition ([4]). The agents perceive this environment with a visual sensor and they have a natural

language interface, which is used by the user to describe the spatial arrangement of currently perceivable objects, e.g., *Der Baum steht rechts vom Haus* (The tree is right of the house).

Modeling the process of concept formation makes use of ideas by Steels (e.g., [14]), who has proposed discrimination games to model the process of situated acquisition of object concepts. While agents play discrimination games they build up discrimination nets for the sensor channels they possess, depending on the experienced input (see Fig. 3 for some nets). Every node in such a net is a feature detector. A detector can get activated in a specific situation by the incoming information and is passed as input to further processing steps. In Locator, feature detectors are not operating on the values of sensor channels, but on five preprocessed perceptual features. Two of these describe the distance (ABS) and the alignment (AUS) between two objects. The other three supply information concerning the three main axis of spatial cognition, i.e. the horizontal (H), the vertical (V), and the 2. sagittal (2S). First, the vector between the center of mass of two objects is calculated (COM). The perceptual features are the angles between each reference axis and this vector. For each perceptual feature a corresponding initial feature detector exists. This initial detector ranges over the whole value range of the corresponding feature. Consequently, it always gets activated when a value for this feature is encountered. During the process of concept formation the initial feature detectors are elaborated, resulting in a number of different discrimination nets, one for each perceptual feature (see Fig. 3). Each node is itself a feature detector that corresponds to parts of the value range of the given perceptual feature and is activated if a value falls into this range. The number and kind of feature detectors employed by a specific agent emerges during the conceptualization process and is due to the experienced multi-modal input, i.e. to the visual as well as to the linguistic input.

The pressure to modify an existing discrimination net results in Steels' approach from the assumption, that no identical objects exist. Thus, a distinctive set of features can be determined for every object. Discrimination nets are build up only on the information supplied by the available sensors without taking linguistic input into account. In a later version ([15]), Steels' agents play so-called *language games*. But these games are independent processes mapping words on concepts and have no impact on the concept formation process. In LOCATOR the linguistic input is an integral part of this process, following an usage-based approach to language acquisition (e.g., [1]). The linguistic input is seen as a deliberately produced utterance by a member of the language group the agent is part of. This utterance realizes a generally accepted way of structuring the spatial domain. By means of the utterance, this way of structuring is given as a positive example to the learner. Consequently, LOCA-

---

[2]LOCATOR is based on the system LOKUTOR which was developed by Milde (e.g., [8]) at Bielefeld University.

TOR utilizes the information inherent in this positive example for the concept formation process, i.e., for building up discrimination nets (see [10]).

4 Simulations and Results

LOCATOR is a testbed for questions concerning the acquisition of concepts like the interplay between different modalities (e.g., vision and speech) or the acquisition of different frames of reference (see [10]). Focusing on dynamic concepts, three simulations are described in this article. In the first two, the environment is not subject to changes during acquisition and use of concepts. The third simulation focuses on the dynamic nature of concepts. It replicates the first simulation and then adds environmental changes that lead to the same conditions as in simulation two. The general setting of the simulations is identical. An agent autonomously explores its environment, i.e. it follows a random path through its environment based on local behaviors to avoid collisions with objects. From time to time the user describes the spatial arrangement, which the agent is able to perceive with its visual sensor. The user input triggers a categorization attempt and, if this fails, a learning step. The success of this categorization attempt is measured. The visual and linguistic input activate concepts that represent the joint meaning of the different types of input. If a single concept is activated, the categorization attempt is successful.

4.1 Simulation 1: rechts vs links

During their exploration, five agents were confronted with utterances that employ a relative frame of reference. The relations[3] were *links* and *rechts*. Each agent experienced 1600 utterances. The number of specific relations each agent was confronted with varies between 790 (49.38%) and 808 (50.5%) for *links* and between 792 (49.5%) and 810 (50.63%) for *rechts*.

4.1.1 Results

Because the number of trials per relation varies across agents, an analysis of covariance is necessary. For this analysis the average performance over all agents is calculated at each reading point. If an input can not be categorized non-ambiguous, then this attempt is not interpreted as a successful categorization attempt. Because two different phases can be distinguished in regard to performance, the analysis is repeated with the last half of the reading points.

---

[3]To distinguish between *relations* and CONCEPTS they are given in two different fonts. Because linguistic input was given in German, the German words for relations and concepts are employed in this article: rechts (right), links (left), vor (front), hinter (back).

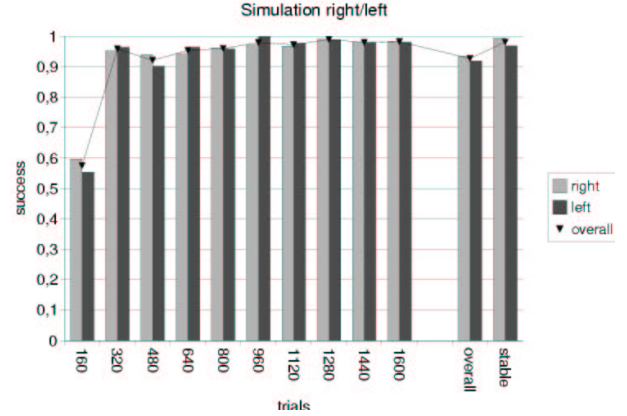|	Trials 1–1600		Trials 801–1600	
	df	F	df	F
Trials	9,79	27.02**	4,39	0.77
Relations	1,79	0.82	1,39	0.02
T x R	9,79	0.42	4,39	0.09

$p < 0.05^*$, $p < 0.01^{**}$

Figure 2. Simulation 1: Categorization performance during trials.

The results can be found in figure 2 which also depicts the categorization performance of the system Locator. A significant effect shows up between different trials if all reading points are taken into account: $F(9,79) = 27.02$, $p < 0.01$. The effect vanishes if the analysis is limited to the stable phase.

4.1.2 Discussion

Categorization performance is low at the beginning. By and by, i.e. when concepts become more stable, this is reversed. Consequently, a significant effect can be found if all reading points are considered. This is due to the fact, that an agent starts without any concept at all. Thus the performance rate is zero at the beginning. Over time, the agent modifies its perceptual system to create suitable features for concept formation. As a result, concepts become better suited for the task at hand, i.e. categorizing relations that hold between objects. Figure 2 shows that the agents have established "working" concepts, i.e. concepts that are suited for the categorization task. The performance rate varies between 95% and 100% after concepts have stabilized (> 800 trials).

Of special interest is the question, how the agents have modified their initial perceptual system, i.e., their initial feature detectorse due to the experienced input. The perceptual sys-
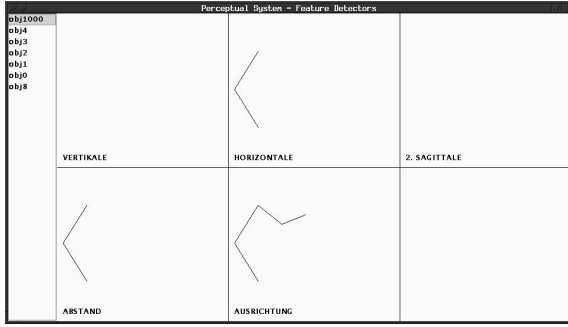
Figure 3. Simulation 1: The perceptual system of one agent.

tem of one agent after 1600 trials is depicted in figure 3.[4] The five feature detectors have been modified to analyze the perceptual features appropriately. In the first row, the feature detectors which analyze the angle between COM and a reference direction are shown, i.e. from left to right V, H, and 2S. The second row depicts the detectors responsible for the features distance and alignment.

All agents of this group have modified their respective perceptual system in slightly different ways. The modifications depend on the experienced situations that vary between agents. A more thorough analysis of the established concepts shows that one feature detector is used primarily for concept formation. This detector evaluates the angle between COM and the horizontal. Thus it corresponds to one of the main axis of human spatial cognition, which is an expected result.

4.2 Simulation 2: rechts, links, vor, hinter

In this simulation, each agent got 3200 utterances during its exploration. Roughly a quarter of these utterances realized one of the four relations *links*, *rechts*, *vor*, and *hinter*. The exact numbers are found in a range between 782 (24.44%) and 842 (26.31%) (*rechts*), 758 (23.69%) and 815 (25.47%) (*links*), 773 (24.16%) and 837 (26.16%) (*vor*), and 761 (23.78%) and 827 (25.84%) (*hinter*).

4.2.1 Results

The results are more or less the same as in simulation one. A significant effect shows up between trials if all reading points are considered which vanishes if the analyses is constrained to the stable phase. Additional effects show up this

---

[4]To test the acquisition of different frames of spatial reference [6], it is necessary to know the origin of the coordinate system in which features are calculated. In all simulations described in this paper, the origin is *obj1000* (the agent itself), This is indicated in every figure on the left side. See, e.g., [11] for simulations concerning absolute frames of reference.

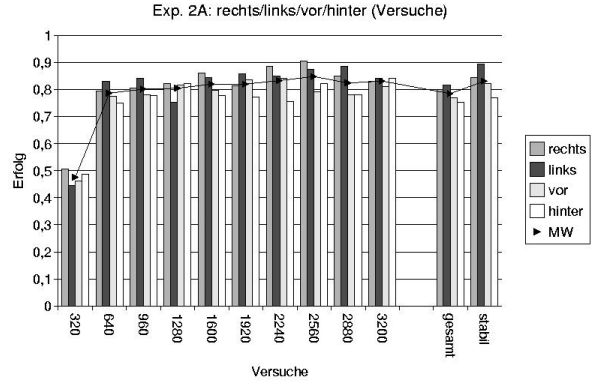|	Trials 1–1601		Trials 1601–3200	
	df	F	df	F
Trials	9,159	45.9**	4,79	0.65
Relations	3,159	6.93**	3,79	8.22**
V x R	27,159	0.89	12,79	1.49

$p < 0.05^*, p < 0.01^{**}$

Figure 4. Simulation 2: Categorization performance during trials.

time between relations that do not vanish in the stable phase: $F(3, 159) = 6.93^{**}, p < 0.01$ and $F(3, 79) = 8.22^{**}, p < 0.01$. Figure 4 shows that performance is on the average better for the relations *rechts* and *links* (0.87, on the average, > 800) than for *vor* and *hinter* (0.79, on the average, > 800).

4.2.2 Discussion

The agents were successful in establishing concepts. The performance is not as high as in the first simulation and there is a significant effect between relations that persists over all trials. A thorough analysis of the situations experienced by the agents reveals that this effect is attributable to the low depth resolution of the visual sensor.

The visual sensor consists of different sensor layers that register objects in different ranges from the agent. Sensor layers are combined into a single represenation, starting with those that have a longer range. In the foveal area of the sensor, resolution of these layers is high but their range is restricted. In the the parafoveal area the resolution is lower but the range is higher. In both areas, five layers exist that correspond to five different depth levels (see Fig. 6). Because they overlap to some extent, there are eight depth levels in all. This low depth resolution especially affects the discrimination between the relations *vor* and *hinter*. If two objects are in a *vor-/hinter*-relation and near to each other, it is very likely that the agent perceives them as being
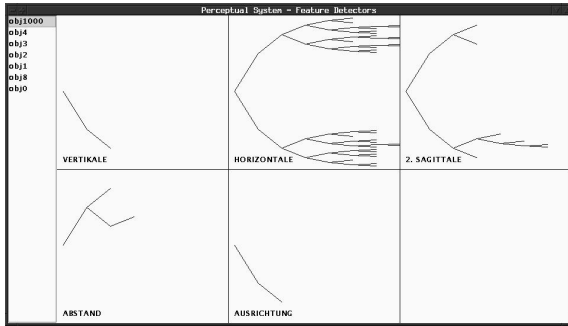
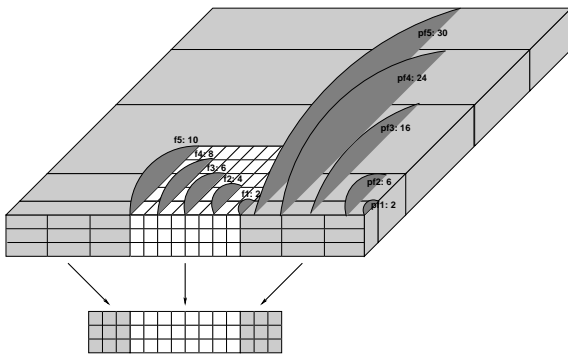**Figure 5. Simulation 2: Perceptual system of one agent.**

Figure 6. Visual sensor of the Locator agents. The white area marks the foveal part of the sensor. The parafoveal area is given in grey. Pf2:16 denotes, e.g., parafoveal sensor layer 2, which has a range of 16.

Figure 7. Simulation 3: Development of categorization performance of one agent during trials 1600–3200.

equally far away. To successfully apply such a relation it is inevitable that this is not the case. Such situations are called ambiguous in this context. Roughly 15 % of the encountered situations were ambiguous in this sense (*vor*: 15.68 %; *hinter*: 15.11 %).

In contrast to the first experiment, the perceptual systems of the agents are more elaborated (Fig. 5). The analysis of the established concepts indicate that the modifications of the perceptual systems in simulation one are not sufficient to solve the more complex learning problem in this simulation. On the one hand, it is necessary to differentiate the initial feature detectors further. On the other hand, it is not sufficient to concentrate on a single feature for the task at hand. Instead, a combination of relevant features is needed. Here, it is a combination of the features H and 2S. H was primarily used in simulation one to establish the concept LINKS and RECHTS. In a further simulation which is not reported here, it was shown that 2S is primarily used in establishing the concepts VOR and HINTER. This clear
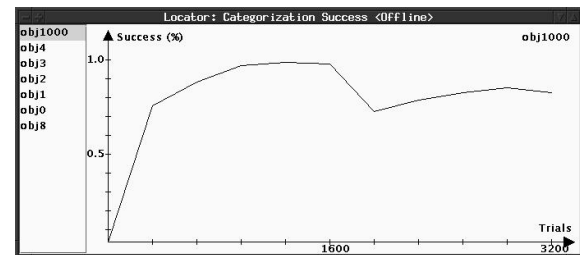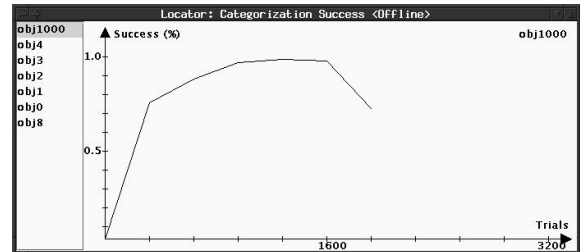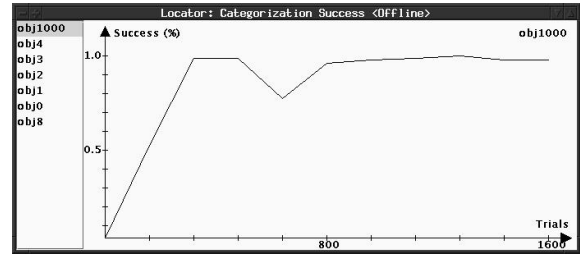
assignment between concepts and features does not persist in this simulation. Both features constitute all of the concepts while different value ranges emerge for the different concepts.

Comparing the concepts from simulation one and two, the same concepts are structurally different. These differences arise from the experienced situations, i.e., dependent on the visual and linguistic input. If the *links-/rechts*-dichotomy has to be learned — like in simulation one — the feature 2S contains no relevant information. Contrasting the relations *vor* and *hinter* — like in simulation two — this feature does not only become relevant to establish the concepts VOR and HINTER but also to establish the concepts LINKS and RECHTS. The content of a concepts can thus not be restricted to a so called *conceptual core*. The structure and the content of a concept can only be determined depending on the specific task and context the concept is used in. The task in simulation two is more complex due to a larger number of concepts that have to be acquired. Consequently, more aspects of the encountered situations become relevant.
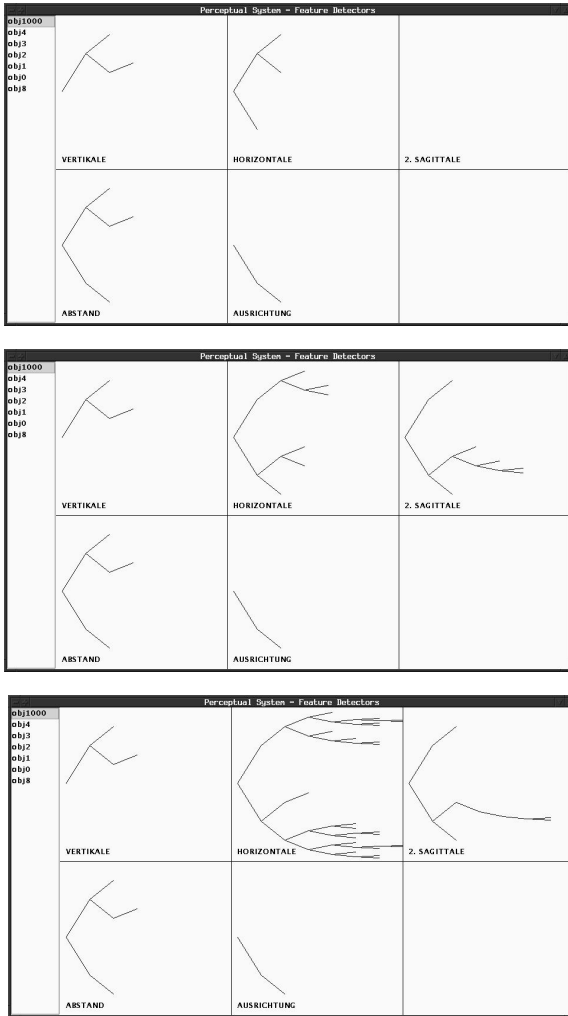
Figure 8. Simulation 3: Development of the perceptual system of one agent during trials 1600–3200.

4.3 Simulation 3: Dynamic concept formation

The first two simulations showed that LOCATOR implements a successful model of situated and individual multimodal concept formation which is a pre-requisite for the claim of dynamic concepts. Simulation three examines this claim in more detail.

A number of agents was confronted with linguistic input during their exploration of the environment. This time, the linguistic input was changed after concepts had stabilized. In the first part of this experiment, simulation one was replicated, i.e., the agents got 1600 utterances that employed the relations *rechts* and *links*. Afterwards, 4800 utterances were presented that used *vor* and *hinter* as additional input. Thus, the agents environment was subject to a radical change be-

cause totally new structurings of the visually perceivable reality are introduced via the linguistic input.

In the first part, the results from the first simulation were replicated. The same holds true for the stable phase of the second part (from 3200 trials onwards) concerning the results of the second simulation. The interesting things happen in between, i.e. from the beginning of the use of the new relations until concepts have stabilized again (from 1600 — 3200 trials).

Figures 7 and 8 show the development of the categorization performance and of the perceptual system of one agent as an example. The use of new and unknown relations in the linguistic input causes at first a breakdown in categorization performance. For this agent, performance drops from 95% to 70%. During the next 1600 trials, performance increases steadily due to a modification of the perceptual system and stabilizes at the level that is known from simulation two. The modification concentrates on two features. On the one hand, the feature H is made more specific. It was already used in the first part of the simulation. On the other hand, the feature 2S is used additionally. This feature was not significant in the first part of the simulation. Consequently, modifying the perceptual system involves a restructuring of the already established concepts. The features that have constituted the concepts LINKS and RECHTS so far, can no longer guarantee a successful categorization. Other features become necessary, too. In the course of this restructuring the categorization performance for the concepts LINKS and RECHTS decreases at first before increasing on the performance level known from simulation two. The overall decrease is thus not solely attributable to the unknown relations for which new concepts have to be established.

Categorization performance should not be affected for the already existing concepts (at least after restructuring is complete). Unfortunately, this cannot be fully guaranteed in LOCATOR because of the limitation of the available sensory equipment (see simulation two). Thus, performance stabilizes on a level known from simulation two. This is a direct result of a situated and embodied approach to the problem of concept formation and is thus an advantage and not a fault of the system.

A difference in performance during the acquisition of the concepts VOR and HINTER was found comparing simulations two and three. The unstable phase at the beginning, that was registered in all simulation done so far, is missing here. Instead, performance stabilizes very rapidly on a high level. The only difference between the first two simulations and this one is found in a segmentation of the learning problem. Acquiring the concept LINKS and RECHTS caused the agents to modify their perceptual systems, i.e., their mechanisms to analyze the visual input, in a way that is compatible with the experienced input. The linguistic input trades the generally accepted conceptual constraints of the agent's lan-

guage group, in this case in the domain of spatial references in a relative system. Consequently, the agent's perceptual system is already pre-structured for analyses in such a relative frame of reference. Acquiring further concepts that are in accordance with the needs of such a system can be based on this already established knowledge. Thus, forming new concepts is alleviated because they are compatible with the already established conceptual structuring.

5 Conclusion

Intelligent agents are used in more and more complex environments. Complex environments are also always inherently dynamic. This poses new challenges to the representational abilities of agents. LOCATOR is a system that realizes a process of multi-modal concept formation which takes the inherently dynamic nature of the environment into account. This is inevitable to be able to react to changes in the environment or to previously unsupposed conditions. It was shown that the agents in LOCATOR react to such radical changes by adapting their existing concepts and if necessary by creating new ones. No additional mechanisms are needed to achieve this result. Instead, in the case of an environmental change, existing concepts are just re-entered into the concept formation process. An interesting result was the finding that a certain pre-structuring allows for a more rapid learning of new concepts that follow the same structuring.

Further research will be conducted mainly in one direction. In the context of dynamic concepts it is very interesting to see which effects will arise if an agent is forced to acquire a second language, that structures the spatial domain in a different way. Such an agent will e.g., first learn the German spatial concepts and afterwards will be confronted with linguistic input in Marquesan. Such an agent will have to be contrasted to a bilingual one, that is exposed to both languages from the beginning.

6 Acknowledgments

References

[1] M. Bowerman and S. Choi. Shaping meanings for language: universal and language-specific in the acquisition of spatial semantic categories. In M. Bowerman and S. C. Levinson, editors, *Language acquisition and conceptual development*, pages 475–511. Cambridge University Press, Cambridge, 2001.

[2] G. H. Cablitz. The Acquisition of an Absolute System: Learning to talk about SPACE in Marquesan. In *Proceedings of the 31st Stanford Child Language Research Forum: Space in Language — Location, Motion, Path, and Manner*, pages 40–49, 2002.

[3] A. Cangelosi and S. Harnad. The adaptive advantage of symbolic theft over sensorimotor toil: grounding language in perceptual categories. *Evolution of Communication (Special Issue on Grounding Language)*, 2000.

[4] S. Franklin and A. Graesser. Is It an Agent, or Just a Program?: A Taxonomy for Autonomous Agents. In J. P. Müller, M. J. Woolridge, and N. R. Jennings, editors, *Intelligent Agents III. Agent Theories, Architectures, and Languages.*, pages 21–35. Springer, 1996.

[5] S. C. Levinson. From outer to inner space: linguistic categories and non-linguistic thinking. In J. Nuyts and E. Pederson, editors, *Language and conceptualization*, pages 13–45. Cambridge University Press, 1997.

[6] S. C. Levinson. Covariation between spatial language and cognition, and its implications for language learning. In M. Bowerman and S. C. Levinson, editors, *Language acquisition and conceptual development*, pages 566–588. Cambridge University Press, Cambridge, 2001.

[7] K. L. Madole and L. M. Oakes. Making Sense of Infant Categorization: Stable Processes and Changing Representations. *Developmental Review*, 19:263–296, 1999.

[8] J.-T. Milde. Lokutor: Towards a Believable Communicative Agent. In J. Rickel, W. L. Johnson, and J. Lester, editors, *Achieving Human-Like Behavior in Interactive Animated Agents*. Fourth International Conference on Autonomous Agents, 2000.

[9] T. Regier. *The Human Semantic Potential - Spatial Language and Constrained Connectionism*. MIT Press, Cambridge, Mass., 1996.

[10] M. Rehm. Language guiding concept formation in artificial agents. In A. Holmer, J. O. Svantesson, and A. Viberg, editors, *Proceedings of the 18th Scandinavian Conference of Linguistics. Travaux de l'Institut de Linguistique de Lund*, pages 241–253. 2001.

[11] M. Rehm. *Lokator — Multimodale Bedeutungskonstitution in situierten Agenten*. Universität Bielefeld, Online publication, URL: http://archiv.ub.uni-bielefeld.de/disshabi/2001/0076, 2001.

[12] P. G. Schyns and L. Rodet. Categorization Creates Functional Features. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(3):681–696, 1997.

[13] G. Senft. Introduction. In G. Senft, editor, *Referring to Space - Studies in Austronesian and Papuan Languages*, pages 1–38. Claredon Press, Oxford, 1997.

[14] L. Steels. Perceptually grounded meaning creation. In M. Tokoro, editor, *Proceedings of the International Conference on Multi-Agent Systems*, pages 338–344. AAAI Press, 1996.

[15] L. Steels and F. Kaplan. Bootstrapping Grounded Word Semantics. In T. Briscoe, editor, *Linguistic evolution through language acquisition: formal and computational models*. Cambridge University Press, 1999.

[16] R. Sun. Symbol Grounding: A New Look At An Old Idea. *Philosophical Psychology*, 13(2):149–172, 2000.