DEVELOPING EPISODIC SEMANTICS

Martin Takáč

Center for Cognitive Science DAI FMFI, Comenius University
Mlynská dolina, 842 48 Bratislava, SLOVAKIA, takac@ii.fmph.uniba.sk

ABSTRACT

In this paper, we describe a variant of the semantics of distinguishing criteria for artificial agents. A distinguishing criterion is a computational abstraction of meaning and is modeled by a locally tuned detector reacting to some part of its input space. All criteria are dynamic in that they are autonomously and incrementally (re)constructed during the life/runtime of the developing agent. We apply this semantic framework to learning from episodes describing both performed and observed actions in a virtual environment by "who did what to whom and with what result" theta role frames. Thanks to their mirror neuron-like nature and auto-associative property, the acquired criteria enable the agent to plan actions and predict their results, infer internal states and possible actions of other agents, and open a route to situated pragmatic language use.

1. INTRODUCTION

The work described in this contribution could be characterized as a constructivist or *developmental* [1] approach to building intelligent systems, which emphasizes the role of incremental and continuous learning. Instead of an inbuilt (pre-programmed) knowledge/ontology, the systems are endowed with learning capabilities that allow them to acquire/construct representation of knowledge relevant to their goals during their lifetime (run-time). This approach is especially suitable for tasks in open and dynamic environments (virtual or real) with characteristics that cannot be exhaustively captured at design-time, such as intelligent search on the world wide web or a robot control on a remote planet.

In [2], we have theoretically elaborated the notion of "meaning" and "understanding" in neutral non-anthropocentric terms so that it can also be used with preverbal living organisms and artificial agents.[1] We have identified several design principles for building understanding agents:

1. Individual representation of meanings of each agent should be incrementally and continually constructed in its interactions with the environment.
2. Possible meanings should also include those capturing the dynamics of the environment.
3. For mutual understanding, individually constructed meanings should be collectively coordinated in social interactions.

Based on Šefránek's idea of distinguishing criteria as basic semantic units and a formal abstraction of the ability to distinguish [3], we have elaborated a computational formalization of the *semantics of distinguishing criteria*[2] (SDC) [4], together with mechanisms of its autonomous construction. The ideas were implemented in computational models of meaning construction driven by (simulated) sensorimotor interaction [5] and social instruction [6], wherein we also explored stability of meanings in iterated intergenerational transmission.

In this contribution, we focus on the individual aspect of meaning acquisition process. We introduce a modification of SDC that features robust and stochastic autonomous action-based construction of episodic representation of meanings. We will suggest how this representation can be used in language grounding, prediction, planning, and for a computational equivalent of empathy.

2. SEMANTICS OF DISTINGUISHING CRITERIA

The semantic knowledge of an agent is represented by a (dynamically changing) set of *distinguishing criteria*. Each distinguishing criterion r represents one meaning and has the following properties:

learnability: r can be incrementally and continually constructed from an incoming sequence of examples $\{x_1, \ldots, x_N\}$,

identification: for an input x, r returns a probability that x is an instance of the concept represented by r. $r(x) \rightarrow [0, 1]$,

auto-associativity: for a (possibly incomplete or noisy) input x, r returns the best example (prototype) of the represented concept.

2.1. Geometrical interpretation

In computational implementation of distinguishing criteria, we were inspired by a notion of *conceptual spaces*

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

[1] In the following, we will use the term *agent* to denote any autonomous entity achieving some goals in a (real or virtual) environment by perception and action, regardless of whether it is a human, animal, robot or a software entity.

[2] In our older works, we also used the nomenclature *discrimination* or *identification* criteria.

[7], in which objects are represented as vectors/points in geometrical spaces with dimensions defined by their attributes, and (natural) concepts correspond to (convex) regions in space. An object (projected to a vector/point) is categorized by finding the closest *prototype* (geometric center of examples of some category).

The crucial difference between the original conceptual spaces and our approach is that we use as a distance measure that determines category membership[3] a *unique* metric stored with each distinguishing criterion, instead of a common (Euclidean) metric. Each metric depends on parameters that are iteratively recomputed from incoming examples of the category and reflect their statistical properties such as (co)variances and frequencies of occurrence of their attributes,[4] so that differences in highly varying characteristics are less important for the total distance than in invariant ones [6]. Hence, distinguishing criteria work as *locally tuned detectors* with adaptively changing activity curves and receptive fields of various shapes, which have a good neurobiological plausibility [8]. In [4, 6], we described how locally tuned detectors with transformed input can serve as representation of various types of concepts (objects, properties, relations, changes, situations).

3. EPISODIC REPRESENTATION

3.1. Frame interface

In all our models, we assume that distinguishing criteria operate on frames – sets of $\langle$*attribute*: *numeric value*$\rangle$ pairs, e.g.

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

Frames can be conceived as vectors in the space with dimensions defined by their attribute names, and are general enough to serve as an interface for describing arbitrary (possibly preprocessed) input, e.g. vectors of sensor readings, matrices of retinal activity, frequency tables in text analysis, database records, etc. This is especially useful for agents that will operate in virtual environments.

We even use frames for information passing between layers of distinguishing criteria in our model architecture. Criteria can be organized in layers, where the criteria of the lowest layer operate directly on the external input and higher-level criteria operate on frames that represent activity configuration of the lower layer.[5] Such an architecture resembles a neural network; however, since nodes correspond to concepts, individual memories or types of episodes, the network is inherently dynamic and has no fixed topology. The set of potential nodes is open – new nodes appear, obsolete disappear, etc.

Hence, our goal can be reformulated as building a learning mechanism that discovers structural regularities in frames with potentially open set of attributes and uses them for categorization and auto-association [9].

a)

```
[ACT  ={eat:1; howMuch:6},
 SUBJ ={dir:2; @energy:10; posX:4; posY:3},
 OBJ  ={nutrition:129; posX:3; posY:3},
 dSUBJ={dir:0; @energy:+6; posX:0; posY:0},
 dOBJ ={nutrition:-6; posX:0; posY:0}]
```

b)

```
{ACT_eat:1; ACT_howMuch:6; SUBJ_dir:2;
 SUBJ_@energy:10; SUBJ_posX:4; SUBJ_posY:3;
 OBJ_nutrition:129; OBJ_posX:3; OBJ_posY:3;
 dSUBJ_dir:0; dSUBJ_@energy:6; dSUBJ_posX:0;
 dSUBJ_posY:0; dOBJ_nutrition:-6; dOBJ_posX:0;
 dOBJ_posY:0}
```

Figure 1. Theta role structure describing a particular episode – a subject placed on the position (4,3) with energy level 10 ate 6 units of an object placed on (3,3) with nutrition value 129, which resulted in +6 increase in the subject's energy level and -6 decrease in the object's nutrition value. No other properties were changed. Attributes with '@' prefix ('@energy') are internal, i.e. only perceivable by the subject itself. The letter 'd' in 'dSUBJ' and 'dOBJ' stands for Δ, and denotes change, i.e. the difference between the corresponding attribute values in two subsequent time steps. Figure a) shows the theta role frame in more human-readable format, but the agents actually use format shown in b), which merges all roles into one frame by using prefixes.

3.2. Episode frame

In particular, we apply this general framework to input of a special type [ACT, SUBJ, OBJ, ΔSUBJ, ΔOBJ] that represents the agent's perception of one elementary behavioral episode in terms of roles defined by a performed/observed action (Figure 1). In the input frame, attributes with SUBJ_ prefix describe perceived properties of the subject (agent) of the action, OBJ_ the object (patient) of the action, ACT_ parameters of the action and ΔSUBJ_, ΔOBJ_ its consequences (resulting change in subject/object's properties). The frame can be incomplete, i.e. missing some attributes or the whole role(s). In that case, auto-associative retrieval mechanism should fill in the missing information, based on the agent's previous experience. By extracting structural regularities in input frames, the agent gradually acquires situated knowledge about its environment, its own abilities, affordances of objects, etc.

3.3. Example experiment

The details can best be illustrated in a particular example. We have deliberately chosen a very simple toy example that, however, well demonstrates the basic principles and possibilities of our approach. We have implemented a simple 5x5 simulated environment with 4 agents characterized by attributes direction, position (X,Y), energy, and

[3]The degree of category membership is an exponentially decaying function of the distance from the category's prototype that is iteratively recomputed as the sliding average of incoming examples.

[4]Due to errors, noise, and/or incomplete or developing perception, sets of attributes of examples of a category do not have to be identical.

[5]Names of attributes in the frame are IDs of active criteria and values are their activities.

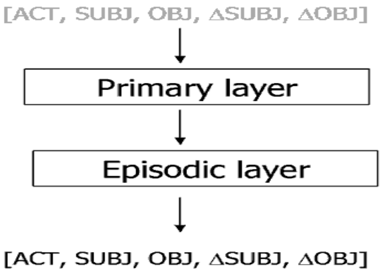

Figure 2. The architecture of the representation consisting of two layers of distinguishing criteria. The gray frame depicts possibly incomplete input frame; the black one represents auto-associative retrieval.

10 objects (food sources) characterized by position (X,Y) and nutrition value (all values were generated randomly). In each time step, each agent could perform a random action with parameters:[6] *turn* (resulting in the change of own direction), *move* (change of own position), or *eat* a bit of a randomly selected object in its vicinity (resulting in increase of own energy and decrease of nutrition value of the object proportionally to the eaten amount).

Initially, the agent had no knowledge of effects of the actions, but could learn it gradually by observing its own actions and those of other agents.[7] Agent could perceive all actions currently happening in its visual field.[8] The roles of the input frame were filled in with attributes as perceived by the observing agent, e.g. the attribute of energy was private (only visible to its possessor but not to other agents). ΔSUBJ and ΔOBJ parts of the frame were constructed by subtracting values of the subject/object's perceived attributes before the action from those after the action.

3.4. The architecture and learning

We used two-layer architecture (Figure 2). The input frame was first fed into the *primary layer*, which transformed real-valued attributes to population coding – configuration of [0,1] activities of primary detectors. The primary layer activity configuration was then fed into the *episodic layer* that stored memories of (types of) events encountered.

3.4.1. Primary layer

Technically, the primary layer consisted of separate *landmark pools*, dynamically created for each newly-encountered attribute. The purpose of each landmark pool was to approximate the continuous domain of its attribute with a limited number of dynamically created *landmarks* (elementary 1-dimensional distinguishing criteria) that reflected the probability distribution of the encountered values (Figure 3).

Consider an attribute a. In the course of time, the agent encounters a sequence of a's values $\{x_1, \ldots, x_N \mid x_i \in R\}$ (in different frames). The landmark pool for a iteratively computes the a's sample variance σ^2 and creates a set of landmarks $L_a = \{v_1, \ldots, v_k \mid k \leq K\}$, where K is the capacity of the pool. For an input value x, each landmark v_i reacts with the activity

$$v_i(x) = \exp\left(-c \cdot d_i(x)\right), \text{where } d_i(x) = \frac{|x - \overline{v}_i|}{\frac{2\sigma}{K}}$$

(c is a positive constant[9] and $\overline{v}_i$ is the sliding average of the values that updated v_i). Hence, any real-valued x can be represented by the landmark reacting to x with the highest activity (the winner).[10] This way, the landmark pools of the primary layer encode all attributes of the input theta role frame and pass it on to the episodic layer.[11]

As the (initially empty) landmark pools are created dynamically upon encountering new values, the algorithm that computes the activities should also provide for learning. We use a simple algorithm that finds a winner and either updates the winner if it is close enough to the input value, or inserts a new landmark equal to the input value otherwise, unless the landmark pool has already reached its maximum capacity (see Figure 4).

Sometimes it happens that two independently created landmarks converge to the same value in the course of time. To prevent this situation, we implemented a pairwise testing for similarity between the landmarks. If any two landmarks with values $\overline{v}_1$ and $\overline{v}_2$ and hit frequencies (of being a winner) N_1 and N_2 react to each other's value with the activity bigger than the identity threshold θ_I, they are merged into one landmark with the hit frequency $N = N_1 + N_2$ and the value $v = \frac{N_1}{N}\overline{v}_1 + \frac{N_2}{N}\overline{v}_2$.

3.4.2. Episodic layer

The episodic layer is a dynamically created set of distinguishing criteria $E = \{e_1, \ldots, e_n\}$ that represent memories of types of events (behavioral episodes). Each criterion $e \in E$ is created from and updated by a sample of its incoming example frames. It maintains the number N of examples seen so far (the size of the sample), a set of attributes A_e (the union of sets of attributes of all example frames), and a pair $\langle f_a, p_a \rangle$ for each attribute $a \in A_e$, where f_a is the frequency of occurrence of a in the example frames and p_a is the sliding average of all encountered values of a. For an input frame

[6] The parameters, such as angle, number of steps, or eaten amount were random too, but constrained by agent's simulated physical limits.

[7] The agent could try to move beyond the borders of the lattice, eat from an empty food source or even try to eat another agent – such actions resulted in no change of attributes.

[8] The visual field of an agent was the segment of the lattice defined by its direction $\pm 45^\circ$ within the distance 4 from its position.

[9] In our experiments, we used $c = -\ln(0.5)/2$ so that the value $v_i(x) = 0.5$ (used as a decision threshold θ_m for adding new landmarks) is returned for $d_i(x) = 2$, which means $|x - \overline{v}_i| = \frac{4\sigma}{K}$, i.e. $\frac{1}{K}$th fraction of the estimated range of the attribute. The other parameters we used were $K = 20$ and $\theta_I = 0.95$.

[10] Possible alternatives are all landmarks with activities bigger than some threshold, or simply the activity configuration of the whole landmark pool.

[11] The names of attributes in the resulting frame combine original attribute names with unique IDs of landmarks in the pools.

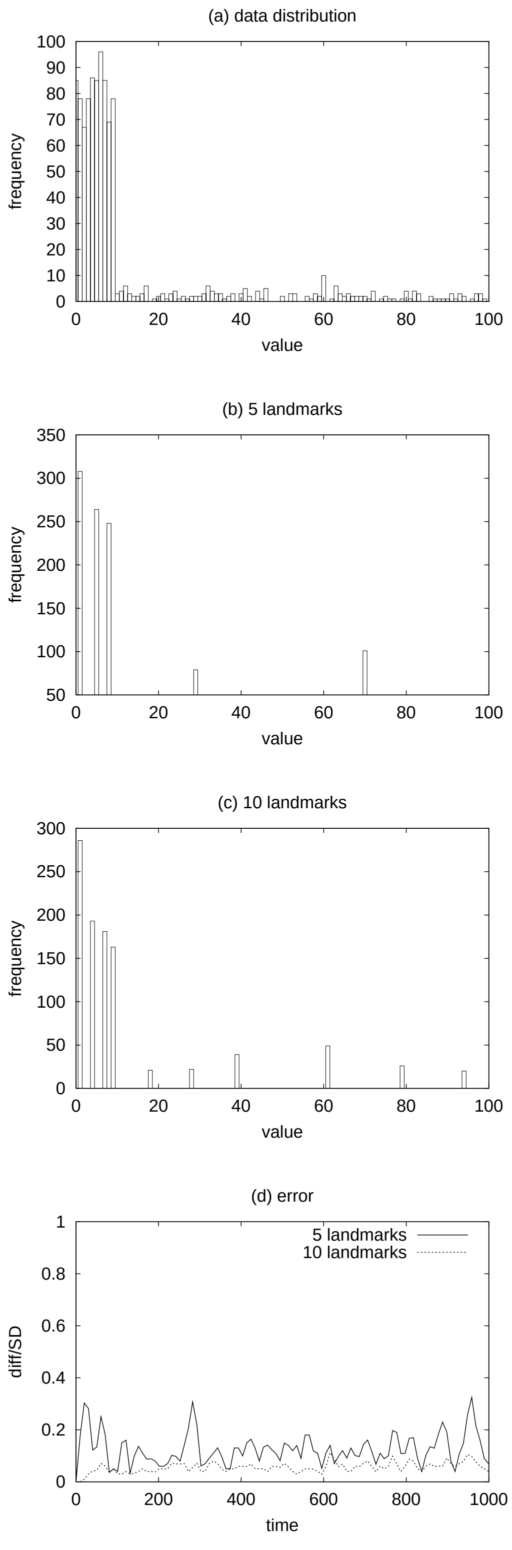

Figure 3. Dynamically created landmark pool approximating a non-uniform distribution of 1000 sequentially presented attribute values (a) with 5 (b) and 10 (c) landmarks. The difference between a presented value and the most active landmark normed by the standard deviation of the whole sample in each time step is depicted in the graph (d).

> For an input attribute-value pair $\langle a : x \rangle$, where $x \in R$:
> Retrieve a landmark pool $L_a = \{v_1, \ldots, v_k\}$ for a
> (if it does not exist yet, create a new one with a single landmark in x).
> Winner $v = v_{i^*}$, where $i^* = \mathrm{argmax}_i\,(v_i(x))$.
> If $k < K$ and $v(i) < \frac{K}{k} \cdot \theta_m$ and $v(i) < \theta_I$,
> then add into L_a a new landmark l in x and return l,
> else update v by x and return v.

Figure 4. The pseudo-code of the learning/retrieval algorithm for the primary layer. K is the maximum capacity of the pool, θ_m is the category membership threshold, and θ_I is the identity threshold. Multiplying the threshold θ_m by K/k in the condition for adding new landmarks promotes bigger precision in case there are still spare resources.

$x = \{\langle a : x_a \rangle \mid a \in A_x,\ x_a \in [0, 1]\}$ that describes the resulting activity of the primary layer, the criterion e returns the activity $e(x) = \exp(-c \cdot d_e(x))$, where

$$d_e(x) = \sqrt{\sum_{a \in A_e - G} (x_a - p_a)^2 \cdot \frac{f_a}{N}}\ . \qquad (1)$$

G is the (possibly empty) set of attributes ("ignore list") that should not contribute to the total distance. If we view auto-association as a sort of search or retrieval from a noisy or incomplete input, the set "G" serves as the wildcard in the search. The retrieval mechanism returns the episode that best matches the input in all other attributes; nevertheless, the retrieved episode may also contain attributes in G (and possibly some others), which can be viewed as prediction of their values. Depending on the content of $A_e - G$, the auto-associative retrieval supports the inference of abilities of an agent (if $A_e - G$ contains attributes describing SUBJ theta role), affordances of objects (retrieval from OBJ theta role), simple planning (retrieval from desired ΔSUBJ, ΔOBJ) and others.

The meaning of the formula (1) should be clearer, if we recall that the input to episodic criteria is a configuration of activities of the primary layer. It means that the real values of attributes in the original (primary-layer) input frames are population-coded, i.e. each $x_a \in [0, 1]$ and different real values are possibly represented by different attributes.[12] Hence, attributes that vary much in the input sample will have low values of the quotient f_a/N and they will not contribute to the total distance so much. On the other hand, invariant attribute values will have high p_a and high f_a/N (close to 1), so the criterion will require also high values of x_a, i.e. better match (bigger precision) of a corresponding primary detector.

The learning/retrieval algorithm of the episodic layer works in the following way: For a new input, the most similar memory (the criterion reacting with the highest activity, the winner) is retrieved. If it is similar enough

[12] In (1), the differences are summed over all attributes $a \in A_e - G$. If some of these attributes are missing in the input frame ($a \notin A_x$) the respective x_a are considered to be zero.

For an input frame x:
Winner $e = e_{i^*}$, where $i^* = \text{argmax}_i\,(e_i(x))$.
If e exists and $similar\_enough(e, x)$,
then update e by x and return e,
else add into E a new e_{n+1} in x and return e_{n+1}.

Figure 5. The pseudo-code of the learning/retrieval algorithm for the episodic layer.

to the input, its internal statistics are updated by the input, otherwise a new criterion with the input as its first example is added to the memory (see Figure 5). The test for being "similar enough" can be completely internal, i.e. only based on the activity of the winner, or can also take into account some external pragmatic feedback. The latter approach is important for *ecological validity*[13] of the agents' categories. In our experiments, we approximated the pragmatic similarity with a combination of the fraction of attributes missing (not predicted) in the retrieved episode with the weighted sum of differences in remaining attributes with more emphasis on attributes representing theta roles of actions and their consequences.[14]

We also implemented merging of episodes that have converged very close to each other. The merging test was based on pair-wise comparisons similar to those of the primary layer. Two criteria e_1 and e_2 were merged into a new criterion e in the following way: $A_e = A_{e_1} \cup A_{e_2}$, $N = N_1 + N_2$, for $a \in A_{e_1} \cap A_{e_2}$: $f_a = f_{a_1} + f_{a_2}$ and $p_a = \frac{f_{a_1}}{f_a} p_{a_1} + \frac{f_{a_2}}{f_a} p_{a_2}$ (frequencies and values of the remaining attributes were just copied into the result).

For cognitive economy purposes, we also implemented forgetting of memories that have not been used for a long time with the exception of those that had been used very often. Each episodic criterion had a counter of time steps since the last hit.[15] If this counter exceeded the threshold 100 and the size of the criterion's sample N was less than 50, it was removed from the layer.

3.5. Simulation results

The presented architecture was implemented and we ran simulations of the experimental setting described in the section 3.3. In each simulation, all agents started with no knowledge (blank memory) and had to discover the environmental causality and construct appropriate representations only by performing random actions and by observing other agents. This was not straightforward even in such a simple environment, because the same action could lead to different outcomes in different contexts, e.g. the change of position caused by moving depended on the agent's direction and the presence of obstacles (other objects or the edge of the lattice) on the route.

The utility of the agent's created representation was measured by its ability to predict the results of actions. Before perceiving the result of an observed/performed action (in the form of the role frame with ΔSUBJ and ΔOBJ parts), the agent made a prediction as the auto-associative retrieval from incomplete [ACT, SUBJ, OBJ] input. The prediction was then compared to the actual episode frame. The results showed that the agents very soon achieved sufficient success in predicting actions of one's own as well as the others (Figure 6a). During its lifetime, each agent observed approximately 2000 episodes. The average number of represented and stored types of episodes was significantly smaller (see Figure 6b and 6c), so the agents managed to generalize successfully.

By inspecting the detailed log of the criteria constructed and used during the simulations, we found out the remarkable fact that the agents, when observing actions of others, were often able to auto-associatively supply the unobservable internal state (energy level) of other agents from their own experience in a similar situation (see Figure 7). As the agents tried to "understand" each newly-perceived episode by recalling the most similar remembered type of event, the result of auto-association could be richer in details than the perceived input. We can imagine more complex models than ours, where internal states include needs, motives, goals, affects, proprioception, and even some kind of "thoughts". The auto-associative mechanism we just described would serve as a useful computational equivalent of a simple form of empathy or "theory of mind" in such models.

4. CONCLUSION

We have proposed a computational-level model of continuous and incremental adaptive knowledge acquisition based on performing and observing actions in a virtual environment. In comparison to related neural networks models, our architecture is dynamic, entertains fast learning, and is amenable to internal analysis of the formed representations.

The important property of auto-associative retrieval from incomplete input gives the representation a potential for simple inferences such as determining abilities of an agent (retrieval from SUBJ), affordances of objects (retrieval from OBJ), planning (retrieval from desired ΔSUBJ, ΔOBJ) and others. Hence, although the criteria represent whole episodes, implicit categories of objects, properties, relations and actions are formed as projections of the criterion's receptive field into respective partial subspaces.

As the very same representation is used for storage and retrieval of both observed and performed action, it can be conceived as a computational equivalent of mirror neurons [10], which in our model enables the auto-associative inference of internal states of other agents in particular situations – a computational equivalent of empathy or "theory of mind".

The presented model accounts for individual action-

[13] For example, it is vital to distinguish edible mushrooms from poisonous ones according to the effects of eating them, regardless of their perceptual similarity.

[14] We used $w_{\text{SUBJ}} = w_{\text{OBJ}} = 1$, $w_{\text{ACT}} = 200$, and $w_{\Delta\text{SUBJ}} = w_{\Delta\text{OBJ}} = 500$.

[15] The counter was set to zero when the criterion became the winner, otherwise it was incremented by 1.

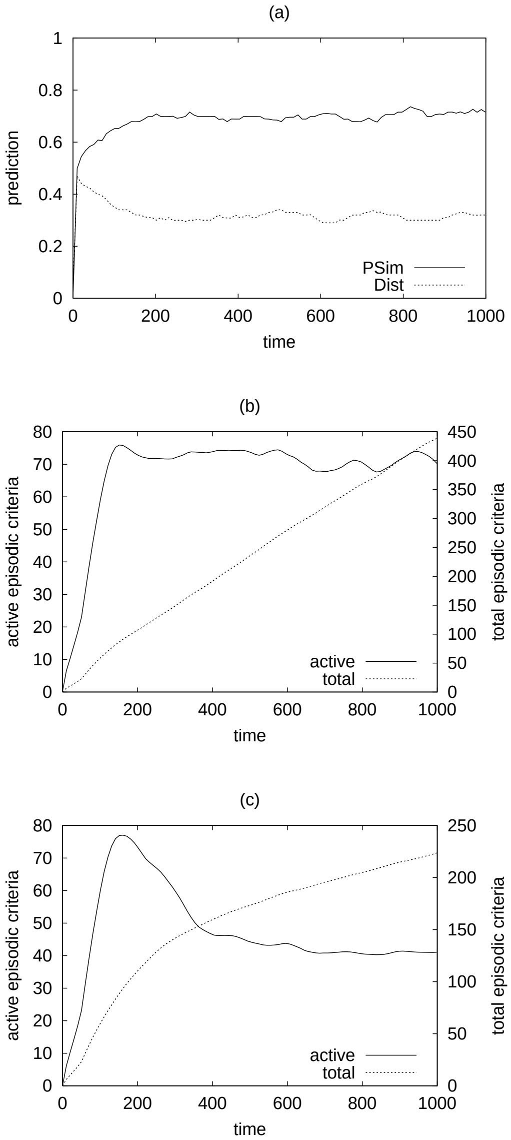

Figure 6. a) Prediction of results of observed actions. The difference between actual and predicted episode was evaluated by “Dist” – the average of differences in individual attributes normed by standard deviations of all their observed values, and “PSim” – the pragmatic similarity as an exponentially decaying function of the weighted distance that highly penalized differences in ΔSUBJ and ΔOBJ parts. b) The number of actively used episodic criteria stabilized around $70-80$. The measure ‘total’ shows cumulated number of all created episodic criteria (including those that have been forgotten). c) The variant of the experiment where objects had low initial nutrition values. At around the time step 200, the agents have already eaten everything, therefore their energy started to fall and after some time they could not move any more. Their world became predictable, the number of actively used criteria needed for interpretation of their world dropped, and the total number of criteria showed a tendency to saturate. (All displayed results were averaged over all 4 agents in 10 simulation runs with different random seeds.)

```
A0 observed (A3 -> O3):
[ACT  ={eat:1; howMuch:4},
 SUBJ ={dir:1; posX:2; posY:0},
 OBJ  ={nutrition:1792; posX:3; posY:0},
 dSUBJ={dir:0; posX:0; posY:0},
 dOBJ ={nutrition:-4; posX:0; posY:0}]

A0 autoassociated:
[ACT  = {eat:1(100%); howMuch:2(50%)},
 SUBJ = {dir:0(50%); @energy:40(46%);
   posX:1(100%); posY:0(100%)},
 OBJ  = {nutrition:1795(98%); posX:3(100%);
   posY:0(100%)},
 dSUBJ= {dir:0(100%); @energy:2.5(45%);
   posX:0(100%); posY:0(100%)},
 dOBJ = {nutrition:-4(99%); posX:0(100%);
   posY:0(100%)}]

Pragmatic similarity = 0.83
```

Figure 7. When observing an action of other agent A3, the agent A0 was able to infer unobservable internal attribute of energy of A3 with pragmatic similarity 0.83. The percentages in parentheses are computed as $p_a f_a/N$ and correspond to degree of certainty of the particular attribute value.

based formation of preverbal meanings. For successful inter-agent communication, the meanings should be mutually coordinated and associated with some signals in a collectively coherent way. The emergence of a common lexicon due to self-organization driven by feedback, e.g. [11], observation [12] or both [13], has already been successfully demonstrated in other models. A straightforward way of extending our model with communication consists in conceiving a speech act as a type of action with parameters describing the content of the speech (signal, utterance). Such utterances will inherently be contextual and pragmatic, by being connected to particular states of the speaker (SUBJ) and the hearer (OBJ), possibly leading to changes of their states (ΔSUBJ, ΔOBJ).[16] We expect that a coordinated communication would emerge by self-organization in the process of mutual observation and imitation. However, experimental testing of this hypothesis remains a topic for our future research, as well as the issue of scaling up to larger populations and more complex environments.

5. REFERENCES

[1] L. Smith and M. Gasser, “The development of embodied cognition: Six lessons from babies,” *Artificial Life*, vol. 11, no. 1–2, pp. 13–30, 2005.

[2] M. Takáč, “Construction of meanings in biological and artificial agents,” in *Agent-Based Societies: Social and Cultural Interactions*, G. Trajkovski and S. G. Collins, Eds. IGI Global, Hershey, PA, in press.

[16] The auto-associative mechanism should then enable prediction/production of different utterances depending on a personal style and affective state of the speaker, or to infer the internal state of the speaker from its utterance in some context.

[3] J. Šefránek, "Kognícia bez mentálnych procesov [Cognition without mental processes]," in *Kognitívne vedy*, J. Rybár, L. Beňušková, and V. Kvasnička, Eds., Bratislava, 2002, pp. 200–256, Kalligram.

[4] M. Takáč, *Construction of Meanings in Living and Artificial Agents*, Ph.D. thesis, Comenius University, Bratislava, Slovakia, 2007.

[5] M. Takáč, "Categorization by sensory-motor interaction in artificial agents," in *Proceedings of the 7th International Conference on Cognitive Modeling*, D. Fum, F. Del Missier, and A. Stocco, Eds., Trieste, Italy, 2006, pp. 310–315, Edizioni Goliardiche.

[6] M. Takáč, "Autonomous construction of ecologically and socially relevant semantics," *Cognitive Systems Research*, in press, DOI: 10.1016/j.cogsys.2007.12.001.

[7] P. Gärdenfors, *Conceptual Spaces*, MIT Press, Cambridge, MA, 2000.

[8] Ch. Balkenius, "Are there dimensions in the brain?," in *Spinning Ideas, Electronic Essays Dedicated to Peter Gärdenfors on His Fiftieth Birthday*. 1999.

[9] T. Kohonen, *Self-organization and associative memory: 3rd edition*, Springer-Verlag, Berlin, 1989.

[10] G. Rizzolatti, L. Fadiga, V. Gallese, and L. Fogassi, "Premotor cortex and the recognition of motor actions," *Cognitive Brain Research*, vol. 3, pp. 131–141, 1996.

[11] L. Steels, "Emergent adaptive lexicons," in *From Animals to Animats 4: Proceedings of the Fourth International Conference On Simulation of Adaptive Behavior*, P. Maes, M. Mataric, J.-A. Meyer, J. Pollack, and S.W. Wilson, Eds., Cambridge, MA, 1996, pp. 562–567, The MIT Press.

[12] A. D. M. Smith, "Intelligent meaning creation in a clumpy world helps communication," *Artificial Life*, vol. 9, no. 2, pp. 559–574, 2003.

[13] Federico Divina and Paul Vogt, "A hybrid model for learning word-meaning mappings," in *Symbol Grounding and Beyond: Proceedings of the Third International Workshop on the Emergence and Evolution of Linguistic Communication*, P. Vogt, Y. Sugita, E. Tuci, and C. Nehaniv, Eds., Berlin/Heidelberg, 2006, pp. 1–15, Springer Verlag.