Available online at www.sciencedirect.com

ScienceDirect

Cognitive Systems Research 9 (2008) 293–311

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

Autonomous construction of ecologically and socially relevant semantics

Action editor: Vasant Honavar

Martin Takáč [*]

Department of Applied Informatics, Comenius University, Mlynská dolina, 842 48 Bratislava, Slovakia

Received 16 March 2007; accepted 16 December 2007
Available online 11 July 2008

Abstract

This article presents a synthetic modeling approach to the problem of grounded construction of concepts. In many computational models of grounded language acquisition and evolution, meanings are created in the process of discrimination between a chosen object and other objects present on the scene of communication. We argue that categories constructed for the purpose of identification rather than discrimination are more suitable for the detached language use (talking about things not present here and now). We describe a semantics based on so-called identification criteria constructed by extracting cross-situational similarities among instances of a category, and present several computational models. In the model of individual category construction, the instances are grouped to categories by common motor programs (affordances), while in the model of social learning, focused on the influence of naming on category formation, entities are considered members of the same category, if they are labeled with the same word by an external teacher. By these two mechanisms, the learner can construct interactionally grounded representation of objects, properties, relations, changes, complex situations and events. We also report and analyze simulation results of an experiment focused on the dynamics of meanings in iterated intergenerational transmission.

Keywords: Language acquisition; Concept formation; Symbol grounding; Cognitive semantics; Computational modeling; Iterated learning

1. Introduction

Constructing computational models of cognitive phenomena has shown to be a fruitful approach in artificial intelligence and cognitive science. Such an approach to cognition is called synthetic modeling and can be characterized as "understanding by building" (Pfeifer & Scheier, 1999). In this article, we adopt the synthetic modeling approach to study the processes of concept formation and language acquisition, and their mutual relations.

Human language is a complex phenomenon that has been co-evolving on different timescales. While necessary physical and cognitive faculties have been shaped by biological evolution, languages themselves undergo changes and evolve in the process of iterated intergenerational cultural transmission on a historical rather than phylogenetic timescale (Deacon, 1997; Kirby & Hurford, 2001). The ontogenetic process of language acquisition determined by innate learning mechanisms imposes further structure on an emerging and evolving language (Briscoe, 2001).

The language acquisition process can, to a large extent, be 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 by the influence of language itself (Whorf, 1956). The nature and origin of linguistic meanings and their relation to the environment is

[*] Tel.: +421 904 842351; fax: +421 2 654 22 263.
E-mail address: takac@ii.fmph.uniba.sk

1389-0417/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.cogsys.2007.12.001

an important issue in computational models of language and communication. Some models (e.g. Kirby & Hurford, 2001; Kvasnička & Pospíchal, 1999) abstract away from this issue by using artificial, predefined and fixed meanings, which bear no relevance to any environment. These models are subject to the *symbol grounding problem* (Harnad, 1990). Other models (Cangelosi, 2005; Roy, 2005; Smith, 2003; Steels & Belpaeme, 2005; Steels & Kaplan, 2001b; Takáč, 2007; Vogt, 2005) address this problem in a constructivist manner: each communication participant constructs its own individual meanings by interacting with the environment and other participants.

In most of the models, the communicative goal is to uniquely identify a chosen static object by discriminating it from all other objects currently present in the communicative situation. The proposed representation-forming mechanisms are tailored to this goal; they are based on capturing differences between the present objects. However, language goes beyond a present situation, and its important function is to enable communication about things not present here and now (Gärdenfors, 1996b). In this article, we argue that representation based on noticing similarities rather than differences is more suitable for this goal. More importantly, meanings should not be limited to object categories; they should also include properties, relations, dynamic changes, situations and events. Adhering to these principles, we study mechanisms of ontogenetic meaning construction with synthetic modeling methodology:

(1) We propose a similarity-based semantic representation of various types of meanings.
(2) We propose individual and social mechanisms of autonomous construction of such semantic representations.
(3) We implement the hypothesized mechanisms in computational models and analyze the results of simulations.

The rest of the paper is organized as follows: After discussing general issues of meaning formation (Section 2), we present several computational models. In Section 3, we describe common features of all the models and propose a semantic representation of various categories based on cross-situational similarities. Then we present experiments with the models of individual (Section 4) and social (Section 5) mechanisms of meaning creation. In Section 6, we let the meaning formation process iterate intergenerationally and we analyze its dynamics. Final two sections contain general discussion and concluding remarks.

2. Meanings and their origin

In the cognitivist semantic tradition, meanings are not objectively existing "out there" in the world but are rooted in people's experience. They are often viewed as mental entities or internal representations (Gärdenfors, 2000). The internal representations bear no intrinsic meanings *per se* (Harnad, 1990); they get it via structural coupling with the world (Maturana & Varela, 1987; Vogt, 2002). It has been argued that lexical meanings are embodied conceptual structures (Gärdenfors, 2000; Lakoff, 1987) grounded in perception and action (Barsalou, 1999; Pecher & Zwaan, 2005), physically corresponding to activations of neural structures, correlated with perceiving, performing, imagining or talking about the content they represent (Feldman, 2006; Pulvermüller, 1999; Rizolatti et al., 1996).

The coupling of the representations with the world has two components: individual and social. The former one, called *physical symbol grounding*, refers to the ability of each individual to create an intrinsic link between world entities and internal representations, while the latter one, called *social* (or *external*) *symbol grounding*, refers to the collective negotiation for the selection of shared symbols and their meanings (Cangelosi, 2006). A corollary of this is that meanings are individually created subjective constructs, which need to be attuned to each other collectively (Takáč, in press). Philosophically, this approach is essentially constructivist: physical symbol grounding roughly corresponds to cognitive constructivism based on the work of Piaget (1937/1955) in that individuals actively construct their own meanings through cognitive processes, based upon their past experiences and their interactions in the world. Social symbol grounding is close to social constructivism (Vygotsky, 1978), in the sense that individually created meanings are motivated and constrained by the social context.

2.1. Individual meanings

The foregoing study of the individual meaning creation process concerns individuals situated in an environment, achieving their goals by sensing and acting. We will call such individuals *agents*. In the most elementary sense, agents attribute meanings to parts of their environment by recognizing, via their sensors and actuators, information useful for achieving their goals (Nehaniv, 2000). In this sense, meanings are preverbal: The embodied knowledge coming from perceiving and acting in the environment can be found in living organisms well before the appearance of a language, both in phylogeny (Kováč, 2000) and ontogeny (Piaget & Inhelder, 1966; Spelke, 1990).

Gärdenfors (1996a) distinguishes between two kinds of representations: *cued* and *detached*. A cued representation must always be triggered by something present in the current situation, while a detached representation may stand for objects and events neither present nor triggered by the current situation of the organism. Any organism reacting to certain states of its environment in certain ways (e.g. eating objects recognized as food and avoiding objects recognized as predators) performs categorization and possesses cued representations of the respective categories. A chimp looking for a (non-present) twig to reach for a banana possesses a detached representation of a twig and its use. It is speculated that the appearance of detached representations in phylogeny is coordinated with the development of neocortex (Gärdenfors, 1996b); in ontogeny it corresponds to *object permanence* (Piaget & Inhelder, 1966). Possession of detached representation is a necessary condition for higher cognitive functions such as planning, deception, self-awareness and linguistic communication (Gärdenfors, 1996a).

Cued representations observable as non-volitional behavioral reactions are innate and have evolved phylogenetically. In this article, we pay attention to representations acquired in ontogeny. We propose a computational model of an agent constructing meanings grounded in its sensorimotor interactions with the environment (see Section 4).

2.2. Social meanings

Human language is a fully detached communication system, in the sense that it enables talking about things not present here and now, even about things that cannot exist physically. According to Gärdenfors (2004), language evolved in order to make cooperation about future goals possible. This seems to create a paradox: cooperation requires socially shared meanings, but meanings are constructed individually by each communication participant. How can the participants understand each other? There are several answers to this paradox.

First, although individual meanings are not identical, they must be sufficiently similar thanks to similar learning mechanisms and experiences in a shared environment (Steels, Kaplan, McIntyre, & Looveren, 2002). If the meanings are not sufficiently similar, the communication ends up in misunderstanding.

Second, the intended meaning of the speaker is inherently ambiguous: it cannot be *transferred*, but it must be *inferred* by the hearer from the pragmatic context. Inference of the meaning is problematic, as it has been stated by Quine (1960) in the famous *Gavagai problem*.[1] In the next section, we review several developmental strategies that tackle this problem.

Third, common social meanings can be viewed as constantly renegotiated moving equilibria emerging from the process of mutual coordination of individual meanings of language users (Gärdenfors, 2000). This self-organizing process was modeled by Steels (2000), showing how a globally coherent language can emerge from scratch as a result of local interactions of language users. The community of language users was modeled by a multi-agent system, in which agents (simulated or embodied in real robots) played various types of *language games* by picking a topic from the environment and describing it with a chosen linguistic form. In the course of time, each agent adapted its linguistic behavior according to the history of previous interactions. A positive feedback between the selection of a language form and its success in use resulted in self-organization and the emergence of a coherent lexicon.

2.3. The Inference of meanings

According to Wittgenstein (1953), words get their meaning in use. The language is primarily about action in the real world. In pragmatics, a distinction is made between the literal meaning of an utterance and the speaker's intended meaning. The speaker can have a variety of communication goals; drawing the hearer's attention to something, describing something, giving information, asking a question, making a request or giving an order are typical examples of *speech acts* (Austin, 1962).

In bootstrapping a language from scratch, it is particularly important to establish shared meanings of referential expressions, such as names, nouns and adjectives (Gärdenfors, 2004). This happens by referential (labeling) acts that draw attention to objects present on the scene of communication, with the help of non-verbal means such as pointing, gaze following or joint attention (Tomasello & Farrar, 1986). The whole matter is complicated by the fact that a word uttered along with a non-verbal reference to an object can label the object, any of its parts or properties, its superordinate class and many other things (Quine, 1960). Children use several strategies to overcome this problem: they assume that novel words refer to *whole objects* (Markman, 1992), that a novel word cannot name an object that already has a name (the *mutual exclusivity constraint* Markman, 1992), that any difference in form marks a difference in the meaning (the *principle of contrast* Clark, 1987). They also disambiguate meanings by occurrences of their referents in multiple situations (Akhtar & Montague, 1999; Waxman & Braun, 2005).

In computational models of language bootstrapping (e.g. Steels & Belpaeme, 2005; Steels & Kaplan, 2001b; Vogt, 2002), agents communicate about objects present on the scene of communication.[2] A typical communicative goal of the speaker is to uniquely identify a particular object present in the environment, chosen as the topic. The speaker first plays a *discrimination game* (Steels & Kaplan, 1999) in order to find internal meanings (a set of categories or features) that distinguish the topic from all other objects in the context. In case of failure, the representation of meanings is refined. In case of success, the speaker tries to lexicalize the selected meanings. In the *guessing game* (Steels & Kaplan, 1999), the speaker utters an expression and the hearer tries to guess what referent the speaker names. In the *observational game* (Oliphant, 1997), the speaker narrows down the set of possible referents by e.g. pointing, and the hearer adapts its lexicon by Hebbian learning. If the context contains more than one object, meanings are disambiguated cross-situationally (Siskind,

[1] Suppose we have a linguist observing a native speaker of a foreign language pointing to a rabbit and saying "Gavagai". The linguist cannot be sure what 'Gavagai" means, because it could mean 'rabbit', 'animal', 'white', 'fur' and many other things.

[2] The objects are geometric shapes of various colors pasted on a magnetic white board (Steels & Kaplan, 2001b), different light sources (Vogt, 2002), or colored Munsell chips (Steels & Belpaeme, 2005).

1996; Smith, 2005a). In all the models, the hearer's inference of the meaning is constrained by the assumption that the scene only contains one referent of the speaker's utterance (semantic hypotheses with more referents are excluded from consideration). In some models, the hearer also assumes mutual exclusivity, i.e. it excludes from consideration all those objects on the scene, for which it already knows an appropriate word (Smith, 2005b).

2.4. Discrimination versus identification

The communicative goal of unique discrimination in a particular context determines the shape of the meanings. The categories that have evolved for the purpose of discrimination do not have to be natural and suitable for other purposes, as has been argued by Harnad (2005), who distinguishes between discrimination and identification (categorization). Discrimination is a *relative* judgment between things that are present simultaneously, while identification (categorization) is an *absolute* judgment of a thing alone answering the question whether or not a given input is a member of a particular category (Harnad, 1990; Harnad, 2005). The importance of identifying categories (kinds) of things is even higher for detached use of language, i.e. talking about things not present, which is apparently a different situation from a discrimination task.

Ecologically valid categorization should reflect what can be done with different kinds of things. People categorize at different levels, of which the *basic level* has a special status (Rosch, 1978). It is the most general level, at which a common perceptual image and a common motor program can be created for members of a category, and the level with the highest intra-cluster similarity and inter-cluster distinctiveness. Basic-level categories support inductive inferences, i.e. deriving further properties of objects from their membership in a category. When adults speak to children, they tend to use words for basic-level categories and these words are understood and acquired by children first (Rosch, 1978). Naming influences the children's conceptual organization and supports discovery of novel concepts: using distinct names for distinct objects motivates looking for *differences* and supports individuation, while using the same name for distinct objects motivates looking for *similarities* and supports categorization (Waxman & Braun, 2005).

In Section 5, we explore the influence of naming on a category formation process. We present a computational model of two agents: a teacher describing various aspects of the scene, and a learner inducing meanings of words by noticing cross-situation similarities between their referents. The important novel contribution of this model is that meanings of words are not only whole objects, but also their properties, mutual relations and dynamic changes in time.

3. Computational modeling of grounded semantics

We have designed and implemented several computational models simulating different mechanisms of meaning construction. In this section, we describe a general framework common to all the models.

A model typically consists of one or more agents and a simulated environment. The environment that contains some entities (objects) is dynamic and open, in the sense that the objects can appear or disappear or change their properties in discrete time steps. In each time step, the agent senses its environment, updates its internal representation and can communicate or perform other actions, depending on a particular application (see Fig. 1).

3.1. Perception

We describe the perceptual input of the agent at time t as a set of frames of the attribute + value pairs with one frame for each object in the environment. To be able to also perceive changes of properties of objects, the agent must track individuals in time. Hence, we establish the correspondence between frames of the same object at different times.

Frames represent all perceptual inputs relevant to the agent, e.g. objects in the physical environment of the agent, incoming data, or even the agent's "proprioceptive" input (values of internal variables, parameters of operations performed, position of an arm, etc.). The names of attributes have no special meaning for the agent, except for establishing a correspondence of attributes with the same name in different frames. The values of attributes are real numbers.

Formally, a perceptual frame f is characterized by the set of attributes A_f and the real-valued attribute accessor function $h_f : A_f \to R$. In the text, we will use a more conventional notation $f.a$ instead of $h_f(a)$.

Frames have a long tradition in artificial intelligence (Minsky, 1975) and cognitive semantics (Fillmore, 1982); they are easily readable by human observer, general enough to describe various data structures, and they can be implemented in the spirit of structured connectionism (Feldman, 2006; Shastri, Grannes, Narayanan, & Feldman, 1999).

Using frames or other arbitrary amodal symbols as semantic representations has been criticized by Barsalou (1999). Barsalou defines representation as amodal, if its internal structure bears no correspondence to the perceptual states that produced it. However, frames in our architecture are not yet a representation, but a computational-level description of structures resulting from a low-level preprocessing of the perceptual input.[3] For example, a frame can represent an array of intensity values of retina or a camera image, or a scene segmentation to perceptual characteristics of particular objects. Hence, two agents perceiving the same scene can receive identical sets of input frames; however, they can categorize and represent them differently.

[3] Autonomous construction of discrete perceptual structures from raw continuous sensory readings was modeled e.g. by Rosenstein and Cohen (1998) and Kuipers et al. (2006).

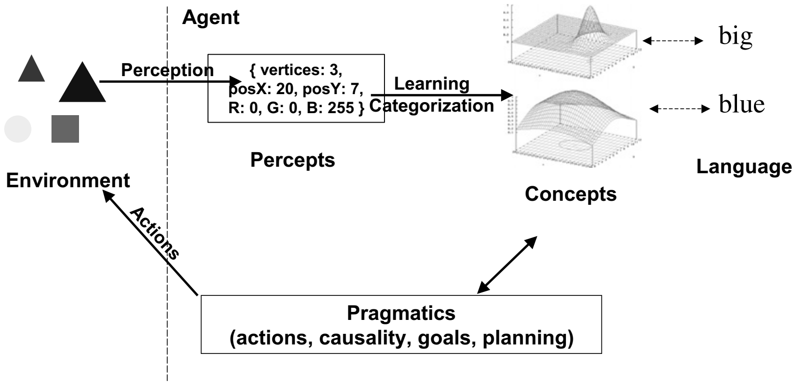

Fig. 1. The cognitive architecture of the agent includes perception, representation, language and pragmatic modules.

3.2. Representation of meanings

The agents gradually learn to distinguish environmental properties and group entities similar in some respect into categories. Each category is represented by an *identification criterion*[4] – an activation function that returns, for some input, the degree of the input's membership in the category. The possible inputs include a perceptual frame of one object (for criteria of objects and properties), perceptual frames of several objects (relational criteria), frames of the same object in different times (change criteria) and output activities of other criteria (compositional criteria of situations and events). The agents construct all their criteria from scratch by extracting common statistical properties of examples of categories encountered during their lifetime.

Extraction of statistical properties and categorization of novel examples is realized in *locally tuned detectors*, which form the core of every identification criterion. A locally tuned detector takes as input one frame and returns a value from the closed interval $[0, 1]$, expressing to what extent the frame is an instance of the category (1 means the best, prototypical example).

3.2.1. Geometrical view on categories

Locally tuned detectors have an intuitive geometric interpretation based on *conceptual spaces* (Gärdenfors, 2000). A conceptual space is a geometric space with dimensions corresponding to attributes of represented objects. Because not all attributes are relevant to all represented entities, dimensions are organized in domains. A particular object is represented as a point (vector of coordinates) in a subspace of one or several domains; the similarity between two objects is inversely proportional to the distance of their point representations in the conceptual space (for the distance to be evaluated, objects must share some domains or a subspace with a common metric). Representation of natural categories is based on the convexity assumption: if two points represent objects that are good examples of a category, then any point in between them must also be a good example of that category (Gärdenfors, 2000). Hence, natural concepts are represented by convex regions in the space. Geometrical centroids of the regions correspond to the best examples – *prototypes* of categories (Rosch, 1978).

Perceptual frames defined in the previous section can be viewed as vectors in the respective subspaces with dimensions determined by the attributes of the frames. A locally tuned detector should react with high activity to the convex hull of the vectors corresponding to examples of the represented category. The detectors represent categories with fuzzy boundaries (with their activity expressing the degree of category membership), but for practical purposes we can establish a decision threshold. In this case, the *receptive field* of a detector $r : D \rightarrow [0, 1]$ in the input space (domain) D, defined as the set $\Psi_\theta(r) = \{x \in D \mid r(x) > \theta\}$ for some decision threshold θ delineates a category. Locally tuned detectors have a high neural and biological plausibility (Balkenius, 1999; Hassoun, 1995; Martin, 1991).

3.2.2. Construction of locally tuned detectors

Functioning of the detectors is based on expressing common statistical properties of category examples in geometric terms and evaluating category membership as a distance in a conceptual space.

Let us assume that the agent has to induce a detector from a sequence of example frames $\{f^{(1)}, \dots, f^{(N)}\}$, each of which can be represented by a point

$$\left(f^{(i)}.a_1, \dots, f^{(i)}.a_{|A_{f^{(i)}}|}\right)$$

in the respective conceptual space with dimensions corresponding to attributes $a_j \in A_{f^{(i)}}$. The induction is based on properties of values of attributes *common to all frames*. Hence, each frame $f^{(i)}$ is projected into a common subspace $\mathscr{A}$ with dimensions from intersection of all attribute sets $\bigcap_{i=1}^{N} A_{f^{(i)}}$. Attributes not present in every example are considered irrelevant for the category membership. From now on, we will represent the sample of the category as a set of projected vectors $\vec{x}^{(i)} = \left(x_1^{(i)}, \dots, x_n^{(i)}\right)$ for $i = \overline{1,N}$ in the common space $\mathscr{A}$ of the dimensionality $n = \left|\bigcap_{i=1}^{N} A_{f^{(i)}}\right|$.

[4] In our previous works, we used the term *discrimination criterion* inspired by Šefránek (2002). We have changed it, because the activation function actually returns the degree of identification of its input with the represented category (see the discussion in Section 2.4).

In line with Gärdenfors (2000), the geometric centroid computed as the mean vector of the sample set

$$\vec{p} = \frac{1}{N}\sum_{i=1}^{N}\vec{x}^{(i)}.$$

will represent a prototype of the category.

The membership of a perceptual frame f in the category represented by a locally tuned detector $r_{\vec{p}}$ will be evaluated as an exponentially decaying function of the distance from the prototype (Shepard, 1987)

$$r_{\vec{p}}(\vec{x}) = \exp(-k \cdot d(\vec{p},\vec{x})), \tag{1}$$

where k is some positive constant, d is some metric and $\vec{x}$ is a projection of the frame f into $\mathscr{A}$ (if f cannot be projected because it lacks some attributes from $\mathscr{A}$, the detector returns 0).

The shape of the receptive field $\Psi_\theta(r) = \{\vec{x} \in \mathscr{A} \mid r(\vec{x}) > \theta\}$ depends on the metric d. In the simplest case of Euclidean metric d_{L_2}, the receptive fields of all detectors are hyperspheres in $\mathscr{A}$ centered at $\vec{p}$ and with the same radius determined by θ. Hence, they have the same shape regardless of the distribution of values in their sample sets. This may be undesirable.

In the original theory of conceptual spaces (Gärdenfors, 2000), the use of weighted Euclidean metric is suggested to express unequal importance of dimensions depending on the context or shifts of attention. In our approach, each detector uses *its own* metric derived from the properties of its sample set, instead of a common metric. Local metrics make similarity judgments essentially asymmetric. In general, for two detectors centered in $\vec{p_1}$ and $\vec{p_2}$, the value of $r_{\vec{p_1}}(\vec{p_2})$ does not have to be equal to $r_{\vec{p_2}}(\vec{p_1})$. People show the same effect, e.g. Tel Aviv is judged more similar to New York than vice versa (Tversky, 1977).

Now we review several metrics and their effect on the representational power of the detectors. Euclidean metric weighted by the inverse of the common variance σ^2 of values of all attributes in the sample set enables representing categories with different levels of generality (hyperspheres with different radii).[5] Normalized Euclidean metric with differences on each dimension weighted by the inverse of the variance of sample values on that dimension makes the detector sensitive to the unequal importance of attributes for the category membership (for technical details, see Appendix A). This is very important for cross-situational disambiguation of meanings. The attributes with nearly the same value in all examples will be considered more important for category membership than attributes with big variances within the sample set. The value of an attribute with zero variance will become mandatory for the category instances (any other value in the input frame would yield zero activity of the detector).[6] The receptive fields of the detectors are $\vec{p}$-centered n-dimensional hyperellipses having axes parallel with those of the input space $\mathscr{A}$ (see Fig. 2). In case of a mandatory attribute value, the receptive field is a degenerate ellipsoid with the corresponding axis of zero length.

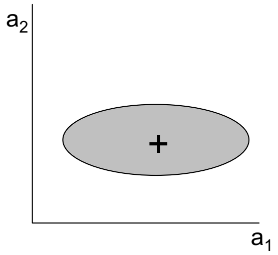

Fig. 2. Variance-based detectors can account for unequal importance of attributes; the hyperelliptic receptive field has a longer axis along the dimension a_1 because of the bigger variance of a_1 values in the sample.

The proposed variance-based detectors can learn to attend to differences in some attributes more than in others. However, they cannot learn correlations between attributes, while people can do so (Medin, Altom, Edelson, & Freko, 1982). For example, to induce the concept of *square* from example frames containing the attributes *vertices*, *sizeX*, *sizeY*, one must not only learn the mandatory value 4 of the attribute *vertices*, but also learn that values of attributes *sizeX* and *sizeY* should be identical. This can be achieved by a detector using squared Mahalanobis distance

$$d^2_{\boldsymbol{\Sigma}^{-1}}(\vec{p},\vec{x}) = (\vec{x}-\vec{p})^\top \boldsymbol{\Sigma}^{-1}(\vec{x}-\vec{p}),$$

where $\vec{p}$ and $\vec{x}$ are column vectors and $\boldsymbol{\Sigma}^{-1}$ is the inverse of the covariance matrix of the sample set (see Appendix B).

As the examples of a category are not usually given all at once, but come sequentially one by one, the mean and the covariance matrix are continuously recomputed by iterative formulas (see Appendix C). The attribute set determining the subspace $\mathscr{A}$ is updated iteratively, too, by intersecting with the attribute set of each new example. If some attributes are removed from the subspace $\mathscr{A}$ this way, the corresponding rows and columns of the covariance matrix and the mean vector are removed, too.

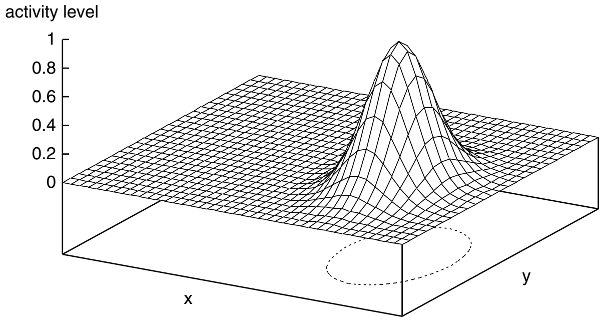

Fig. 3. A 2-dimensional locally tuned detector with the multivariate Gaussian activity curve. The detector's receptive field with the threshold $\theta = 0.1$ is shown in (x, y) plane.

[5] If we allow infinite weights and define $\infty \cdot 0 = 0$, a category with zero variance will have 1-point receptive field in $\vec{p}$ and will represent an individual.

[6] For example, triangles can have various sizes, positions, orientations, etc., but they all must have 3 vertices.

By setting $k = \frac{1}{2}$ and using squared Mahalanobis distance in Eq. (1), we get a detector with (not normalized) multivariate Gaussian tuning curve (see Fig. 3) and the receptive field with the rotated hyperelliptic shape.

Covariance-based detectors also support dealing with mandatory attributes, ignoring irrelevant ones and generalization by capturing the most important features (for technical details, see Appendices B and D).

3.2.3. Sign pattern based detectors

In everyday reasoning, people often abstract away from numerical values and use a simpler qualitative calculus based on ordinal relations and invariant sign patterns (Kuipers, 1994). The sign structure of attributes is often constitutive for relational concepts, e.g. the relation $bigger(f_x, f_y)$ can be expressed as $f_x.size > f_y.size$, or equivalently, $\text{sgn}(f_x.size - f_y.size) = +1$.

Now we define a qualitative detector sensitive to the sign pattern of attributes in the sample set. The subspace $\mathscr{A}$ is defined by a set of attributes present in all examples with the same sign. The projection of frames to $\mathscr{A}$ is composed with the operator sgn. The sign pattern is recorded in the prototype

$$\vec{p} = (p_1, \ldots, p_n), \text{where } p_j = \text{sgn}\left(x_j^{(i)}\right) \ \forall\, i = \overline{1, N}.$$

The sign pattern is recorded only once, upon seeing the first example. Later updates of the criterion only remove from $\mathscr{A}$ the attributes not occurring in new examples with the same sign as recorded. The detector returns a binary result: 1, if an input frame has the same sign pattern as $\vec{p}$ for all attributes in $\mathscr{A}$, and 0 otherwise.[7]

3.2.4. Identification criteria based on locally tuned detectors

Categories of individual objects, classes and properties of objects can be directly represented by locally tuned detectors (as their argument is one frame describing an object). Criteria having more input arguments (relations, changes) can be reduced to locally tuned detectors by transforming their input. Binary relational criteria, e.g. $larger(f_1, f_2)$ for frames f_1 and f_2, can be represented by detectors operating on a transformed input $\Delta(f_1, f_2)$, where $\Delta(f_1, f_2) = f$ is a frame of differences of aligned attributes (Markman & Gentner, 1993), defined by

$$A_f = A_{f_1} \cap A_{f_2} \text{ and } f.a = f_1.a - f_2.a \quad \forall\, a \in A_f.$$

Complex situations or properties of the whole scenes can be built as hierarchical networks of locally tuned detectors. Detectors of the bottom level operating on perceptual frames represent components of the situation and their required mutual relations. Vectors of output activities of the elementary detectors serve as an input to aggregate detectors of the higher level (Takáč, 2007), which can attribute unequal importance to the elementary detectors and/or detect their mutual correlations.

3.2.5. Representation of environmental dynamics and verb semantics

Criteria expressing changes of properties of an object in time are relational criteria applied across time to frames $f^{(t)}, f^{(t-1)}$ of the same object. They can be represented by detectors applied to a transformed input $\Delta f_{t,t-1} = \Delta(f^{(t)}, f^{(t-1)})$.

Environmental dynamics expressed by the change criteria is an important part of meanings of verbs. Some changes can be captured by qualitative relations, e.g. *grow*(*f*) can be expressed by $\text{sgn}(\Delta f_{t,t-1}.size) = +1$, others require encoding of a typical change's magnitude, e.g. movement criteria for *crawl*, *walk*, *run* could differ in mean values of $\Delta f_{t,t-1}.position$. The criteria with zero sign pattern of some attributes can represent a state or a persistence of a property, e.g. *stay*.

Of course, this is just a part of the picture. Embodied verb representation is connected to actions and includes the manner of performance, e.g. the representation of *jumping* can refer to a non-declarative procedural representation of an invariant motor stereotype together with a frame representing variable parameters of the action, e.g. the velocity or joint angles (Bailey, Feldman, Narayanan, & Lakoff, 1997).

As the embodied meanings are grounded in sensorimotor interactions with the environment, they also include situated causal knowledge about preconditions of successful actions and their possible consequences. For example, the action of *lifting* performed in the same manner (with the same force) can lead to different outcomes (changes) depending on the object of the action, e.g. lifting a ball or lifting a 200 kg piece of furniture. Such propositional knowledge can be suitably represented by cross-categorical associations of the type

$$(preconditions,\ action\ \rightarrow\ consequence),$$

where *preconditions* are criteria for objects of the action, *action* is a criterion representing the action's manner, and *consequences* are change criteria of the resulting dynamics. A BDI[8] agent can represent its goals by criteria of desired situations and can plan sequences of actions presumably leading from the current situation to a desired one.

4. Construction of environmentally and ecologically relevant meanings

In the course of time, the agent perceives a mixed sequence of instances of many concepts and it must somehow determine which of the existing criteria should be updated by an incoming example (and when to create a new criterion). If the agent has no additional information,

[7] The sign pattern based detector can be viewed as a special case of the variance-based detector with all attributes mandatory (as their signs have zero variance).

[8] Belief–Desire–Intention (Bratman, 1987).

it can group similar frames to categories by unsupervised clustering techniques, e.g. *distributed clustering algorithm* (Hulth & Grenholm, 1998), *1-nearest neighbor*, or others (Everitt, Landau, & Leese, 2001), maximizing inter-cluster and minimizing intra-cluster distances.

4.1. The model

Construction of ecologically relevant categories must be based on pragmatic criteria. Now we review our experiment focusing on construction of ecologically relevant categories by sensorimotor interactions with the environment (Takáč, 2006a). In the experiment, the simulated environment (2-dimensional lattice) contained frames of toys, fruits, furniture, and one agent. The agent was actively exploring its environment. In each time step, it could randomly choose an action from its action repertoire (lift, put down) and perform it with different parameters (force, arm angle) upon an object randomly chosen from its surrounding. The effects of the action on the chosen object were simulated by the environment. In the case of lifting, the vertical position of the object was increased by the value proportional to the arm angle, if the exerted force was greater than the weight of the object, otherwise the action had no effect. In the case of putting down, the vertical position of the object was decreased by the value proportional to the arm angle, unless the object was already on the ground.

After performing an action, the agent observed the resulting changes and represented the knowledge of causal relations between actions, objects and changes in the form of associations among their respective categories (see the previous section). Categories of each type were represented by identification criteria with variance-based detectors, stored in three separate categorical systems. Categories were constructed from scratch, i.e. the agent started with the empty categorical systems. Objects and actions were grouped to categories by the change. That is, if an action led to the same change on several objects, they would all fall in the same object category (and be in the sample set that updates the same detector) and vice versa. A significantly[9] different outcome of the action triggered creation of new categories.[10]

4.2. The results

We let the agent interact with its environment for 5000 time steps. To evaluate the usefulness and adequateness of the agent's representation, we measured it's ability to predict the correct result of the action. After choosing an object and an action, the agent predicted the resulting change. The resulting activity of the predicted change criterion applied to the actually perceived change was recorded as the *prediction*. However, the more general change criteria give higher similarity values, thus we also measured the *generality* of the prediction expressed by the average standard deviation of attributes of the criterion used for prediction (a lower value means higher accuracy of the prediction). We have also measured the number of criteria in the agent's representation.

The results are summarized in Fig. 4. As we can see in the graph, while the prediction change 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. This value corresponds to the average distance σ, which is an average intra-cluster distance of the category. Hence, this means that the criteria give correct predictions and the agent has constructed ecologically relevant categories.

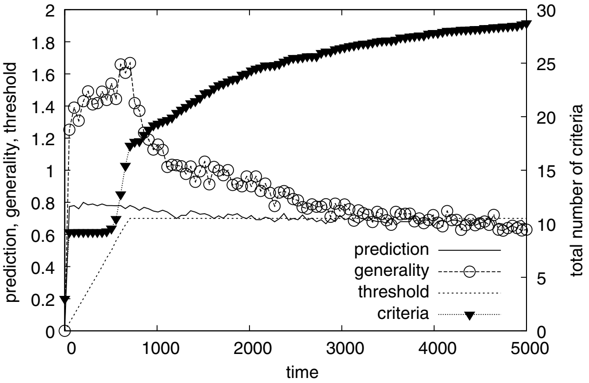

Fig. 4. Results of the experiment simulating construction of categories by sensorimotor interactions. The number of criteria saturates and the prediction value converges to that of the average intra-category distance (i.e. predictions are correct).

Table 1
A fragment of the agent's associations of categories

Category	Action	
	putDown(5 ± 3)	*liftUp*$(6 \pm 2, 10 \pm 6)$
C1	No change	
C2		No change
C3	$\Delta = \{posZ : -6 \pm 1\}$	$\Delta = \{posZ : 7 \pm 1\}$
C4	$\Delta = \{posZ : -4 \pm 2\}$	$\Delta = \{posZ : 5 \pm 2\}$

Object categories are in rows, action categories are in columns, associated change criteria are in intersections of rows and columns. Arguments of actions are *putDown*(*armAngle*) and *liftUp*(*armAngle*, *force*). Change criteria express the difference in vertical position of the involved object (for brevity, the attribute values are written as the mean ± the standard deviation σ).

[9] In order to model developmentally growing sensitivity to environmental differences, we used detectors with linearly increasing decision threshold θ for change criteria.

[10] Details of the algorithm can be found in the original paper (Takáč, 2006a).

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

Object type	Category			
	C1	C2	C3	C4
Agent	1			
Fruit			8	2
Toy		1	3	6
Furniture	5	5		

Actual meanings of the constructed categories can be guessed by inspecting the internal representation of the agent. In Table 1, we can see a fragment of the agent's representation from an example run of the experiment. Table 2 shows the object criteria applied to 31 objects in the environment. Numbers in a row express the object counts of a given type most similar to the criterion in a column. We can see that the agent constructed categories such as "objects too heavy to be lifted" (C2) or "objects that cannot be put down, because they are already on the ground" (C1). Category C3 represents mostly fruits and C4 mostly toys.

5. Construction of socially relevant meanings

The experiment described in the previous section modeled individual construction of preverbal meanings. In this section, we study the influence of verbal instruction (naming) on category formation process, in line with empirical observations of Waxman and Braun (2005). Also, we test the Sapir-Whorf hypothesis (Whorf, 1956) stating that language alone can lead to construction of relevant categories.

5.1. The model

The model consisted of two agents situated in a simulated environment: a teacher describing various aspects of the present situation, and a learner inducing meanings of the teacher's words by noticing cross-situation similarities between their referents.

The induction of meanings was guided by *no true synonymy* and *no true homonymy* assumptions (Takáč, 2006b). The former assumed that different words had different meanings, even if they shared a referent (in that case they expressed different aspects of the referent).[11] The latter assumption was crucial for cross-situational disambiguation of the meaning: all referents of a single word across different situations were considered instances of the same category denoted by the word. The more contexts of the word's use, the bigger the probability that the referents would vary in the properties irrelevant for the meaning of the word.

The simulated environment consisted of 2D geometrical shapes characterized by five attributes: the number of vertices ($vertices \in [2, 5]$), coordinates of the centroid of the shape ($posX, posY \in [0, 50]$) and the size of the bounding rectangle ($sizeX, sizeY \in [0, 25]$). The initial values of attributes were uniformly randomly generated integer numbers from the respective intervals. The environment was dynamic in that in each time step, randomly selected objects could be resized, moved, or removed from the environment and newly generated objects could be added (there were usually 2–4 objects simultaneously present on the scene). Multiple changes could happen simultaneously in one time step. In each time step, the teacher, using the predefined ontology and the lexicon, commented the current scene (including the changes) to the learner. The teacher's lexicon included 2 nouns, 3 adjectives and 2 verbs (see Table 3).

Table 3
A predefined ontology and lexicon of the teacher

Word	Meaning
Square	$vertices = 4 \wedge sizeX = sizeY$
Triangle	$vertices = 3$
Big	$sizeX > 15 \wedge sizeY > 15$
Slim	$sizeX < 0.2 sizeY$
Small	$sizeX < 10 \wedge sizeY < 10$
Grow	$sizeX^{(t)} > sizeX^{(t-1)} \wedge sizeY^{(t)} > sizeY^{(t-1)}$
Shrink	$sizeX^{(t)} < sizeX^{(t-1)} \wedge sizeY^{(t)} < sizeY^{(t-1)}$

The experiment was run for 5000 learning epochs (time steps). The learner used detectors based on the Mahalanobis metric, with the receptive field threshold $\theta = 0.1$ and SVD-filtering with the threshold $b = 10\%$ (see Appendix D).

5.2. The results

To evaluate the fidelity of meaning transmission, before receiving the scene description from the teacher, the learner described a scene in each time step too, and the two descriptions were compared for the *correctness* and the *completeness* of the learner's description. The measure of the *description similarity* was the average of the correctness and the completeness.

The *correctness* of the learner's description was computed as $1 - w/L$, where w was the number of wrong words in the learner's description of the scene and L was the total number of words in the learner's description. A word in a learner's utterance describing some referent was considered *wrong*, if it was not used by the teacher in its utterance describing the referent.

The *completeness* of the learner's description was computed as $1 - m/T$, where m was the number of teacher's words missing in the learner's description of the scene and T was the total number of words in the teacher's description. A word in a teacher's utterance describing some referent was considered *missing*, if it was not used by the learner in its utterance describing the referent.

We also evaluated a pragmatic *usage* of the learner's ontology in guessing games. In each time step, the teacher

[11] This assumption corresponds to the *Principle of Contrast* (Clark, 1987).

uttered a verbal description of a referent randomly picked up from the scene and the learner guessed the referent. The learner's guess was a set $\mathscr{L}$ of possible referents of the utterance, as understood by the learner. In case $\mathscr{L}$ did not contain the referent meant by the teacher, the usage was zero. Otherwise, the success in the guessing game was evaluated by comparing $\mathscr{L}$ with the set $\mathscr{T}$ of referents that the teacher itself would guess from the utterance (as the agents did not play discrimination games in our model, the teacher's description did not have to be unique either). The usage was computed as $1/(1+r)$, where r was the number of referents in $\mathscr{L} - \mathscr{T}$. Hence, even in the case of a correct guess, the usage was lowered by any extra referents that could not be meant by the teacher.

The *uncertainty* inherent in the teacher's descriptions was measured as $1 - 1/|\mathscr{T}|$. For example, if the teacher described the chosen object by the utterance "triangle", and the scene contained two objects categorized as triangles by the teacher, the uncertainty of the teacher's description would be 50%. If the teacher's utterance had a unique referent, uncertainty would be zero.

Besides playing guessing games during the learning, the agents played 200 guessing games after every 500 learning epochs. The guessing games were only played for evaluation purposes and did not have any influence on the learning process.

The simulation results (Fig. 5) show that cross-situational learning is fast and reliable: the learner soon acquires an ontology and lexicon of a sufficient quality, which remains stable for the rest of the simulation.

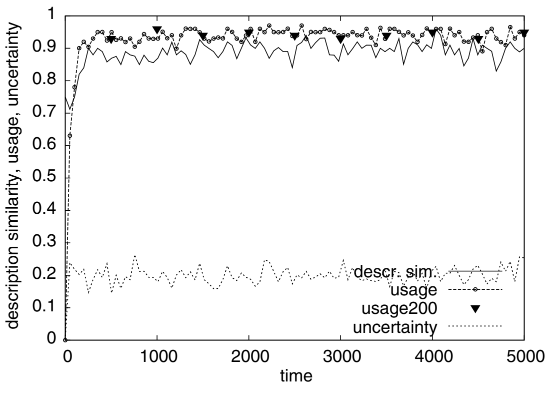

Fig. 5. The quality of the lexicon acquired by cross-situational learning within one generation. The measure "usage200" is an average usage of 200 guessing games played after learning. Each measure in the graph has been averaged over the time window of 30 last steps. The results of the experiment were averaged over 10 simulation runs with different random seeds.

6. Evolution of meanings within iterated intergenerational transmission

We have shown how a learner can construct meanings sufficiently similar to those of its teacher by verbal instruction. Now the question is, whether meanings constructed this way remain stable, if we let the acquisition process iterate intergenerationally. In this section, we describe an extension of the previous experiment, based on the iterated learning model (ILM) framework (Kirby & Hurford, 2001), designed to study how meanings change across generations.

The iterated learning model, which involves vertical cultural transmission of language between generations, has primarily been designed for modeling the emergence of grammar. In the ILM framework, language develops by flowing between two forms of private language competence and externalized utterances by processes of acquisition and production. A learner builds up its own internal language representation by observing external language input from its teacher, later the learner becomes a teacher and produces utterances, which are the input for the next generation learner, etc.

6.1. The model

The first generation setting in the model was identical to the experiment of the previous section, except that we varied the number of learning epochs. After a certain number of epochs, the teacher with predefined ontology was removed and the learner became a teacher for a new agent with an empty ontology and lexicon. We let this process iterate for 50 generations. We ran two versions of the experiment: in Experiment 1, the agent could neither modify nor add any new meanings, once it became a teacher (it only used the meanings acquired from it's own teacher). In Experiment 2, the teacher could invent new meanings or extend old ones, in case it had no meanings applicable to describe some object on the current scene.

6.2. Experiment 1

Fig. 6 shows the results of Experiment 1 run for 50 iterations of 500 learning epochs. We measured the description similarity and the pragmatic usage of the learner's lexicon in each generation. As we can see, the quality of the lexicon is quickly retained after each drop caused by a generation exchange.

To explore intergenerational meaning shifts, we also inspected the agents' internal representations of categories. We found out that categories *grow* and *shrink* represented by sign pattern based detectors remained the same in all generations. The agents were also successful in inducing the correct criteria for *triangle* (3 vertices, all other attributes irrelevant) and *square* (4 vertices and equal side lengths), which remained the same in all generations, too. Though the exact parameters of the criterion for *slim* varied across generations, the property of having horizontal size small in comparison to the vertical one has been correctly captured and retained. Criteria for *big* and *small*, based on certain intervals of uncorrelated attribute values, did not turn out to be so stable. Either they were overspecialized in some simulation runs and they died out (their receptive

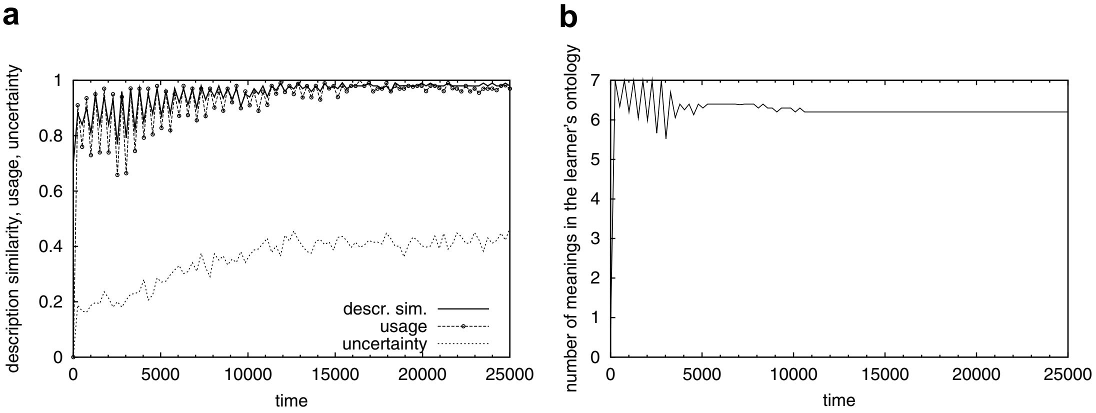

Fig. 6. The quality of the learner's lexicon in iterated learning setting. Generations exchanged every 500 time steps. The results were averaged over 10 simulation runs with different random seeds. (a) The description similarity and the pragmatic usage quickly stabilized at a very high value close to 1, at the cost of a higher uncertainty caused by overgeneralization of some meanings. (b) The number of meanings stabilized on the value of 6 – one meaning died out because of overspecialization.

field gradually shrank to zero), or they were overgeneralized (due to a takeover of some attributes, see Fig. 7).

In order to explore causes of this instability, we varied the number of learning epochs in each generation (Fig. 8). We can see that a smaller number of learning epochs causes smaller sample sets. If sample sets are too small, concepts are unstable, some of them get overspecialized and disappear (the number of total meanings gets smaller), others get overgeneralized (the uncertainty rises). This is the case of simulations with less than 300 learning epochs per generation (corresponding to less than 25 examples for the detector with the smallest sample set). In simulations with more than 300 learning epochs, the average number of meanings stays between 6 and 7 and the uncertainty is around 30–35%. These results are further discussed in the following sections.

6.3. The influence of the meaning bottleneck

As objects and their changes are generated randomly within the fixed number of learning epochs, the sample size for a learner's category depends on the probability of

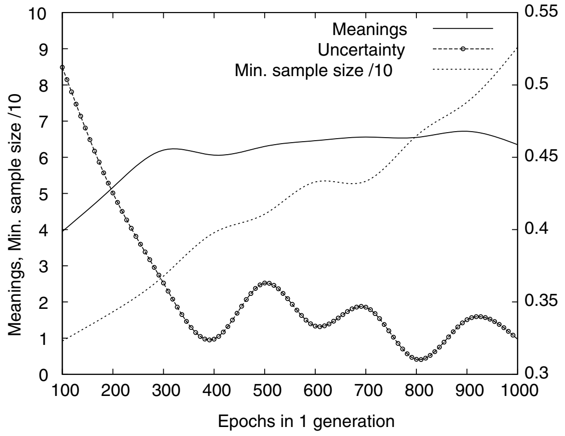

Fig. 8. The influence of the number of learning epochs per generation on the stability of meanings. Data for each number of learning epochs are averaged over 10 simulation runs with different random seeds. "Meanings" is the total number of the learner's meanings, averaged over all learning epochs in all generations. "Min. sample size/10" is the size of the sample size of the learner's criterion with the smallest sample set, averaged over all learning epochs in all generations (and scaled by 10). "Uncertainty" expresses the referential ambiguity of the teacher's descriptions (cf. Section 5.2), averaged in the same way as the two previous measures.

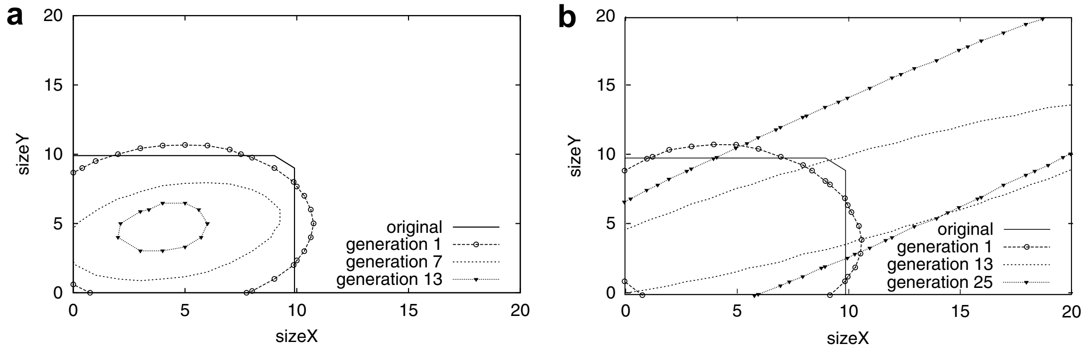

Fig. 7. Categories unstable over generations. The receptive field of each category was projected into the plane with dimensions *sizeX*, *sizeY*. The projecting plane crossed other dimensions in mean values of the dimensions recorded in the category's detector. In some simulation runs, the categories *big* and *small* showed one of the following behaviors: (a) overspecialization – the size of the receptive field converged to zero over generations, (b) overgeneralization – a random correlation of some attributes in the sample overtook other attributes that became overgeneralized (ignored).

occurrence of instances of the teacher's category on the scene. This creates an implicit meaning bottleneck. In iterated learning models of grammar emergence, the learning bottleneck leads to the emergence of compositionality, because compositional rules are more likely to be transmitted through the bottleneck (Kirby & Hurford, 2001; Vogt, 2005). In our model, instances of more general categories are more likely to appear on the scene within the learning period than those of very specific categories or even categories representing individual objects. Also, our model shows the same frequency effects as those reported by Kirby and Hurford (2001): if examples of a very specific category appear often on the scene because of a biased random generator, they can get preserved over generations, otherwise they would probably die out.

Catching and amplifying randomly occurred regularities is the inherent property of iterated learning. While this property is desirable for the emergence of grammar, it can lead to distortion of meanings in our model. The smaller the sample, the bigger the chance that it will contain random correlations that are not a part of the original meaning and a covariance-based detector would not reconstruct the original meaning properly. Once a random correlation becomes a part of the meaning, it gets reinforced in the next generation, because the teacher will pick up as instances of the category only examples containing the correlation. This way the meaning gets overspecialized (see Fig. 7a). Overspecialization is dumped by SVD-filtering (see Appendix D) that captures the properties most invariant in the sample and filters out the others. However, a random invariance in a small sample can lead to overgeneralization due to truncating some relevant attributes (see Fig. 7b).

Meaning transmission can be viewed as an evolutionary process with meanings as competing replicators. The selection pressure is imposed by the meaning bottleneck. Meanings pass through the bottleneck, if they are relevant to the environment. Special meanings describing situations that occur very rarely have smaller survival chances than frequently applicable general meanings. However, the model used in Experiment 1 corresponds with the replicator view only partially: meanings can die out, if they are no longer relevant, but there is no mechanism of creation of new meanings in the model. If the agents were suddenly relocated to a completely different environment, the teacher would remain silent because of the lack of adequate meanings, and the whole language would die out. To make the model more realistic, the teacher was allowed to coin new meanings and words in such situations. The experiment is described in the following section.

6.4. Experiment 2

The model setting was similar to that of Experiment 1, in that the teacher had to describe a randomly generated and modified scene to the learner. The teacher of the first generation started with the criteria for *triangle*, *square*, *small*, *big*, and *slim* (see Table 3). If, for some object on the scene, there was no criterion with an above-threshold activity, the object was approximately described by the word associated with a criterion returning the highest non-zero activity. If there was no criterion returning non-zero activity for the object, a new criterion (named by a new random word) was created with the object as the first example.

The learning process was iterated for 50 generations with 200, 500, or 1000 learning epochs in one generation. The results we got confirmed our replicator hypothesis. The meaning of *triangle* was stable and remained the same throughout all generations in all three versions of the experiment. The meaning of *square* remained stable in the experiments with 500 and 1000 learning epochs and died out in some simulation runs of the experiment with 200 learning epochs. In comparison to *triangle* and *square*, the criteria for *small*, *big*, and *slim* were more likely to return a non-zero activity for a random object. Hence, they were more often used in an approximate sense, which caused their instability.

Using criteria in approximate senses caused the extension of their receptive fields in the next generation. Meanings with under-threshold activity competed for selection and the meaning with highest activity was selected and extended subsequently. This created *rich-get-richer* dynamics (a positive feedback loop) and led to the formation of very general criteria. Indeed, in every version of the experiment, we observed the appearance and survival of general criteria[12] with meanings such as "objects with 2–5 vertices" or "objects with horizontal position *posX* between 0 and 50" applicable to all referents in all situations. Other newly created meanings defeated in the competition had very small sample sets and have not survived.

7. Discussion

7.1. Mechanisms of meaning construction

We have proposed and explored several models of meaning construction. In the model of individual meaning construction, we have shown how ecologically relevant categories can be constructed from sensorimotor interactions with the environment. The resulting categories reflected the structure and dynamics of the environment. In accordance with Rosch (1978), objects were grouped to categories by common interaction programs. All action categories associated with some object category represented *affordances* of the object, i.e. the perceivable possibilities for acting on the object (Gibson, 1979). On the other hand, object and change categories associated with action categories can be viewed as verb islands in line with the *verb island hypothesis* (Tomasello, 1992) stating that the

[12] These included overgeneralized meanings of *small*, *big*, and *slim* as well as some newly created criteria.

early more complex lexical constructions of children are organized in verb-centered structures with verb-specific arguments. The adequateness of the constructed representation was proved by the agent's ability to use it for predicting the results of its actions on objects.

Children learn by interacting with the world, but, at the same time, they are exposed to the linguistic production of their caregivers. Naming influences the children's conceptual organization and supports discovery of novel concepts (Waxman & Braun, 2005). We have shown how a linguistic instruction accompanied by a non-verbal reference can lead to cross-situational construction of the learner's meanings, which are similar enough to those of the teacher, to be used for pragmatic purposes. A high similarity has been achieved very rapidly, which is in line with the observed phenomenon of *fast mapping* (Carey & Bartlett, 1978).

However, meanings do not stay intact, if we let the acquisition process iterate. While a high similarity between teacher's and learner's meanings is maintained within each generation, meanings do change throughout the generations. These results suggest how real languages can change historically, while still preserving their communicative function.

Meanings constituted by simple structural relations and invariant attribute values have shown to be more stable in the iterated transmission than meanings based on interval values of uncorrelated attributes. In our model, we used adjectives *big* and *small* with the meaning of having the size bigger or smaller than a fixed value. This is not realistic: most adjectives are semantically dependent on the nouns they modify (Warren, 1988), e.g. an adjective *big* refers to very different absolute measures in the phrases "big mouse" and "big elephant". The meanings of such adjectives are constituted by structural relations that are mapped onto a particular domain generated by the modified noun. We can speculate that the semantic dependency of adjectives, observed in real languages, is the result of the dynamics of the selection process within the iterated language transmission, where meanings based on structural relations are much more persistent.

If we allow extension of meanings to novel referents, the dynamics of the iterated transmission inevitably leads to the erosion toward more and more general meanings. This phenomenon was also observed in other models (Smith, 2001, 2005a). The occurrence of the meaning drift in our model can be explained by the lack of other selection forces, as the meanings were only optimized for their expressive coverage. In real situations, utterances and their meanings serve pragmatic purposes including identification and discrimination. Hence, optimal meanings should reflect the trade-off between coverage and distinctiveness (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976).

To explore the nature of meaning creation mechanisms, we have deliberately studied each of them in isolation. However, in real situations, meaning formation processes are coupled and interact with each other.

7.2. Related works

There are lot of models related to our work in different aspects. The similarities and differences can be analyzed on the level of representation, learning mechanisms and the overall dynamics.

7.2.1. Representation

In the models of language bootstrapping, meanings are usually represented by collections of intervals from discrimination trees (Smith, 2005a; Steels, 2000), prototypes in one-dimensional (Vogt & Divina, 2007) or multidimensional (Vogt, 2005) conceptual spaces (Gärdenfors, 2000), adaptive networks (Steels & Belpaeme, 2005), or weight configurations in artificial neural networks (Borghi, Parisi, & Di Ferdinando, 2005; Cangelosi, 2005).

Nodes of a discrimination tree represent features – subintervals of the range of a particular sensory channel. The initial range $[0, 1]$ is adaptively refined, based on the results of discrimination games. Hence, meanings are related to particular sensory channels (attributes) and correspond to locally tuned detectors evaluating the similarity in one-dimensional subspace (based on one attribute) in our model. In some models, if a single feature cannot identify a topic referent uniquely, a set of features is chosen. However, the discrimination trees are mutually independent and the construction of the feature set is situational and not persistent. Hence, discrimination trees have smaller expressive power, because they cannot persistently represent meanings based on correlations of different attributes.

Prototypes in conceptual spaces have the potential to capture correlations of attributes. However, they are used in a different way in the above-mentioned models. Prototypes in one-dimensional spaces related to perceptual features have the same expressive power as features in discrimination trees. Categories in multidimensional space are constructed by placing prototypes on each dimension separately and combining them together (Vogt, 2005). A category of an object is constructed as a vector of the closest prototypes on each dimension and the prototypes generate a grid in the conceptual space. Unlike in our model, the density of each dimension is the same for all categories. Hence, the representation is not sensitive to unequal importance of dimensions for a category. Categories based on correlated attributes could in principle be represented as multiple nodes of the grid, but this is not used in the model.

Adaptive networks used by Steels and Belpaeme (2005) are most closely related to our representation. An adaptive network consists of a set of locally reactive units, each with a Gaussian activation function centered at some point (widths of all Gaussians are the same and fixed to some constant). The resulting activity of the network is a weighted linear combination of activities of all reactive units applied to a common input. Each category is represented by one adaptive network; an input is categorized as a member of the category represented by the network giving the highest activity.

Adaptive networks can be considered variants of Radial Basis Function Networks (RBFN, Poggio & Girossi, 1990b). Locally reactive units of RBFN correspond to our locally tuned detectors with the Euclidean metrics $d_{L_2,\sigma}$ (see Appendix A).[13] A single reactive unit cannot capture correlations of attributes, but the whole network can. It has been proved that RBFN is a universal approximator in that it can approximate any multivariate continuous function, given a sufficient number of reactive units (Poggio & Girossi, 1990a). Adaptive networks and RBFN represent categories with fuzzy boundaries. Receptive fields of the networks do not have to be convex; they can even consist of disconnected regions.

7.2.2. Learning mechanisms

The adaptive networks are not trained, but adapted by adding or removing a locally reactive unit and by changing weights of the units. The adaptation of networks is guided by their success or failure in discrimination games and language games. Other models of Steels' group (for overview see Steels, 2000) are based on the same principle of learning from feedback about success or failure in various games.

Learning in our models is based on extracting cross-situational similarities between examples of a category. Input frames are considered examples of the same category, if they can be interacted with in the same way (in the case of individual learning), or if they were named by the same word (in case of social learning). A seminal model of cross-situational learning was published by Siskind (1996). In this model, the algorithm learns mappings between word symbols and conceptual symbols such as GO, *John*, *ball*. A hypothesis set of all possible conceptual symbols is given in advance, and "lexical acquisition is simply a process of learning the mapping between two pre-existing mental representation languages" (p. 47, Siskind, 1996). In this aspect, we can view our model as cross-situational learning of meanings of Siskind's atomic conceptual symbols, while learning in the model of Siskind works more on the sentence level by eliminating meaning mappings incoherent across situations. Siskind's model deals with the referential indeterminacy, noise and homonymy by employing the mutual exclusivity assumption (Markman, 1992). Cross-situational learning is also used in the model of Smith (2005a). Semantic hypotheses (represented by nodes of discrimination trees) are not given in advance, but are constructed by playing discrimination games prior to the language acquisition phase. In the language acquisition phase, agents learn mappings between the constructed meanings and words from word-meaning co-occurrence frequencies. Mathematical properties of cross-situational learning are analyzed by Smith, Smith, Blythe, and Vogt (2006). Cross-situational learning is combined with learning based on corrective feedback in the model of Divina and Vogt (2006) and Vogt and Divina (2007).

Among recent connectionist models of action-based categorization, that of Borghi et al. (2005) seems to be most closely related to our model of individual interactionist meaning construction. In their model, an organism with a visual system and a two-segment arm (simulated by a neural network) reaches different points in space, depending on the object seen and on the context. Constructed categories reflect characteristics of the output actions to be performed rather than perceptual characteristics of the input. 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.

The role of social learning in the acquisition of concepts and language has been studied by Steels and Kaplan (2001a). In their experiment, a human teacher (mediator) interacted with a Sony AIBO robot, trying to teach it the names of three objects. The perceptual input of the robot was in the form of camera images taken from different angles and under different light conditions. The robot used simple instance-based learning and this was compared with unsupervised clustering techniques. The experiment has shown that categories obtained by unsupervised techniques did not match the three objects, while categorization directed by naming has been much more successful. Active and rich social interactions served the role of narrowing the context and reducing the noise. If the teacher just estimated at which object the robot was looking, and uttered a name for it, the input was much more noisy and the performance deteriorated. The learning algorithm used in the experiment was not sensitive to unequal importance of dimensions. The authors state that if there were more objects on the scene represented by perceptual images in more dimensional space, methods for computing correlations of dimensions with categories and the intercorrelations among dimensions should be employed (Steels & Kaplan, 2001a, p. 26).

7.2.3. Iterated learning

In the paradigmatic iterated learning model (Kirby & Hurford, 2001) focused on the emergence of compositional structures on the syntax level, meanings were predefined and artificial structures. This has been refined in the iterated learning model of Vogt (2005), in which meanings were created in discrimination games. Cross-situational learning of meanings was combined with iterated vertical transmission in the model of Smith (2005a). Though, in this model, meanings were created individually by each agent in discrimination games, the experiment led to results similar to ours (high intra-generational meaning similarity, decreasing intergenerational meaning persistence, and the drift toward more general meanings).

[13] Generalized RBFN (Poggio & Girossi, 1990b) use units with covariances, which are equivalent to the pseudoinverse version of our covariance-based locally tuned detectors.

7.3. Limits and future work

Our models have been simplified in many aspects. First, categories were constructed by taking into account attributes common to *all* examples. This approach works well for basic-level categories, but can be problematic for some superordinate categories. Moreover, instances of some concepts are related by *family resemblances* rather than by common properties of all members, as exemplified by Wittgenstein (1953) with the concept of a *game*. In its current form, the model cannot cope with homonyms and with noise. If a sample set erroneously contained an instance with a set of attributes completely different from other instances, it would result in a category with an empty attribute set. This could be amended by recording frequencies of attributes and by splitting concepts in case of homonymy detection.

Second, our models do not account for hierarchic and taxonomic relations that exist among real concepts. Third, in the model of social learning of concepts, an explicit reference to instances of the named category was given along with the linguistic input. We used this simplification deliberately, in order to show that even in the absence of *referent* indeterminacy, the learner has to solve *sense* indeterminacy, because different words can describe different aspects of the same (known) referent. However, in the later phases of the language acquisition process, the explicit reference could be substituted by the inference from the linguistic or pragmatic context.

Meaningful categories should be useful for the agent in achieving its goals (Nehaniv, 2000). In our action-based model, the agent had no goals and performed actions randomly. The next research step is to endow the agent with needs, need-driven goals and an action planning mechanism. Also, we plan to study the interplay of the meaning construction mechanisms in a model with coupled individual and social learning.

According to the syntactic bootstrapping theory (Lidz, Gleitman, & Gleitman, 2004), children acquiring a language use grammatical cues to constrain possible meanings of words. Incorporating some form of grammar into our model is a topic for future research.

8. Conclusion

In this article, we described a grounded cognitive semantics for representing concepts of objects, properties, relations, changes, complex situations and events, based on *identification criteria*. The identification criteria are constructed individually by each agent, based on interactions with the environment and other agents. Unlike in most of the related models, construction of criteria is based on cross-situational *similarities* among instances of a category rather than on *differences* between a chosen object and other objects present on the scene of communication. We argue that categories constructed for the purpose of identification rather than discrimination are more suitable for the detached use of language (talking about things not present here and now).

Learning in our models is incremental and permanent. The learning mechanism is sensitive to correlations of attributes of instances with categories and the intercorrelations among attributes. We have implemented and experimentally tested meaning construction by individual and social learning, and explored the dynamics of meanings in iterated intergenerational transmission.

We would like to emphasize that, in the presented models, categories are not given and interpreted by an external designer, but are constructed by and meaningful to the agents themselves. Such models can have important practical applications in the areas involving agents that need to coordinate their activities in unknown, dynamic and open environments. As all possible meanings cannot be anticipated in design-time, the agents' ability to acquire (and continuously reconstruct) relevant meanings is critical.

Acknowledgement

This research has been supported by VEGA grants 1/3105/06 and 1/0361/08.

Appendix A. Variance-based metrics

Euclidean metric weighted by the inverse of the common variance σ^2 of all attributes

$$d_{L_2,\sigma}(\vec{p},\vec{x}) = \sqrt{\sum_{i=1}^{n} \frac{1}{\sigma^2}(x_i - p_i)^2} = \frac{1}{\sigma}\, d_{L_2}(\vec{p},\vec{x})$$

leads to hyperspheric receptive fields with radii proportional to σ.

The natural extension of the previous case is to record variances individually for each dimension and use the weighted distance $d_{L_2,\vec{w}}$ with

$$\vec{w} = \left(\frac{1}{\sigma_1^2},\ldots,\frac{1}{\sigma_n^2}\right),$$

i.e. use the normalized Euclidean distance

$$d_{L_2,w}(\vec{p},\vec{x}) = \sqrt{\sum_{i=1}^{n} \frac{(x_i - p_i)^2}{\sigma_i^2}}.$$

Now the receptive fields of the detectors are $\vec{p}$-centered n-dimensional hyperellipses having axes of lengths proportional to σ_i. The axes are parallel with those of the input space $\mathscr{A}$.

Appendix B. Covariance-based metric

Let

$$\left\{\vec{x}^{(i)} = \left(x_1^{(i)},\ldots,x_n^{(i)}\right) | i = \overline{1,N}\right\}$$

be the sample set with the mean vector $\vec{p}$ and the square-symmetric $n \times n$ covariance matrix

$$\boldsymbol{\Sigma} = (\sigma_{ij})_{i,j=\overline{1,n}}, \text{ where}$$
$$\sigma_{ij} = \frac{1}{N-1}\sum_{k=1}^{N}(x_i^{(k)} - p_i)(x_j^{(k)} - p_j).$$

Because $\boldsymbol{\Sigma}$ is a square-symmetric positive semi-definite matrix, it can be decomposed to

$$\boldsymbol{\Sigma} = \mathbf{U}\mathbf{D}\mathbf{U}^{\top}, \tag{B.1}$$

where $\mathbf{U}$ is an orthonormal rotation matrix of eigenvectors and $\mathbf{D} = diag(\lambda_1, \ldots, \lambda_n)$ is a diagonal matrix of eigenvalues of $\boldsymbol{\Sigma}$ with $\lambda_1 \geqslant \cdots \geqslant \lambda_n \geqslant 0$ (Ientilucci, 2003). Then the squared Mahalanobis distance

$$\begin{aligned} d^2_{\boldsymbol{\Sigma}^{-1}}(\vec{p},\vec{x}) &= (\vec{x}-\vec{p})^{\top}\boldsymbol{\Sigma}^{-1}(\vec{x}-\vec{p}) \\ &= (\mathbf{U}^{\top}(\vec{x}-\vec{p}))^{\top}\mathbf{D}^{-1}(\mathbf{U}^{\top}(\vec{x}-\vec{p})). \end{aligned}$$

Rotation does not change the shape and the size of the receptive field, which are completely determined by the diagonal matrix D. Hence, the receptive field of a detector using the squared Mahalanobis distance will be a hyperellipse with axes of lengths proportional to λ_i and the orientation determined by the rotation matrix $\mathbf{U}^{\top}$. Variance-based metrics are special cases of Mahalanobis metric with diagonal covariance matrices $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$ and $\boldsymbol{\Sigma} = diag(\sigma_1^2, \ldots, \sigma_n^2)$, respectively.

B.1. Singular case

In case the covariance matrix is singular, hence non-invertible, the Moore–Penrose pseudoinverse $\boldsymbol{\Sigma}^{+}$ is often used instead of $\boldsymbol{\Sigma}^{-1}$. Computation of the pseudoinverse matrix is based on singular value decomposition (SVD) of the matrix $\boldsymbol{\Sigma}$, which takes the form (B.1). In case of a singular $\boldsymbol{\Sigma}$,

$$\mathbf{D} = diag(\lambda_1, \ldots, \lambda_k, 0, \ldots, 0) \text{ for some } k < n.$$

Then

$$\boldsymbol{\Sigma}^{+} = \mathbf{U}\mathbf{D}^{+}\mathbf{U}^{\top}, \text{ where } \mathbf{D}^{+} = diag\left(\frac{1}{\lambda_1}, \ldots, \frac{1}{\lambda_k}, 0, \ldots, 0\right).$$

A detector using the pseudoinverse will ignore the very dimensions that are constant in the whole sample set instead of considering them mandatory – their weights will be zero instead of infinity, and the corresponding axes of the degenerate hyperellipse will have infinite lengths (see Fig. 9). This is against the philosophy of capturing the regularities of the sample set. For example, the constituting property of the category *triangle* is "having 3 vertices". However, as all the examples of the category have the same number of vertices regardless of other properties, the covariance matrix will be singular and the whole dimension *vertices* will be ignored, because of receiving a zero weight in $\mathbf{D}^{+}$.

Hence, instead of $\mathbf{D}^{+}$, we shall use the standard inverse

$$\mathbf{D}^{-1} = diag\left(\frac{1}{\lambda_1}, \ldots, \frac{1}{\lambda_k}, \infty, \ldots, \infty\right) \tag{B.2}$$

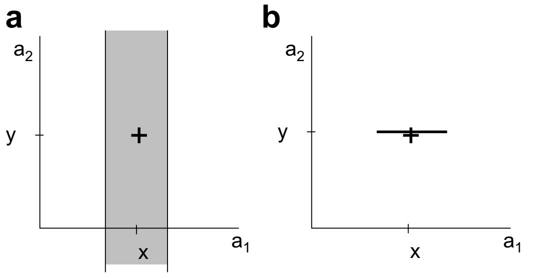

Fig. 9. The example of a category with the mean (x,y) and the zero variance of the attribute a_2. A detector using the pseudoinverse will consider the attribute a_2 unimportant (a), while a detector using the inverse with infinite values will consider the value $a_2 = y$ mandatory (b).

allowing infinite elements $\frac{1}{\lambda_i} = \infty$ for $\lambda_i = 0$. The result of the distance function will be finite, only if the rotated vector $\mathbf{U}^{\top}(\vec{x}-\vec{p})$ has a zero coordinate on the respective dimensions corresponding to infinite elements of $\mathbf{D}^{-1}$.

However, detectors based on the pseudoinverse can be useful for distinguishing the figure from the background, if examples contain attributes that are constant throughout the sample set, but irrelevant for the category.

Appendix C. Iterative formulas for computing the mean and the covariance matrix

Let $N-1$ be the number of examples seen so far, $\vec{x}^{(N)}$ be a novel (Nth) example. Then for $N = 1$,

$$\vec{p}^{(1)} = \vec{x}^{(1)}$$
$$\boldsymbol{\Sigma}^{(1)} = (0)_{n\times n}, \text{ or } \sigma^2 I_n,$$

where σ^2 is some initial estimate of the variance. For $N > 1$,

$$\vec{p}^{(N)} = \frac{N-1}{N}\vec{p}^{(N-1)} + \frac{1}{N}\vec{x}^{(N)}$$
$$\boldsymbol{\Sigma}^{(N)} = \frac{N-2}{N-1}\boldsymbol{\Sigma}^{(N-1)} + \frac{N}{(N-1)^2}(\vec{x}^{(N)} - \vec{p}^{(N)})(\vec{x}^{(N)} - \vec{p}^{(N)})^{\top}.$$

Appendix D. Category generalization based on SVD-filtering

After seeing a certain number of examples of some concept, people can decide which properties are relevant for the concept by comparing their variances. For example, if people had to induce the concept *small* from a set of small things of all shapes and colors, they could observe that, albeit finite, variances of shape and color are significantly larger than that of size.[14] Hence, shape and color could be ignored.

In our model, this type of generalization corresponds to finding those diagonal elements of $\mathbf{D}^{-1}$ that are very small

[14] In this simplified example, we abstract away from the problem of semantic dependency of the adjective *small* (see Section 7.1).

in comparison to others, and replacing them with zero. Because $\frac{1}{\lambda_1} \leqslant \ldots \leqslant \frac{1}{\lambda_n}$, we can find the largest L such that

$$\frac{\sum_{i=1}^{L} 1/\lambda_i}{\sum_{i=1}^{n} 1/\lambda_i} < b, \tag{D.1}$$

where b is some percentage threshold, e.g. $b = 10\%$, and we can set the first L elements of $\mathbf{D}^{-1}$ to zero. The idea is to abstract away those components that contribute little to the total distance. This can be viewed as an opposite process to principal component analysis (PCA), which reduces the dataspace to components with largest variances (Haykin, 1999).

D.1. Telling the figure from the ground

It follows from (D.1) that if $\mathbf{D}^{-1}$ contains some infinite elements (corresponding to totally invariant properties of the sample set), all finite ones will be abstracted away. This is right for most concepts, but sometimes the infinite elements can be an artifact of taking into account some constant but irrelevant attributes.

This is a common problem of all induction algorithms that only learn from positive examples of a category (Gold, 1967). A property shared by all positive examples should be considered irrelevant, if it is also shared by negative examples. However, our algorithm does not receive and utilize any information about negative examples.

References

Akhtar, N., & Montague, L. (1999). Early lexical acquisition: The role of cross-situational learning. *First Language, 19*, 347–358.

Austin, J. L. (1962). *How to do things with words*. Cambridge, MA: Harvard University Press.

Bailey, D., Feldman, J., Narayanan, S., & Lakoff, G. (1997). Modeling embodied lexical development. In Proceedings of the 19th cognitive science society conference, pp. 19–24.

Balkenius, C. (1999). Are there dimensions in the brain? In Spinning Ideas, Electronic Essays Dedicated to Peter Gärdenfors on His Fiftieth Birthday. Retrieved from URL <http://www.lucs.lu.se/spinning/categories/cognitive/Balkenius/Balkenius.pdf>.

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

Bratman, M. (1987). *Intention, plans, and practical reason*. Cambridge, MA: Harvard University Press.

Briscoe, T. (Ed.). (2001). *Linguistic evolution through language acquisition: Formal and computational models*. Cambridge, UK: Cambridge University Press.

Cangelosi, A. (2005). Approaches to grounding symbols in perceptual and sensorimotor categories. In H. Cohen & C. Lefebvre (Eds.), *Handbook of categorization in cognitive science* (pp. 719–737). Elsevier.

Cangelosi, A. (2006). The grounding and sharing of symbols. *Pragmatics and Cognition, 14*(2), 275–285.

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development, 15*, 17–29.

Clark, E. (1987). The principle of contrast: A constraint on language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 1–33). Hillsdale, NJ: Lawrence Erlbaum Associates.

Deacon, T. W. (1997). *The symbolic species: The co-evolution of language and the brain*. New York, NY: W.W. Norton & Co.

Divina, F., & Vogt, P. (2006). A hybrid model for learning word-meaning mappings. In P. Vogt, Y. Sugita, E. Tuci, & C. Nehaniv (Eds.), *Symbol grounding and beyond: Proceedings of the third international workshop on the emergence and evolution of linguistic communication* (pp. 1–15). Berlin/Heidelberg: Springer-Verlag.

Everitt, B.S., Landau, S., & Leese, M. (2001). *Cluster analysis*. London: Arnold.

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

Fillmore, C. J. (1982). Frame semantics. In *Linguistics in the morning calm* (pp. 111–137). Seoul: Hanshin Publishing Company.

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

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

Gold, E. M. (1967). Language identification in the limit. *Information and Control, 10*(5), 447–474.

Gärdenfors, P. (1996a). Cued and detached representations in animal cognition. *Behavioral Processes, 35*, 263–273.

Gärdenfors, P. (1996b). Language and the evolution of cognition. In V. Rialle & D. Fisette (Eds.), *Penser l'esprit: Des sciences de la cognition a' une philosophie cognitive* (pp. 151–172). Grenoble: Presses Universitaires de Grenoble.

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

Gärdenfors, P. (2004). Cooperation and the evolution of symbolic communication. In K. Oller & U. Griebel (Eds.), *The evolution of communication systems* (pp. 237–256). Cambridge, MA: MIT Press.

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

Harnad, S. (2005). Language and the game of life. Commentary on Coordinating perceptually grounded categories through language. A case study for colour. L. Steels & T. Belpaeme. *Behavioral and Brain Sciences, 28*(4), 497–498.

Hassoun, M. H. (1995). *Fundamentals of artificial neural networks*. Cambridge, MA: MIT Press.

Haykin, S. (1999). *Neural networks: A comprehensive foundation* (2nd ed.). Upper Saddle River, NJ: Prentice Hall.

Hulth, N., & Grenholm, P. (1998). A distributed clustering algorithm. *Lund University Cognitive Studies*, 74.

Ientilucci, E.J. (2003). *Using the singular value decomposition*, retrieved from URL <http://www.cis.rit.edu/ejipci/Reports/svd.pdf>.

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* (pp. 121–148). Berlin: Springer-Verlag.

Kováč, L. (2000). Fundamental principles of cognitive biology. *Evolution and Cognition, 6*, 51–69.

Kuipers, B. (1994). *Qualitative reasoning: Modeling and simulation with incomplete knowledge*. Cambridge, MA: MIT Press.

Kuipers, B., Beeson, P., Modayil, J., & Provost, J. (2006). Bootstrap learning of foundational representations. *Connection Science, 18*(2), 145–158.

Kvasnička, V., & Pospíchal, J. (1999). An emergence of coordinated communication in populations of agents. *Artificial Life, 5*(4), 319–342.

Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.

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

Lidz, J., Gleitman, H., & Gleitman, L. R. (2004). Kidz in the Hood: Syntactic bootstrapping and the mental lexicon. In D. G. Hall & S. R. Waxman (Eds.), *Weaving a lexicon* (pp. 603–636). Cambridge, MA: MIT Press.

Markman, A. B., & Gentner, D. (1993). Splitting the differences: A structural alignment view of similarity. *Journal of Memory and Language, 32*, 517–535.

Markman, E. (1992). Constraints on word learning: Speculations about their origins and domain specificity. In M. R. Gunnar & M. Maratsos (Eds.), *Modularity and constraints in language and cognition* (pp. 59–101). Hillsdale, NJ: Lawrence Erlbaum Associates.

Martin, J. H. (1991). Coding and processing of sensory information. In E. R. Kandel, J. H. Schwartz, & T. M. Jessel (Eds.), *Principles of neural science* (pp. 329–340). New York: Elsevier.

Maturana, H. R., & Varela, F. J. (1987). *The tree of knowledge: The biological roots of human understanding*. Boston, MA: Shambhala.

Medin, D. L., Altom, M. W., Edelson, S. M., & Freko, D. (1982). Correlated symptoms and simulated medical classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*, 37–50.

Minsky, M. (1975). A framework for representing knowledge. In P. M. Winston (Ed.), *The psychology of computer vision* (pp. 211–277). New York: McGraw Hill.

Nehaniv, C. (2000). The making of meaning in societies: Semiotic and information-theoretic background to the evolution of communication. In: Edmonds, B., Dautenhahn, K. (Eds.), *Proceedings of the AISB 2000 symposium: Starting from society – the application of social analogies to computational systems*. AISB, pp. 73–84.

Oliphant, M. (1997). *Formal approaches to innate and learned communication: Laying the foundation for language*. Ph.D. Thesis, University of California, San Diego, CA.

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

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

Piaget, J. (1937/1955). *The child's construction of reality*. Routledge and Kegan Paul, London, originally appeared as La construction du réel chez l'enfant. Neuchâtel, Switzerland: Delachaux et Niestle.

Piaget, J., & Inhelder, B. (1966). La Psychologie de L'enfant (The Psychology of the Child). Paris: PUF.

Poggio, T., & Girossi, F. (1990a). Networks and the best approximation property. *Biology of Cybernetics, 63*, 169–176.

Poggio, T., & Girossi, F. (1990b). Networks for approximation and learning. *Proceedings of IEEE, 78*(9), 1484–1487.

Pulvermüller, F. (1999). Words in the brain's language. *Behavioral and Brain Sciences, 22*(2), 253–279.

Quine, W. (1960). *Word and object*. Cambridge, MA: MIT Press.

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* (pp. 27–48). Hillsdale, NJ: Lawrence Erlbaum Associates.

Rosch, E. H., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382–439.

Rosenstein, M.T., & Cohen, P.R. (1998). Concepts from time series. In *Proceedings of the fifteenth national conference on artificial intelligence*, pp. 739–745.

Roy, D. (2005). Grounding words in perception and action: Computational insights. *Trends in Cognitive Sciences, 9*(8), 389–396.

Šefránek, J. (2002). Kognícia bez mentálnych procesov (Cognition without mental processes). In Rybár, J., Beňušková, L., & Kvasnička, V. (Eds.), Kognitívne vedy. Kalligram, Bratislava, pp. 200–256.

Shastri, L., Grannes, D., Narayanan, S., & Feldman, J. (1999). A connectionist encoding of schemas and reactive plans. In G. K. Kraetzschmar & G. Palm (Eds.), *Hybrid information processing in adaptive autonomous vehicles. Lecture Notes in Computer Science*. Berlin: Springer-Verlag.

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

Siskind, J. M. (1996). A computational study of cross-situational techniques for learning word-to-meaning mappings. *Cognition, 61*(1–2), 1–38.

Smith, A.D.M. (2001). Establishing communication systems without explicit meaning transmission. In Kelemen, J., & Sosík, P. (Eds.), Advances in artificial life. Proceedings of the 6th European conference on artificial life ECAL 2001. Lecture Notes in Computer Science (pp. 381–390). Prague: Springer.

Smith, A. D. M. (2003). Intelligent meaning creation in a clumpy world helps communication. *Artificial Life, 9*(2), 559–574.

Smith, A. D. M. (2005a). The inferential transmission of language. *Adaptive Behavior, 13*(4), 311–324.

Smith, A. D. M. (2005b). Mutual exclusivity: Communicative success despite conceptual divergence. In M. Tallerman (Ed.), *Language origins: Perspectives on evolution* (pp. 372–388). Oxford University Press.

Smith, K., Smith, A. D. M., Blythe, R. A., & Vogt, P. (2006). Cross-situational learning: A mathematical approach. In Vogt, P., Sugita, Y., Tuci, E., Nehaniv, C. (Eds.), *Symbol grounding and beyond: Proceedings of the third international workshop on the emergence and evolution of linguistic communication* (pp. 31–44). Berlin/Heidelberg: Springer.

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* (pp. 17–26). Berlin: Springer-Verlag.

Steels, L., & Belpaeme, T. (2005). Coordinating perceptually grounded categories through language: A case study for colour. *Behavioral and Brain Sciences, 28*(4), 469–529.

Steels, L., & Kaplan, F. (1999). Situated grounded word semantics. In Dean, T. (Ed.), *Proceedings of the sixteenth international joint conference on artificial intelligence* (pp. 862–867). San Francisco, CA: Morgan Kauffmann.

Steels, L., & Kaplan, F. (2001a). AIBO's first words: The social learning of language and meaning. *Evolution of Communication, 4*(1), 3–32.

Steels, L., & Kaplan, F. (2001b). Bootstrapping grounded word semantics. In T. Briscoe (Ed.), *Linguistic evolution through language acquisition: Formal and computational models* (pp. 53–73). Cambridge, UK: Cambridge University Press.

Steels, L., Kaplan, F., McIntyre, A., & Looveren, J. V. (2002). Crucial factors in the origins of word-meaning. In A. Wray (Ed.), *The transition to language* (pp. 252–271). Oxford: Oxford University Press.

Takáč, M. (2006a). Categorization by sensory–motor interaction in artificial agents. In Fum, D., Del Missier, F., & Stocco, A. (Eds.), *Proceedings of the 7th international conference on cognitive modeling* (pp. 310–315). Trieste, Italy: Edizioni Goliardiche.

Takáč, M. (2006b). Cognitive semantics for dynamic environments. In Hitzler, P., Schärfe, H., & Øhrstrøm, P. (Eds.), *Contributions to ICCS 2006 – 14th international conference on conceptual structures* (pp. 202–215). Aalborg, Denmark: Aalborg University Press.

Takáč, M. (2007). *Construction of meanings in living and artificial agents*. Ph.D. Thesis, Comenius University, Bratislava, Slovakia.

Takáč, M., in press. Construction of meanings in biological and artificial agents. In Trajkovski, G., & Collins, S. G. (Eds.), *Agent-based societies: Social and cultural interactions*. Hershey, PA: IGI Global.

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

Tomasello, M., & Farrar, J. (1986). Joint attention and early language. *Child Development, 57*, 1454–1463.

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

Vogt, P. (2002). The physical symbol grounding problem. *Cognitive Systems Research, 3*(3), 429–457.

Vogt, P. (2005). The emergence of compositional structures in perceptually grounded language games. *Artificial Intelligence, 167*(1–2), 206–242.

Vogt, P., & Divina, F. (2007). Social symbol grounding and language evolution. *Interaction Studies, 8*(1), 31–52.

Vygotsky, L. S. (1978). Mind in society the development of higher psychological processes. In M. Cole et al. (Eds.). Cambridge, MA: Harvard University Press.

Warren, B. (1988). Ambiguity and vagueness in adjectives. *Studia Linguistica, 42*(2), 122–171.

Waxman, S. R., & Braun, I. E. (2005). Consistent (but not variable) names as invitations to form object categories: New evidence from 12-month-old infants. *Cognition, 95*, B59–B68.

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

Wittgenstein, L. (1953). *Philosophical investigations*. New York: Macmillan.