Categorization by Sensory-Motor Interaction in Artificial Agents

Martin Takáč (takac@ii.fmph.uniba.sk)
Dept. of Applied Informatics FMFI, Comenius University
Mlynská dolina, 842 48 Bratislava, Slovakia

Abstract

In this paper we propose a cognitively relevant computational model of categorization, grounding the categories in sensory-motor interaction with a dynamical environment. Simple perceptual categories are represented as discrimination criteria – membership functions based on distributional information about intra-cluster variances of properties of a category, the representation more effective for predictions than prototypes. Complex categories are represented as cross-categorial associations of criteria of objects, actions and changes, hence, they support action-based inferences and can serve as grounded meanings not only for nouns and adjectives, but also for at least elementary verbs in models of language evolution and acquisition.

Introduction

How is it that we speak? How does a child acquire a language and how has the ability to communicate in human languages appeared in the history of mankind? These puzzling questions have been explored and discussed for more than hundred years. In the tradition of cognitive semantics, the language acquisition has been to a large extent viewed as a problem of acquiring correct mappings between elements of overt form, such as words, sentences, gestures etc. and covert meanings (Langacker, 1991). Meanings, as mental concepts, are taken to be inborn (Fodor, 1981), acquired in the course of interaction with the world (Bloom, 2000), or formed solely by the influence of language itself (Whorf, 1956). If we took an extreme view that all concepts exist in advance, language acquisition would be just a labeling problem – learning the names for existing concepts. If we took the opposite extreme view, no thinking could exist without a language. While in our opinion some concepts are inborn and some are formed or reformed by the influence of the language, in this paper we want to emphasize and explore the process of forming concepts by sensory-motor interaction with the environment. Concepts, in this view, are non-verbal embodied structures supporting the elementary ability to discriminate between objects, situations, states, etc., helping to survive in a real world and providing grounding for the language. Physically they correspond to activations of certain neural structures correlated with perceiving, imagining or talking about the content they represent (Barsalou, 1999).

Empirical findings suggest that infants have rich mental life well before they start to speak, they learn various skills and motor schemata (Piaget & Inhelder, 1966) and learn to discriminate between various objects and situations (Spelke, 1990). Also, many species in animal kingdom possess abilities to perceive and categorize scenes, and the language is a very late invention on an evolutionary time-scale. This proves that at least some mental concepts can exist independently of the language. We believe that studying these prelinguistic structures and their formation is a necessary, although not a sufficient, condition for understanding the intricate problem of our language.

In recent years, a number of computational models of language evolution and acquisition have been presented, e.g. (Kirby & Hurford, 2001; Steels, 2000). However, meanings used in these models are either artificial, predefined symbolic structures, such as the predicate *loves(john,mary)* lacking any grounding in the environment (Harnad, 1990), or they represent only static objects on the scene, thus constituting meanings only for nouns and adjectives, while meanings for verbs are completely absent.

The first words of a child label not only things, but also relations, actions and internal states (Tomasello & Merriman, 1995) and the child's first more complex lexical constructions are organized around verbs – *verb island hypothesis* (Tomasello, 1992). Although verbs are more difficult to acquire and understand correctly than nouns (Gentner, 1982), a child must have some semantic prelinguistic structures, to which verbs can ground.

In this paper we propose a computational model of grounded action-based formation of categories, which can serve as more cognitively plausible meanings of nouns, adjective and verbs in language models.

Categorization

Since the very infancy, each of us has encountered millions of different experiences of performing different actions with different objects under different circumstances with different outcome. Remembering each such experience individually could be just too hard and beyond our cognitive limits. Categorization serves the cognitive economy purposes – stored information about the category is reduced, nevertheless it should enable inducing further information about category members. More important, our personal experience would be intransferrable verbally to another person, unless he or she has the very same experience or unless our individual experiences are grouped to similar categories with shared conventional names. Thus the ability to categorize is vital both for our effective functioning in the world and for the effective communication.

Rosch (1978) has found out that people categorize at different levels leading to basic-level, superordinate and subordinate categories. Basic level is the most general level, at which a perceptual image of a category can be created, the level that adults use for naming when speaking to children, the level that children acquire the names for first and the level with the highest intra-cluster similarity and inter-cluster distinctiveness. Basic-level categories show prototype effects, i.e. some members of categories are considered more representative than the others.

In computational models, categories can be represented by a collection of intervals from discrimination trees (Steels, 2000), vectors in a multidimensional conceptual space with dimensions corresponding to attributes of objects (Gärdenfors, 2000), or weight configurations in artificial neural networks (Borghi, Parisi, & Di Ferdinando, 2005). In our approach, a category is represented by a *discrimination criterion*. A discrimination criterion for some concept c models the elementary ability to determine, to what extent the perceptual[1] input i is an instance of the concept c (Šefránek, 2002).

Perceptual Input

Objects Shortly after birth, a normally developing child can perceive and distinguish individual objects (Spelke, 1990). In our model, a cognitive agent is placed on a non-toroidal 2D grid together with other objects, which have properties that can change in time. The time changes in discrete steps. In each time step the agent perceives a *scene* – entities on the grid within some distance from the perceiver. The agent perceives each entity (including itself) as a frame of $\langle attribute : value \rangle$ pairs, e.g. $\{\langle weight : 10\rangle, \langle size : 3\rangle, \langle posX : 2\rangle, \langle posY : 6\rangle\}$. Formally a scene with k_t objects perceived at a time t is a set

$$Scene(t) = \{f_1, f_2, \ldots, f_{k_t}\}$$

of percepts, where each percept is a frame

$$f = \left\{\left\langle a, h_a^f \right\rangle \mid a \in A_f,\ h_a^f \in \mathbf{R}\right\}.$$

A_f is a set of all attributes[2] of the percept frame f and h_a^f denotes a numeric value of the attribute a of the frame f.

The frame is not yet a representation, but models the sensory input preprocessed by low-level sensory routines, such as the scene segmentation to particular objects, etc. If two agents perceived the same scene, they would have the same sets of input frames; however, they can categorize and represent them differently.

[1]In this model, we use only elementary criteria operating directly on the perceptual input. Composite criteria can categorize relations between objects, situations, events and can operate on an output of other criteria (Šefránek, 2002).

[2]Perceptual frames of different objects can have different attribute sets. However, in this paper, we assume that perceptual frames of the same object at different times will have the same attribute set.

Actions The agent has a repertoire of actions (e.g. touch, lift, move), each of which can be performed with particular parameters, e.g. pushing with different forces, stretching the arm at a different angle, walking with different sizes of step etc. The action type, e.g. lift or touch, is in our model the abstraction of a non-declarative implicit agent's knowledge of the action – a motor stereotype of invariant characteristics of the action, while action parameters are varying characteristics of a particular execution of the action.

Actions are in the model represented by pairs $\langle actionType, f\rangle$, where f is a frame of parameter values of the action, e.g. $\langle moveBy, \{\langle x : 0\rangle, \langle y : -8\rangle\}\rangle$. Actions performed in the environment change the properties of involved objects. For the sake of simplicity, an action in our model is always performed on a single object and can only change the properties of that object (e.g. lifting an object can change it's altitude, unless the object is too heavy).

Changes Children segment events in points of maximum change in input data (Zacks & Tversky, 2001). The agent in our model interacts with the environment by performing actions on objects and it learns from this experience by observing the resulting changes. Let f be a percept of an object o at time t and f' be a percept of the same object at the previous time step $t-1$. Then the observed change of the object o at time t is represented as a frame of attributes with non-zero changes:

$$\Delta f = \left\{\left\langle a, \Delta h_a^f \right\rangle \mid a \in A_f,\ \Delta h_a^f = h_a^f - h_a^{f'} \neq 0\right\}.$$

Discrimination Criteria

Each discrimination criterion representing a particular concept is a membership function that takes a frame as an argument and returns a value from the closed interval $[0, 1]$, expressing to what extent is the frame an instance of the concept (0 means not at all, 1 means the best, prototypical example).

A discrimination criterion records the mean and variance of each attribute common to all instances of the concept seen so far – a set of frames $\{f_1, f_2, \ldots, f_n\}$. Formally

$$r = \langle \{\langle a \in A_r, h_a^r, var_a^r \in \mathbf{R}\rangle\}, n \in N\rangle,$$

where

$$A_r = \bigcap_{i=1}^{n} A_{f_i},\ h_a^r = \overline{\left\langle h_a^{f_i} \right\rangle_{i=1}^{n}},\ var_a^r = \sigma^2 \left\langle h_a^{f_i} \right\rangle_{i=1}^{n}.$$

Upon encountering a new instance of the concept, the criterion is *updated* by the instance, i.e. the attribute set, means and variances are iteratively recomputed and n is increased. When *creating a new criterion* r from the first example f, $A_r = A_f$, variances of all attributes are set to zero and the means are set to the attribute values of the example.

The discrimination criterion r as a membership function then evaluates the similarity of the input percept f

with the mean case of the category inversely weighted by the variances of particular attributes:

$$r(f) = sim(r, f) = e^{-k\,dist(r,f)}, \tag{1}$$

where

$$dist(r, f) = \sqrt{\frac{1}{|A_r|} \sum_{a \in A_r} \frac{(h^f_a - h^r_a)^2}{var^r_a}} \tag{2}$$

and k is a constant defined below. Viewing the similarity as an exponentially decaying function of the distance is common in the psychological literature (Shepard, 1987). The distance function evaluates an average distance of the percept's attribute values from the attribute values of the mean case in standard deviation units (i.e. if a frame would differ from the mean by $m\sigma_a$ in each attribute a, the distance function would return m). If we further define in equation (1) $k = -\frac{1}{2}\ln(0.5)$, the similarity function returns 1 for zero distance and 0.5 for the distance 2σ, which could be taken as a decision threshold deciding whether the percept is an instance of the category or not. The instance of the category with an average distance σ would have the similarity approximately 0.7.

This representation resembles the conceptual spaces (Gärdenfors, 2000) with attributes corresponding to dimensions of the conceptual space and the mean case being a central member (prototype).[3] The dimensions corresponding to attributes with low variance would be "stretched out", thus having a bigger decisive weight for category membership. On the other hand, if observed instances of a category vary much in an attribute, it would be unimportant for category membership. If we define $0/0 = 0$ and $x/0 = \infty$ for $x \neq 0$, the mean value of an attribute with zero variance will be mandatory for instances of the category (any other value would yield zero similarity). However, this would lead to implausible behavior for criteria with $n = 1$ (induced from just one example). As variances of all attributes have yet been zero, this criterion would yield similarity 1 for percepts exactly the same to the example and zero for all others, regardless of how close they are to the example. On the contrary, children can generalize even from just one example (Bloom, 2000). If variances still remain zero after the agent has seen many examples of the category, then it would be reasonable to induce that this category represents an individual. We overcome this problem by substituting var^r_a in the equation (2) with the estimated variance $evar^r_a = var^r_a + v_0/n^s$, where v_0 is an initial estimated variance and s is the speed of decay toward var^r_a.

Using the above described criteria with variances has many interesting and cognitively plausible effects. First, the categories can represent concepts with different levels of generality from individuals to general concepts with large variances. Second, the concepts with very large variances in all but one attribute can semantically represent an adjective (e.g. if all examples had big size, while differing in all other attributes). Third, the similarity judgments will be asymmetric, i.e. the mean case of one category can be more similar to the other category than vice versa.[4]

Simple Categories

By perceiving objects, their changes and performed actions, our agent will learn and store their categorical representations[5] in three separate sets of discrimination criteria C_o, C_c, C_a. This is cognitively plausible, as similar separate representational systems exist in humans (Ungerleider & Mishkin, 1982; Orban et al., 1995; Rizolatti et al., 1996), and the representations remain perceptual (Barsalou, 1999). The core problem of categorization is to decide when to update the existing discrimination criterion most similar to the percept and when to create a new one. The decision can be problem-dependent and we describe it in the following section.

Complex Categories

According to Gibson (1979), we perceive in order to operate on the environment. Gibson called the perceivable possibilities for action *affordances*. The cognitive agent in our model interacts with the environment in each time step by choosing an action from it's repertoire, performing it with random parameters[6] on an object randomly picked up from the currently perceived scene and observing the resulting change of the object.[7]

The agent represents the knowledge of causal relations between actions, objects and changes in the form of associations among their respective categories (analogy to associative areas of the cortex). Formally the agent's association system V is a set of triples

$$V = \{\langle r_a \in C_a, r_o \in C_o, r_c \in C_c \rangle\}.$$

Objects and actions are grouped to categories by the change. That is, if an action leads to the same change on several objects, they will all fall in the same category and vice versa. All action categories associated with some object category represent agent's knowledge of affordances of the object, while all object categories associated with an action category form the precursor of a verb-centered semantic representation – a verb island (Tomasello, 1992).

The Association Algorithm Initially, the agent starts with the empty category sets C_o, C_c, C_a and the

[3] Unlike concepts of Gärdenfors, regions of concepts with weighted dimensions need not be convex, they even need not be connected (Moravčík, 2005).

[4] People show the same effect, e.g. Tel Aviv is judged more similar to New York than vice versa (Tversky, 1977).

[5] Discrimination criteria for actions store also the action type and return zero for all action percepts, which do not have the same action type.

[6] The values of parameters can be limited by "physical" possibilities of the agent, i.e. reaching only to some height or pushing with limited force.

[7] The agent 'assumes' that all observed changes of the object were caused by it's action. Children have the similar attitude, called "magical causality" (Piaget & Inhelder, 1966), even short after their sensory-motor stage.

empty association set V. These sets can be modified in each time step, after the agent perceives a triple $\langle f_a, f_o, \Delta f_o \rangle$ of percepts of a performed action, object and it's change, in the following way:

1. Find in V the most similar[8] association $v^* = \langle r_a, r_o, r_c \rangle$ to the input triple $\langle f_a, f_o, \Delta f_o \rangle$. (If there is no association with non-zero similarity, create a new one either by reusing existing categories, if they are individually similar enough to the percepts, or by creating new categories.)

2. If $sim(r_c, \Delta f_o) > \theta(t)$, update r_a by f_a, r_o by f_o and r_c by Δf_o, otherwise:

3. if $r_o(f_o) > r_a(f_a)$, create a new action category from f_a, else create a new object category from f_o and use it to form a new association.

In step 2, if the change category of the association is similar enough to the perceived change, the percepts are considered to be the instances of the associated categories and all three categories are updated by the percepts. Otherwise, a new category is created for the less similar percept of either the object, or the action (step 3).

The prediction threshold $\theta(t)$ determines the precision of the representation. It can be constant during the whole simulation or it can increase in time to model the child's growing ability to distinguish differences in the environment.

Merging It can happen that some categories, which started independently, become very similar after being updated by more examples. In our model, such similar categories are sought and merged in the following way: if any of the systems C_o, C_c, C_a contains two criteria r_1, r_2 with mean cases f_1, f_2, such that $\min(sim(r_1, f_2), sim(r_2, f_1)) > 0.9$, they are replaced by a new criterion r with $A_r = A_{r_1} \cap A_{r_2}$. The means and variances of attributes of r are computed from those of r_1 and r_2 as if they were characteristics of the union of example sets of the original criteria. The merging can propagate to the association level – if, after merging some categories, there exist associations $\langle r_a, r_o, r_{c_1} \rangle$ and $\langle r_a, r_o, r_{c_2} \rangle$, such that $r_{c_1} \neq r_{c_2}$, r_{c_1} and r_{c_2} are merged.

Experimental Results

Environmental Setting

In our experiment the agent and 30 other objects – 10 "fruits", 10 "toys" and 10 "pieces of furniture" were placed on random positions of a 25×25 grid. The initial values of object attributes were randomly generated as uniformly chosen integers from respective intervals of the pattern {*weight*: 20, *age*: 3, *posX*: [0, 24], *posY*: [0, 24], *posZ*: 0} for the agent, {*weight*: [1, 3], *size*: [1, 49], *color*: [0, 4], *roundness*: [0, 9], *posX*: [0, 24], *posY*: [0, 24], *posZ*: 0} for fruits, {*weight*: [1, 9], *color*: [0, 9], *cries*: [0, 1], *dressed*: [0, 1], *posX*: [0, 24], *posY*: [0, 24], *posZ*: 0} for toys, and {*weight*: [20, 49], *size*: [20, 49], *legs*: [0, 4], *material*: [0, 9], *posX*: [0, 24], *posY*: [0, 24], *posZ*: 0} for pieces of furniture.

The agent was trying to either lift objects, or put them down. In each time step he randomly chose one of the objects and performed on it an action randomly generated from the pattern ⟨*actionType*: *liftUp*, {*armPosIncrease*: [1, 9], *force*: [1, 19]}⟩ or ⟨*actionType*: *putDown*, {*armPosDecrease*: [1, 9]}⟩.

The effects of the action on the chosen object were simulated by the environment. In the case of *liftUp* action, the *posZ* attribute of the object was increased by the value of *armPosIncrease*, if the force was greater than the weight of the object, otherwise the action had no effect. In the case of *putDown* action, the *posZ* attribute of the object was set to $\max(0, posZ - armPosDecrease)$.

Measures and Parameters

In each time step after performing an action, the agent adapted it's representation according to the association algorithm described above. To evaluate the usefulness and adequateness of the representation, we measured it's ability to predict the correct result of the action. After choosing an object and an action, the agent found the association with the highest similarity of object and action.[9] The change criterion of that association was then applied to the perceived change and the resulting similarity was recorded as a *prediction*. However, the more general change criteria give higher similarity values, thus we also measured a *generality* of the prediction expressed by average standard deviation of attributes of the criterion used for prediction (lower value means higher accuracy of the prediction). We have also measured the number of criteria in the agent's representation. Each measure has been averaged over the time window of 20 last steps.

The parameters of the association algorithm were $v_0 = 30$, $s = 1$, $w_o = 1$, $w_a = 100$, $w_c = 1000$. We used the prediction threshold $\theta(t)$ linearly increasing from $\theta(0) = 0$ to $\theta(700) = 0.7$ and constantly equal to 0.7 for $t > 700$. Results of experiments were averaged over 30 simulation runs.

Results

General Results In the first experiment (Figure 1a), the agent did not use merging. As we can see in the graph, while the prediction threshold is low, the agent only uses a few basic criteria. After the threshold rises over a certain value (around 0.5), the number of criteria starts to rapidly increase, which leads to a better accuracy of the prediction. As the threshold stabilizes at the value of 0.7, the total number of criteria slowly saturates, together with the generality exponentially decaying to a certain value. The prediction value converges to approximately 0.7. Recall that this value corresponds to the

[8] The similarity is computed using equation (1) from the weighted sum of distances $w_a dist(r_a, f_a) + w_o dist(r_o, f_o) + w_c dist(r_c, \Delta f_o)$. However the overall similarity is zero, if r_a, f_a do not have the same action type.

[9] Again computed from their weighted distances, see the footnote 8.

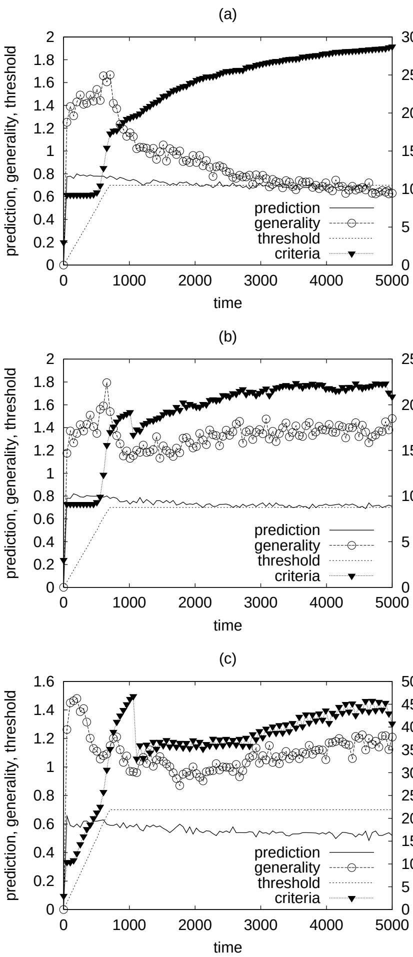

Figure 1: (a) No merging, number of criteria saturates and prediction value converges to that of average intra-category distance (i.e. predictions are correct). (b) Merging keeps the number of criteria lower at the cost of lower accuracy (higher generality). (c) Prototypes (criteria without variances) need much more criteria and still have lower prediction value.

average distance σ, which is an average intra-cluster distance of the category, hence, this means that the criteria give correct predictions.

Merging In the second experiment (Figure 1b), the agent merges similar criteria every 50th time step since the time 1000. This decreases the total number of the criteria at the cost of more general predictions (the generality does not exponentially decay, but stays between 1 and 1.5). The prediction value again converges to approximately 0.7, i.e. the criteria give correct predictions.

Comparison to Prototypes The advantage of the proposed representation is that it can cope with different importance of attributes (or scaling of different dimensions) by recording their intra-category variances. In order to compare it with a standard prototype representation, we ran an experiment, where criteria behaved like

Table 1: Example of object criteria (above) and associations (below).

	posX	*posY*	*posZ*	*weight*	*color*
C1	13 ± 7	13 ± 8	0 ± 0	37 ± 23	
C2	14 ± 7	13 ± 8	4 ± 13	39 ± 22	
C3	11 ± 6	10 ± 7	35 ± 28	4 ± 3	2 ± 2
C4	11 ± 6	10 ± 7	25 ± 23	4 ± 3	

	Action	
Category	$putDown(5 \pm 3)$	$liftUp(6 \pm 2, 10 \pm 6)$
C1	no change	
C2		no change
C3	Δ=$\{posZ\colon -6 \pm 1\}$	Δ=$\{posZ\colon 7 \pm 1\}$
C4	Δ=$\{posZ\colon -4 \pm 2\}$	Δ=$\{posZ\colon 5 \pm 2\}$

Table 2: Number of objects of each type for a category they are most similar to.

	Category			
Object type	C1	C2	C3	C4
agent	1			
fruit			8	2
toy		1	3	6
furniture	5	5		

prototypes in conceptual spaces in that the terms of the sum in the equation (2) were not divided by the variances (Figure 1c). Despite that the criteria were merged as in the second experiment, the number of criteria is almost double and the prediction value is lower than in the case with variances.

Representation in Detail In an example run of the experiment with merging, the agent acquired 4 object criteria, 7 action criteria, 10 change criteria and formed 13 associations. In Table 1 we can see the four object criteria and a fragment of the associations (for brevity the attribute values are written as $h_a^r \pm \sigma$, where $\sigma = \sqrt{var_a^r}$). Table 2 shows the object criteria applied to 31 objects on the grid. Numbers in a row express the object counts of a given type most similar to the criterion in a column. As we can tell from Table 1, the category C1 represents objects on the ground (they cannot be put down) and C2 objects too heavy to lift, C3 mostly fruits and C4 mostly toys. As attributes other than *weight* or *posZ* are present in the criteria too, they could help the agent in classification (e.g. if all heavy objects were in the same part of the grid, or had some specific color). Hence, the representation is situated and encoding the learning context.

Discussion

In this paper we have proposed a cognitively relevant representation of categories based on discrimination criteria and their cross-categorial associations, allowing for graded category membership and prototype effects, sensitivity to differences in intra-category variability of particular properties, hierarchical categories of different generality, and situated perceptual representation of objects, changes and actions.

The categories and their cross-associations are formed by sensory-motor interaction with the environment. The process is action-based, in that objects are grouped in the same category, if an action performed on them yielded the similar result. Once an agent can represent predictions about the outcome of it's actions, it can use them for planning sequences of actions – macro-affordances (Borghi, 2005) to satisfy it's needs and goals. Hence, cross-categorial representation is inherently contextual – different affordances are picked-up depending on the desired effect on the environment.

The proposed representation provides necessary grounding for models of acquisition and transmission of meanings and language. Although we use computational-level description in the form of frame-like structures, these correspond to neurally embodied perceptual symbols and can be in principle implemented in the spirit of structured connectionism (Feldman, 2006). Among recent connectionist models of action-based categorization, the simulations of Borghi et al. (2005) are the most closely related to our work. In their simulations, an organism with a visual system and a two-segment arm reaches different points in space depending on the object seen and on a context. However, the organism is selected from a population of non-learning neural networks by genetic algorithm, which is not plausible as a model of ontogenetic acquisition of categories.

Our current research is directed toward exploring the interactions between categorization and language acquisition.

Acknowledgments

Many thanks to J. Šefránek, I. Farkaš and Ľ. Beňušková for their good advices. This work was supported by the grant VEGA 1/3105/06 and by Agency for Promotion Research and Development under the contract No. APVV-20-P04805.

References

Barsalou, L. W. (1999). Perceptual symbols systems. *Behavioral and brain sciences*, *22*, 577-660.

Bloom, P. (2000). *How children learn the meanings of words.* Cambridge, MA: MIT Press.

Borghi, A. M. (2005). Object concepts and action. In D. Pecher & R. A. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thinking.* Cambridge University Press.

Borghi, A. M., Parisi, D., & Di Ferdinando, A. (2005). Action and hierarchical levels of categories: A connectionist perspective. *Cognitive Systems Research*, *6*, 99–110.

Feldman, J. (2006). *From molecule to metaphor: A neural theory of language.* Cambridge, MA: MIT Press.

Fodor, J. (1981). *Representations: Philosophical essays on the foundations of cognitive science.* Cambridge, MA: MIT Press.

Gärdenfors, P. (2000). *Conceptual spaces.* Cambridge, MA: MIT Press.

Gentner, D. (1982). Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. In S. Kuczaj (Ed.), *Language development: Language, cognition, and culture.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston: Houghton Mifflin.

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

Kirby, S., & Hurford, J. (2001). The emergence of linguistic structure: an overview of the iterated learning model. In D. Parisi & A. Cangelosi (Eds.), *Computational approaches to the evolution of language and communication.* Berlin: Springer-Verlag.

Langacker, R. (1991). *Concept, image and symbol: The cognitive basis of grammar.* Mouton de Gruyter.

Moravčík, A. (2005). *Conceptual spaces with weighted prototypes.* Unpublished master's thesis, Comenius University, Bratislava.

Orban, G. A., et al. (1995). A motion area in human visual cortex. *Proc Natl. Acad. Sci. USA*, *92*, 993–997.

Piaget, J., & Inhelder, B. (1966). *La psychologie de l'enfant.* Paris: PUF.

Rizolatti, G., et al. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, *3*, 131–141.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Shepard, R. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*, 1318–1323.

Spelke, E. S. (1990). Principles of object perception. *Cognitive Science*, *14*, 29–56.

Steels, L. (2000). Language as a complex adaptive system. In M. Schoenauer (Ed.), *Proceedings of PPSN-VI.* Berlin: Springer-Verlag.

Tomasello, M. (1992). *First verbs: A case study of early grammatical development.* Cambridge: CUP.

Tomasello, M., & Merriman, W. E. (Eds.). (1995). *Beyond names for things: Young children's acquisition of verbs.* Mahwah, NJ: Lawrence Erlbaum Associates.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(4), 327–352.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior.* Cambridge, MA: MIT Press.

Šefránek, J. (2002). Kognícia bez mentálnych procesov. In L. Beňušková et al. (Eds.), *Kognitívne vedy* (pp. 200–256). Bratislava: Kaligram.

Whorf, B. L. (1956). *Language, thought and reality: Selected writings of Benjamin Lee Whorf* (J. B. Carrol, Ed.). Cambridge, MA: MIT Press.

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, *127*, 3–21.