Cognitive Semantics for Dynamic Environments

Martin Takáč

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

Abstract. In this paper, we propose and formally define a new semantics based on perceptually grounded conceptual structures – discrimination criteria, suitable for representing concepts in dynamic and open environments. The main novel contribution of the proposed structures is their ability to represent not only static objects, but also properties, relations, changes and situations in a simple and unified way. Moreover, the semantics includes mechanisms of acquisition and continuous updating of the concepts by sensory-motor and linguistic interactions, which increases its usability in real applications ranging from cognitive models to BDI architectures and industrial multi-agent systems. In comparison to other semantic paradigms, the discrimination criteria have shown to be more effective than prototypes in conceptual spaces, as they utilize the distributional information of intra-category variances of attributes.

1 Introduction

Many current technologies involve autonomous agents that need to negotiate, exchange data and coordinate their activities. As their environment is dynamic and open, no predefined ontology and/or language can capture all potential communication topics. The working system of agents able to autonomously create and continuously update ontologies relevant to their interactions with the environment and among themselves, and create a shared language from scratch, would be much more useful than any predefined fixed code.

Important insights for communication systems of artificial agents can be drawn from computational models of human language acquisition and emergence. In recent years, two modeling approaches have become paradigmatic: iterated learning [1] and self-organization [2], both simulating the community of language users with a multi-agent system. However, the underlying semantics in either approaches provide no means for representing dynamic concepts such as processes, events, actions and situations, which are necessary for verb meanings. In the elaborate semantics of conceptual spaces [3], concepts are represented as regions in a multidimensional space. However, the representation of verbs by adding the time dimension is problematic [4, 5].

In this paper, we formally define a new cognitive semantics amending the above-mentioned weaknesses and providing a simple unified framework for representation and acquisition of concepts of objects, properties, relations, changes and complex situations. Meanings of verbs are represented by cross-categorial associations of concepts. We analyze mutual relations of the perceptual, representational, linguistic and pragmatic levels, as well as effects of coupled processes of concept formation and language acquisition.

2 Levels of Description

Any agent, be it an amoeba, a web-searching agent, or a human, that needs to react to certain states of its environment in certain ways, must be able to distinguish between environmental states relevant to its task. This elementary ability to perceive some states as identical (with respect to some criteria) and different from the others, is the core of categorization.

In the simplest cases of purely reactive agents, there is no internal representation and the behavior is realized in the form of pre-programmed (or genetically coded) direct "stimulus-response" links from perceptual input to behavioral output. In this paper, we will focus on more complex cognitive agents performing tasks that require flexibility, learning and mutual co-operation and communication in open environments.

We will distinguish four levels of description:

Perceptual level. This level is an interface between the external environment of the agent and higher levels. In embodied agents, it represents the signal from the agent's sensors pre-processed by low-level perceptual routines. In software agents, it represents the input data the agent operates with, translated to the description processable by the representational level.

Representational level. This is a level of categories/concepts. Each concept is represented by a *discrimination criterion* – the function that maps a perceptual[1] input to a probability value expressing to what extent the perceptual input is an instance of the concept.

Language level. The agent's discrimination criteria are private and are not directly transferable to other agents. The agents communicate by exchanging conventionally established signals of the language level. The meanings of the signals are the perceptually grounded criteria of the representational level. The communication is successful, only if the private meanings of the agents are sufficiently similar. This occurs, if the agents use similar concept formation mechanisms and have similar experiences in the shared environment.

Pragmatic level. On this level, the agent plans and achieves its goals in the environment. It uses representations of causal knowledge about its actions and their consequences in the form of cross-categorial associations of criteria, own goals as desired situations, and plans as sequences of actions leading from the current situation to a desired one [4].

[1] More complex criteria expressing situations and events can operate on an output of other criteria (see Sect. 4.7).

3 Perceptual Level

We assume that the dynamic environment of the agent changes in discrete steps. In each time step, the agent perceives a *scene* – a set of objects described by frames of the attribute-value pairs, e.g.

```
{employeeID:105; age:30; salary:1200; accountNo:3212} .
```

Frames represent all perceptual input 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). Formally, a perceptual frame f is characterized by the set of attributes A_f and the attribute accessor function $h_f : A_f \to \mathbf{R}$. In the text we will use a more conventional notation $f.a$ instead of $h_f(a)$.

The frame is not yet a representation. If two agents perceived the same scene, they would have the same sets of input frames; however, they can categorize and represent them differently.

4 Representational Level

In our approach, each category is represented by a membership function expressing to what extent its input is a member of the category. Categories meaningful for the agent can represent individual objects, classes and properties of objects, relations between objects, changes in time, or their complex combinations.

4.1 Object Criteria

Object criteria operate on single frames. The criterion representing an object concept takes a frame as an argument and returns a value from the closed interval $[0, 1]$ expressing to what extent the frame is an instance of the concept (1 means the best, prototypical example). The object criteria can represent individual objects, if they return zero for all but one particular frame, e.g. $\mathit{JohnSmith}(f)$, properties of objects, e.g. $\mathit{married}(f)$, $\mathit{large}(f)$ or $\mathit{credible}(f)$, and classes of objects, e.g. $\mathit{student}(f)$, $\mathit{fruit}(f)$, $\mathit{desktopComputer}(f)$.[2] Actually, there is no formal difference between criteria of properties and classes.

The criterion records attribute values required in the input frame, together with weights expressing their importance for the category membership. Formally, the criterion r is characterized by the set of attributes A_r, the attribute accessor function $h_r : A_r \to \mathbf{R}$, and the weight accessor $w_r : A_r \to \mathbf{R}$. We will use the notation $r.a$ and $w_{r.a}$ instead of $h_r(a)$ and $w_r(a)$, respectively.

The result of the criterion function is inversely proportional to the weighted sum of squares of the differences of the expected values from attribute values of the input frame:

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

[2] Mnemonic identifiers of criteria should not be confused with language expressions (words), which differ by font and quotation marks, e.g. *large* vs. "large".

where

$$dist(r, f) = \sqrt{\frac{1}{|A_r|} \sum_{a \in A_r} w_{r.a} \cdot (r.a - f.a)^2} \tag{2}$$

is the weighted Euclidean distance and k is a positive constant (see the next section).

As the distance function only depends on differences in attributes recorded in the criterion, the input percept can have arbitrary many other attributes having no influence on the result. A different situation occurs, if an attribute recorded in the criterion is missing in the input percept. In such a case, the distance function can return infinity (yielding the criterion result 0), or some penalty distance depending on the weight of the attribute (remember that attributes with low weights are not very important for category membership).

4.2 Induction of Object Criteria

In the course of time, the agent perceives a mixed sequence of instances of many concepts. Even a single perceptual frame can be a good example of several concepts, e.g. a particular apple can be a good exemplar for concepts *red*, *round*, *small*, *apple*. The problem of determining which percept is an instance of which category is analyzed in Sect. 5. For now, let us assume that the agent has to induce an unknown object concept from a sequence of its examples $\{f_1, f_2, \ldots, f_n\}$. The induction is based on noticing common properties of the frames in the sample set.[3] 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. Attributes not present in every example are excluded from consideration.

The set of required attributes of the criterion can be defined as intersection of attribute sets of example frames:

$$A_r = \bigcap_{i=1}^{n} A_{f_i} \ ,$$

the required values of the criterion can be defined as mean values of the common attributes:

$$\forall a \in A_r : \ r.a = \overline{\langle f_i.a \rangle_{i=1}^{n}}$$

and the weights of attributes are inversely proportional to intra-category variances:

$$\forall a \in A_r : \ w_{r.a} = \frac{1}{var_{r.a}} = \frac{1}{\sigma^2 \, \langle f_i.a \rangle_{i=1}^{n}} \ . \tag{3}$$

The criterion also records the number (n) of instances of the category seen so far.

[3] This approach does not work for disjunctive concepts (e.g. the concept *match-box or elephant*).

Now we can rewrite (2) as:

$$dist(r,f) = \sqrt{\frac{1}{|A_r|}\sum_{a\in A_r}\frac{(r.a - f.a)^2}{var_{r.a}}} \ . \tag{4}$$

The distance function defined this way evaluates an average distance of the percept's attribute values from the mean attribute values of the sample in standard deviation units (i.e. if a frame differed from the mean by $m \cdot \sigma_a$ in each attribute a, the distance function would return m). If we further define $k = -\frac{1}{2}\ln(0.5)$ in (1), the criterion function will return 1 for zero distance and 0.5 for the distance 2σ, which could be considered the threshold for deciding, whether the percept is an instance of the category or not.

This representation resembles the conceptual spaces [3] with attributes corresponding to dimensions of the conceptual space. The criterion function then expresses the similarity of an input percept to a central member (prototype) having the mean attribute values recorded in the criterion.[4] The dimensions corresponding to attributes with low variances are "stretched out", thus acquiring a bigger decisive weight for category membership.

As the examples of a category usually come one by one, the attribute set, mean and variance of the sample are actually computed by iterative formulas.

1. If f is the first example of a new criterion ($n = 1$), create a new r in f:

$$A_r = A_f \ , \quad \forall a \in A_r : \ r.a = f.a \wedge var_{r.a} = 0 \ .$$

2. If f is the n-th example of an existing criterion ($n > 1$), update r by f:

$$\begin{aligned} A_r^{\text{new}} &= A_r \cap A_f \\ \forall a \in A_r : \ r.a^{\text{new}} &= \left(1 - \frac{1}{n}\right) r.a + \frac{1}{n} f.a \\ \forall a \in A_r : \ var_{r.a}^{\text{new}} &= \left(1 - \frac{1}{n}\right) var_{r.a} + \frac{1}{n-1}\left(f.a - r.a^{\text{new}}\right)^2 \ . \end{aligned}$$

As implied by (3), the attributes with zero variances in the sample have infinite weights. If we define $\infty \cdot 0 = 0$, the value of an attribute with zero variance will be mandatory for instances of the category (any other value in the input percept would yield zero similarity). This would lead to implausible behavior for criteria induced from just one example. As variances of all attributes have yet been zero, this criterion would yield similarity 1 for percepts exactly identical with the example and zero for all others, regardless of how close they match the example. However, if the variances still remain zero after the agent has seen many examples of the category, then it would be reasonable to induce that

[4] As the criterion does not need to have all attributes of an input percept, the distance is evaluated in the projection of the conceptual space to the subspace with dimensions A_r (e.g. to the subspace with one dimension *size* in the case of criterion *big*).

this category represents an individual. We overcome this problem by substituting $var_{r.a}$ in (3) and (4) with the estimated variance $evar_{r.a} = var_{r.a} + v_0/n^s$, where v_0 is an initial estimated variance and s is a rate of decay toward $var_{r.a}$.

Using the criteria with variances has many interesting and cognitively plausible effects. First, the categories can represent concepts with different levels of generality from individuals to general concepts with large variances. Second, the concepts with very large variances in all but one attribute can semantically represent adjectives. Third, the similarity judgments will be asymmetric, i.e. the mean case of one category can be considered a better instance of another category than vice versa.[5]

4.3 Relational Criteria

Relational criteria represent binary relations among objects, e.g. $larger(f_x, f_y)$ or $near(f_x, f_y)$. In their *aligned* version, the criteria can only refer to relations (differences) between respective values of the same[6] attributes present in both input frames ($\Delta f_{x,y}.a = f_x.a - f_y.a$).

Like the object criteria, the relational criteria are characterized by a set of attributes A_r and the accessor functions h_r and w_r. The formulas for computing the criterion function are

$$r(f_x, f_y) = e^{-k \cdot dist(r, f_x, f_y)} \tag{5}$$

where

$$dist(r, f_x, f_y) = \sqrt{\frac{1}{|A_r|} \sum_{a \in A_r} w_{r.a} \cdot (r.a - \Delta f_{x,y}.a)^2} \; . \tag{6}$$

4.4 Induction of Relational Criteria

The relational criterion r induced from a sequence of pairs

$$\{(f_{x1}, f_{y1}), (f_{x2}, f_{y2}), \ldots, (f_{xn}, f_{yn})\}$$

is characterized by

$$\begin{aligned} A_r &= \bigcap_{i=1}^{n} (A_{f_{xi}} \cap A_{f_{yi}}) \\ \forall a \in A_r: \; r.a &= \overline{\langle \Delta f_{xi,yi}.a \rangle_{i=1}^{n}} \\ \forall a \in A_r: \; w_{r.a} &= \frac{1}{var_{r.a}} = \frac{1}{\sigma^2 \, \langle \Delta f_{xi,yi}.a \rangle_{i=1}^{n}} \; . \end{aligned}$$

Like the object criteria, the relational criteria are computed by iterative formulas and variances are substituted by estimated variances (see Sect. 4.2).

[5] People show the same effect, e.g. Tel Aviv is judged more similar to New York than vice versa [6].

[6] In general, the criteria could refer to cross-attribute relations between any values of the two input frames (differences $f_x.a_1 - f_y.a_2$, where $a_1 \neq a_2$). In this paper we restrict our attention to the aligned criteria.

4.5 Qualitative Relations

The above described relational criteria are good for representing relations defined by a magnitude (e.g. *near*, *muchLarger*, *almostTheSame*). However, they cannot capture relations only based on signs or ordering of attribute values (so-called qualitative relations), e.g. $\mathit{bigger}(f_x, f_y)$, where $f_x.\mathit{size} > f_y.\mathit{size}$, or $\mathit{leftOf}(f_x, f_y)$, where $f_x.\mathit{posX} < f_y.\mathit{posX}$.

The qualitative version of relational criteria is sensitive to the sign pattern of the differences of attributes, $\operatorname{sgn}(\Delta f_{x,y}.a)$. The sign of an attribute difference that remains the same in the sample becomes mandatory for category membership. Attributes not present in each pair of examples and attributes not having the same difference sign across the whole sample are excluded. The criteria are computed by iterative formulas:

1. If (f_x, f_y) is the first example pair of a new relation criterion ($n = 1$),

$$A_r = A_{f_x} \cap A_{f_y} \ , \ \forall a \in A_r : \ r.a = \operatorname{sgn}(\Delta f_{x,y}.a) \ .$$

2. If (f_x, f_y) is the n-th example pair of an existing criterion ($n > 1$),

$$A_r^{\text{new}} = \left\{ a \in A_r \cap A_{f_x} \cap A_{f_y} \mid \operatorname{sgn}(\Delta f_{x,y}.a) = r.a \right\} \ .$$

The sign pattern is recorded only once, upon seeing the first example. Later updates of the criterion only remove from A_r the attributes not occurring in the new example pair with the same difference sign as recorded.

The criterion function uses a sign version of (6)

$$\mathit{dist}(r, f_x, f_y) = \sqrt{\frac{1}{|A_r|} \sum_{a \in A_r} w_{r.a} \cdot \delta\left(r.a, \operatorname{sgn}\left(\Delta f_{x,y}.a\right)\right)^2}$$

where all weights $w_{r.a}$ are infinite (because all signs have zero variance in the sample) and the sign difference $\delta(s_1, s_2)$ is 0 for $s_1 = s_2$, and 1 otherwise. This way the result of the criterion function is always binary: 1, if $\operatorname{sgn}(\Delta f_{x,y}.a) = r.a$ for all attributes $a \in A_r$, and 0 otherwise.

4.6 Change Criteria

The ability to perceive and represent changes is very important for any agent operating in a dynamic environment. In a continuous world, even 4-month-old children can eye-track moving objects and develop the concept of *object continuity* [7], which is a necessary condition for noticing changes of objects.

As the time is discrete in our model, we must ensure that the agent does not perceive scenes in subsequent time steps as independent, but as the sets with established correspondences between frames representing percepts of the same object at different times. This can be managed by redefining the *scene* – the perceptual input of the agent at time t (see Sect. 3) to a set

$$S_t = \left\{ \left(f_1^{(t)}, f_1^{(t-1)}\right), \left(f_2^{(t)}, f_2^{(t-1)}\right), \ldots, \left(f_n^{(t)}, f_n^{(t-1)}\right) \right\}$$

of percepts linked with their one-step history. If an object just appeared on the scene at time t, its history frame $f^{(t-1)}$ will be assigned a special value $\perp$. If an object had been present on the scene in the time $t-1$ and now disappeared, $f^{(t)} = \perp$.[7] Otherwise both $f^{(t-1)}$ and $f^{(t)}$ are standard frames defined in Sect. 3 (the scene does not contain pairs with $\perp$ on both positions).

Technically, the change criteria are the relational criteria, which are always applied to frames $f^{(t)}, f^{(t-1)}$ of the same object. Some changes can be captured by qualitative relations, e.g. *grew* can be expressed by $\operatorname{sgn}(\Delta f_{t,t-1}.size) = 1$, others require encoding of a typical change's magnitude, e.g. movement criteria for *crawled*, *moved*, *jumped* can differ in mean values of $\Delta f_{t,t-1}.position$.[8] The criteria with zero sign pattern of some attributes can represent a state or a persistence of a property, e.g. *stayed*.

4.7 Criteria of Situations

Criteria representing complex situations or properties of the whole scenes, e.g. concepts *riskyInvestment* or *catOnHotTinRoof*, can be built as compositions of elementary criteria describing the properties of objects, which should be present on the scene, together with their required mutual relations. The composite criterion can prescribe values that the elementary criteria should return and weights of their contributions to the total result.

We say that a situation criterion has arity k, if it describes the properties of and relations among k objects.[9] Formally, a situation criterion s of arity k is characterized by a set of elementary[10] criteria C_s, the value accessor $h_s : C_s \to [0,1]$, the weight accessor $w_s : C_s \to \mathbf{R}$ and the argument descriptor $d_s : C_s \to \{1,\ldots,k\} \times \{1,\ldots,k\}$. For a criterion $r \in C_s$, the argument descriptor $d_s(r)$ determines indices of the frames in the input k-tuple

$$K = \left\langle \left(f_1^{(t)}, f_1^{(t-1)}\right), \left(f_2^{(t)}, f_2^{(t-1)}\right), \ldots, \left(f_k^{(t)}, f_k^{(t-1)}\right) \right\rangle$$

that r should be applied to, and $h_s(r)$ expresses the value that r should return. From now on, we will use the notation $s.r$ and $w_{s.r}$ instead of $h_s(r)$ and $w_s(r)$.

Let $\phi(K, d_s(r), r)$ denote the result of projection of K to particular frames determined by indices $d_s(r)$ and the type[11] of r. Then the distance function of

[7] Special change criteria *appeared* and *disappeared* are based on detecting $\perp$ on the respective position in the input pair.

[8] Real semantics of motion verbs is much more complicated and depends on the manner of motion and other aspects.

[9] In fact, a situation criterion has only one input argument – the whole scene S_t and the function is evaluated for all ordered subsets of S_t with k elements. The criterion then returns maximum of the results for the k-subsets.

[10] Hierarchical compositions of arbitrary depth could be obtained by allowing situational criteria in C_s.

[11] For example, $\phi(K,(i,i),r) = f_i^{(t)}$, if r is an object criterion, but it is equal to $\left(f_i^{(t)}, f_i^{(t-1)}\right)$, if r is a criterion of change.

the situation criterion s can be defined as

$$dist(s,K) = \sqrt{\frac{1}{|C_s|}\sum_{r\in C_s} w_{s.r}\cdot\left(s.r - r\left(\phi\left(K, d_s(r), r\right)\right)\right)^2}\ .$$

Like in the other types of criteria, $s.r$ can express the mean of values returned by r on a sample and $w_{s.r}$ can be inversely proportional to the variance $var_{s.r}$ in the sample. However, working out the details of induction of situational criteria is a task for our future research.

5 Learning of Criteria

So far, we have assumed the learning setting where the agent had a sequence of examples of a criterion to be induced. In reality the agent perceives a mixture of frames that are examples of various concepts and it must somehow determine, which criteria to update (or whether to create a new criterion).

The simplest unsupervised learning procedure works as follows:

1. For an input percept[12] f and a set of criteria C, find $r^* \in C$ such that $\forall r \in C: \ r^*(f) \geq r(f)$.
2. If $r^*(f) > \theta$, update r^* by f, else create a new criterion r^{new} with f as its first example (θ is a threshold parameter).

This algorithm clusters the input by its distributional properties. More environmentally relevant algorithm should group the percepts by some pragmatic criteria, e.g. discriminating between two types of similar mushrooms, if one is edible and the other poisonous. In [8] we have described a model of the agent actively exploring its environment and grouping percepts to categories by their *affordances* [9], e.g. forming a category of objects too heavy to be lifted.[13]

However, words expressing various aspects of objects and their relations often capture finer distinctions that are best acquired in an instructional setting. In this paper, we focus on the induction of criteria guided by linguistic instructions. An agent – novice in the environment can be instructed by other agents, which already share a common language, or by a human in the case of human–computer interface applications.[14] This way the process of criteria formation is coupled with the language acquisition and we can explore their mutual influence.

[12] The formulation for pairs of input percepts and for scenes is analogical.

[13] In the implemented experiment, the scene contained frames of (simulated) toys, fruits and furniture. The agent was trying to either lift objects, or put them down and observed the resulting changes. The objects were considered members of the same category, if an action performed on them yielded the similar result.

[14] Categories acquired by instructions in a natural language are culturally dependent – a well-known example is the existence of different color categories in world languages [10]. Our model has no culture-specific bias built-in and can learn equally well from instructions in an artificial language. The only bias is the assumption that categories are convex, i.e. an average of two examples of a category should be a member of that category too (see also footnote 3).

6 The Language Level

In each time step, the agent perceiving a scene also receives an utterance describing some aspects of the scene and a *focus*. The focus is a non-verbal reference directing the attention of the agent to the parts of the scene described by the utterance. In child language acquisition, it corresponds to joint attention of the child and the mother, gaze following and pointing [11]; in a later phase, when the agent already understands some words, the context of reference can be narrowed by meaning of the known parts of the utterance, e.g. the utterance "*big X*" can narrow the context to big objects on the scene. Formally, the focus ϕ is a projection of the scene

$$S_t = \left\{ \left(f_1^{(t)}, f_1^{(t-1)}\right), \left(f_2^{(t)}, f_2^{(t-1)}\right), \ldots, \left(f_n^{(t)}, f_n^{(t-1)}\right) \right\}$$

of the form ϕ_i or $\phi_{i,j}$, where

$$\phi_i(S_t) = \begin{cases} f_i^{(t)} & \text{if } 1 \le i \le n \\ \left(f_{-i}^{(t)}, f_{-i}^{(t-1)}\right) & \text{if } -n \le i \le -1 \\ \emptyset & \text{otherwise ,} \end{cases}$$

$$\phi_{i,j}(S_t) = \begin{cases} \left(f_i^{(t)}, f_j^{(t)}\right) & \text{if } 1 \le i, j \le n \\ \emptyset & \text{otherwise .} \end{cases}$$

Negative indices of the foci refer to relations across time. The definition of the focus can be straightforwardly extended from projections on single elements and pairs to any ordered subset of the scene. We shall call the result of the projection function a *referent* of the focus.

6.1 Coupled Acquisition of Language and Concepts

In the simplest learning situation, the agent receives a scene S_t, a one-word utterance w and a focus ϕ. The agent's lexical knowledge is stored in the form of associations of *words* (sequences of letters) and their *meanings*. A meaning of each word is a discrimination criterion induced from all the contexts (referents), in which the word has been used.[15]

Although natural languages do contain words with multiple meanings (homonyms) and multiple expressions for a single meaning (synonyms), in case of the coupled acquisition of language and concepts from the scratch, it would be very useful to start with no homonymy and synonymy.[16] In fact, these will be the crucial assumptions guiding the acquisition in our model:

[15] Remember that referents can be also pairs of frames, e.g. the referents for meanings of words "*leftOf*", "*biggerThan*", or "*movedUp*". As there is no grammar in our model, phrases such us "*leftOf*" are written as single words.

[16] Children acquiring a language use a similar *mutual exclusivity* constraint assuming that novel words cannot name objects, which have already had a name [12].

1. *No true homonymy:* A single word has a single meaning, even if it is used with more referents. The referents are considered instances of the same category.[17] Put in practice, if the agent already knows some meaning r of a word w and w is now used with a new referent $\phi(S_t)$, r is updated by $\phi(S_t)$ (instead of creating a new criterion).
2. *No true synonymy:* Different words have different meanings, even if they share a referent (in that case they express different aspects of the referent). Put in practice, if the agent hears an unknown word w in the context of a referent $\phi(S_t)$, a new criterion r is created with $\phi(S_t)$ being its first example and r is associated with w in the agent's lexicon.

Example. Let us consider an agent living in a world of geometric shapes placed on a 50×50 grid with the point coordinates $(1, 1)$ on the left bottom and $(50, 50)$ on the right top. If the agent perceives an object

```
f = {vertices: 3; size: 18; color: 3; posX: 1; posY:23}
```

denoted by words "*left*", "*big*", "*triangle*", it creates three discrimination criteria, which are initially identical and represent the "snapshot" of the perceived object f. The criteria begin to differentiate, when they are updated by more and more instances. E.g. the "snapshot" criterion associated with the word "*triangle*" will be updated by frames of various objects having all kinds of colors, positions, sizes and other properties, but all having 3 vertices. Attributes not common to all instances will be removed from the criterion and others will gain lower importance because of their high variance in the sample. Hence, the property of having 3 vertices (with zero variance in the sample) will become decisive in the criterion associated with the word "*triangle*". Also the word "*left*" will be heard with many different objects sharing the property of low value of the attribute *posX*, etc. The more contexts of the word's use, the bigger the probability that the referents will vary in the properties irrelevant for the meaning of the word. However, if e.g. all triangles in the agent's world are big, then having a big size will become part of the meaning of the word "*triangle*".

6.2 Multi-word Utterances

The principles used for acquisition of meanings of single words can also be applied to multi-word noun phrases: each word of a phrase is assumed to denote a different aspect of the referent. In the acquisition model with no grammar, the word order of the phrase is unimportant and induction from the phrase "*left big triangle*" (or any of its permutations) has the same effect as three subsequent inductions from single words described in the example in the previous section.

[17] If a word has been used in apparently different contexts (e.g. if the referents have nothing in common), the agent can detect homonymy and associate multiple criteria with the word. However, handling homonymy has not yet been implemented in our model.

However, the word order of a noun phrase can make a difference in languages with grammar (such as English).[18] Moreover, the meaning of an adjective in adjective-noun phrases can be contextually dependent on the noun, e.g. in English the word "*little*" refers to different sizes in phrases "*little mouse*" and "*little elephant*". The above described acquisition process would simply update the criterion associated with the word "*little*" by snapshots of an elephant and a mouse, which would yield a no-good criterion having the attribute *size* with value of the average of the mouse's and elephant's sizes and a big variance.

Also, even if the agent has acquired the correct meanings of words such as "*cat*", "*on*", "*hot*", "*tin*", and "*roof*" in the single-word induction setting, it cannot understand the meaning of a phrase "*cat on hot tin roof*", unless it knows the rules of grammatical word composition.

7 Pragmatic Level

How to put it all together? How can the agent use the learned criteria and language expressions? The answer depends on the particular application. However, we can define several functions, which can be useful for the purpose of communication with other agents.

Let us assume that the agent has acquired a lexicon of one-to-one associations $\mathcal{L} \subset W \times C$, where W is a set of learned words and C is a set of induced categories. Let S be a scene and ϕ be a focus. Then we define:

Understanding. The function $U : W \rightarrow C$ returns a criterion r that is the meaning of the word w. For $w \in W$,

$$U(w) = r, \text{ such that } (w, r) \in \mathcal{L} \ .$$

Expression. The function $E : C \rightarrow W$ expresses the criterion r by the word w. For $r \in C$,

$$E(r) = w, \text{ such that } (w, r) \in \mathcal{L} \ .$$

Interpretation. The function I returns a set of concepts that a referent is an instance of, with the degree of membership determined by the threshold parameter θ. For a referent $\phi(S)$,

$$I_\theta(\phi(S)) = \{r \in C \mid r(\phi(S)) > \theta\} \ .$$

Naming. The function N is a composition $I \circ E$ and returns names of all categories that the referent is an instance of. For a referent $\phi(S)$,

$$N_\theta(\phi(S)) = \{w = E(r) \mid r \in I_\theta(\phi(S))\} \ .$$

[18] Children acquiring a language use also other grammatical cues to constrain possible meanings of words (the syntactic bootstrapping theory [13]). Implementing a model of coupled acquisition of concepts and a language with grammar is a topic for future research.

Reference. The function R returns a set of foci determining the referents of the meaning r present on the scene S. Strictness of the membership is given by the threshold parameter θ. For a meaning $r \in C$ and a scene S,

$$R_\theta(r, S) = \{\phi \mid r(\phi(S)) > \theta\} \ .$$

Pragmatic understanding. The function P is a composition $U \circ R$ and returns the set of foci determining the referents of the word w present on the scene S. For a word $w \in W$,

$$P_\theta(w, S) = R_\theta(U(w), S) \ .$$

The reference function can have a contrastive version $R^*(r, S)$ returning foci of referents on the scene S, for which r gives the maximum value (regardless of any threshold). Using this function in P enables the agent to understand also the contrastive use of words, e.g. if the agent heard a word "*big*" uttered along with the scene S containing only small objects, $P_{0.5}(\text{"}big\text{"}, S)$ would return an empty set, while contrastive $P^* = U \circ R^*$ would return best matching objects, i.e. the biggest of the small ones.

7.1 Goals, Actions, Causality and Planning

The proposed representation can also serve for the agent's internal reasoning related to its operations in the environment, e.g. the agent's behavior can be controlled by a set of situational rules composed by discrimination criteria.

The agent, being able to actively change its environment by performing actions, can represent causal knowledge about the actions and their consequences as the cross-categorial associations of the type

$$(\textit{preconditions}, \textit{action} \rightarrow \textit{consequence})$$

where *preconditions* are represented by the situation criterion, *action* by the object criterion representing the action's parameters, and *consequences* by the change criteria. The agent can represent its own goals as desired situations and then plan a sequence of actions presumably leading from the current situation to a desired one.

8 Conclusion

The discrimination criteria provide a unified framework for representation and acquisition of various types of concepts. The process of their formation is grounded in perception and tightly connected with language, which makes them a good candidate for language semantics. Based on recording intra-category variances of attributes, the criteria are more sensitive and flexible representation than prototypes in conceptual spaces, as we demonstrated also experimentally in [8].

To be closer to real semantics of verbs, the representation needs to be extended with composite structures for encoding perspective, manner, roles and other verb parameters. The induction process in its current form is not robust against mistakes and noise (a wrong example in the sample can fatally affect the criterion), thus it should be supplemented with stochastic induction. Future research should also work out the details of situational criteria and their induction and extend the language level to multi-word phrases.

Nevertheless, the proposed semantics provides a solid representational basis for cognitive models of language acquisition, BDI architectures, and industrial applications in open artificial environments, such as the world wide web.

9 Acknowledgments

This work was supported by Agency for Promotion Research and Development under the contract No. APVV-20-P04805 and by the grant VEGA 1/3105/06.

References

1. Kirby, S., Hurford, J.: The emergence of linguistic structure: an overview of the iterated learning model. In Parisi, D., Cangelosi, A., eds.: Computational Approaches to the Evolution of Language and Communication, Berlin, Springer-Verlag (2001)
2. Steels, L.: Language as a complex adaptive system. In Schoenauer, M., ed.: Proceedings of PPSN-VI, Berlin, Springer-Verlag (2000)
3. Gärdenfors, P.: Conceptual Spaces. MIT Press, Cambridge, MA (2000)
4. Šefránek, J.: Kognícia bez mentálnych procesov. In Beňušková, L., et al., eds.: Kognitívne vedy, Bratislava, Kaligram (2002) 200–256
5. Geuder, W., Weisgerber, M.: Verbs in conceptual space. In Katz, G., Reinhard, S., Reuter, P., eds.: Sinn und Bedeutung 6, Proceedings of the sixth meeting of the Gesellschaft für Semantik. Volume 1., Osnabrück, Publications of the Institute of Cognitive Science (2002)
6. Tversky, A.: Features of similarity. Psychological Review **84** (1977) 327–352
7. Johnson, S.P., Amso, D., Slemmer, J.A.: Development of object concepts in infancy: Evidence for early learning in an eye tracking paradigm. Proceedings of the National Academy of Sciences USA **100** (2003) 10568–10573
8. Takáč, M.: Categorization by sensory-motor interaction in artificial agents. In: Proceedings of the 7th International Conference on Cognitive Modeling, Trieste, Italy (2006)
9. Gibson, J.J.: The Ecological Approach to Visual Perception. Houghton Mifflin, Boston (1979)
10. Berlin, B., Kay, P.: Basic Color Terms: Their Universality and Evolution. University of California Press, Berkeley, CA (1969)
11. Tomasello, M., Farrar, J.: Joint attention and early language. Child Development **57** (1986) 1454–1463
12. Markman, E.: Constraints on word learning: Speculations about their origins and domain specificity. In Gunnar, M.R., Maratsos, M., eds.: Modularity and constraints in language and cognition, Hillsdale, NJ, Lawrence Erlbaum Associates (1992)
13. Lidz, J., Gleitman, H., Gleitman, L.R.: Kidz in the 'Hood: Syntactic bootstrapping and the mental lexicon. In Hall, D.G., Waxman, S.R., eds.: Weaving a Lexicon, Cambridge, MA, MIT Press (2004) 603–636