Cognition, Meaning and Truth

MARTIN TAKÁČ, JÁN ŠEFRÁNEK[1]

Abstract. In this chapter we define a cognitive semantics representation framework. We specify meanings connected to cognitive agents, rooted in their experience and separable from language, covering a wide spectrum of cognitions ranging from living organisms (animals, pre-verbal children and adult humans) to artificial agents and covering a broad, continuous, spectrum of meanings. We suggest the way from subjective to collectively accepted meanings with the help of a construct of instrument. The instrument serves as a tool for inclusion of sophisticated meanings, e.g. of scientific constructs. In zooming out, we analyse different paradigms of cognitive science in terms of their instruments, and position our approach in this context.

1 Introduction

An ability to acquire and use (approximately) true knowledge about the external world is a crucial feature of cognitive agents.[2] In order to understand this ability a deep theory of meanings is needed – a construction of appropriate meanings is an indispensable instrument of knowing. We believe that an increased attention devoted to semantic and epistemological problems is necessary for the interdisciplinary cognitive science.

Traditional semantic theories almost exclusively dealt with meanings of linguistic expressions. Elements of language were either mapped to sets of objects and relations in the world (in extensional semantics, e.g. Tarski, 1933) or to mappings from possible worlds to sets of objects and relations (in intensional semantics, e.g. Kripke, 1963; Montague, 1960). In any case, meanings were seen as something objectively existing regardless of any interpreting subjects. This so called *objectivist* approach has been

---

[1] Department of Applied Informatics, Faculty of Mathematics, Physics and Informatics, Comenius University, 84248 Bratislava, Slovakia, e-mail: {takac,sefranek}@ii.fmph.uniba.sk

[2] By agent we mean any autonomous entity that achieves some goals in its environment by sensing and acting (Kelemen, 2003); this includes virtual/simulated actions in virtual environments too.

criticized by Lakoff (1987)[3] who proposed an alternative called *experientialist* approach. Within this approach, meanings are rooted in experience of physically embodied[4] beings, and this experience is richly structured even before language and independently of it. Lakoff's book has started an entirely new research program called cognitive semantics that no longer places meanings in the outside world. Meanings are conceptualized by notions closer to the experience of cognitive agents, such as *image schemas* and *idealized cognitive models* (Lakoff, 1987), geometrical or topological structures in so called *conceptual space* (Gärdenfors, 2004a), or *force dynamics patterns* (Talmy, 2000). Relation of meanings to language, especially grammar, has been further elaborated e.g. in (Langacker, 1987, 1991; Bergen & Chang, 2003; Feldman, 2006).

Cognitive semantics in its various forms has been around for about 25 years; still it has not given a satisfactory account of many issues. It has been criticized for absence of a satisfactory account of semantics of verbs and sentences/propositions and no theoretical account of how the proposed conceptual structures can be constructed; the proposed structures were intuitively plausible only for a small subset of basic cases and solutions for more complex cases were often described vaguely and in an ad-hoc manner (Wildgen, 1994, chap. 2). In the next two sections we will specify problems that we try to address in this article and a quick view on their proposed solution. After that, more thorough motivation is given.

Our basic problem and goal is to show how the semantic constructs can be developed on the basis of a subjective experience of an agent. We find as interesting to describe how it is possible to integrate purely subjective meanings with intersubjective meanings, meanings accepted by a society, meanings assigned to abstractions created in terms of a language and/or corresponding in a way to an external environment (and how those meanings may coexist).

We propose a solution based on an ability of cognitive agents (and more generally, an ability of living organisms and their parts) to distinguish. This ability is demonstrated also as a selection/construction of a schematic view on a complex chunk of percepts (or more abstract entities). Our basic semantic constructs are schemata and distinguishing criteria, abstractions of the ability to distinguish. Distinguishing has a holistic dimension – a cognitive agent selects and constructs a simplified, but focused on essential and relatively invariant features, view on a complex chunk of inputs. Schemata represent this holistic aspect of distinguishing. A schema is built from some more elementary constructs, distinguishing criteria.

A background idea behind our approach is an assumption that some meanings may be independent of (or even prior to) a language.

Similarly, we view and construct the world of meanings as a continuous one, containing a broad spectrum of meanings, from those which can be ascribed to animals or preverbal infants on one side to meanings, assigned to a language with a rich

[3] See also Gärdenfors (2004a) for an analysis of shortcomings of traditional semantic theories.

[4] Meanwhile, so called embodied approach (e.g. Varela et al, 1992; Barsalou, 2008; Chemero, 2009; Adams, 2010) is becoming a dominant paradigm in modern cognitive science.

syntactic structure on the other hand.

This approach to cognitive semantics was first proposed by Šefránek (2002). Distinguishing criteria conceived as functions were defined as abstraction of the ability to distinguish. Basic types of criteria of objects, classes, properties, relations, situations, changes, and plans were proposed, together with the way from pre-verbal biological roots through semantics of two-word language to language with full syntax and reasoning. The theory was further enhanced with more elaborated situation criteria (Takáč, 2006) and short term (focus, situation, problem, event) and long-term (situation/event types, situation/action rules) distinguishing criteria (Šefránek, 2008). The paper (Šefránek, 2008) also analysed a case study of animal behaviour (Bräuer et al, 2006) in terms of the semantics of distinguishing criteria. Takáč (2008a) described a computational implementation of the framework focusing on autonomous construction of distinguishing criteria in interactions with the environment. The semantics of events was further elaborated and implemented in a computational model (Takáč, 2008b).

The semantic framework of distinguishing criteria has been being developed for ten years. So far we and our students have produced 28 papers, eight master theses and one dissertation thesis elaborating various aspects of the theory and simulating its partial computational models. However, much of our work has only been published in Slovak yet, hence inaccessible to the wider audience. The theory presented here is substantially refined and different from our previous work. A comparison of the presented theory to the previous work is described in Takáč & Šefránek (2012).[5]

1.1 Motivation

Our approach to cognitive semantics starts from subjective meanings (meanings adopted by an animal, by a preverbal infant, by me etc.). Hence, we start from the first person perspective and on that basis we try to reconstruct intersubjective meanings – meanings common to more agents and also meanings which may be understood as objective entities, metaphorically located in the realm of ideas.

We believe that such construction could be fruitful for cognitive semantics: the first person perspective enables a kind of grounding (embodiment) of meanings in cognitions of agents and the reconstruction of intersubjective meanings on the basis of subjective meanings completes the picture. Actually, there are many roles of meanings, from understanding of a local environment of an agent to mutual communication of ideas in a society or to an exactly verified view on the world.

Now a more detailed motivation for our basic semantic constructs follows. Our goal is to propose a semantic framework joining (integrating) all meanings, from purely subjective ones to intersubjective meanings, supported by a somehow codified status.

The central building blocks of the framework – semantic constructs called distinguishing criteria and schemata, cover a broad spectrum of meanings – from meanings which can help us to explain behaviour of animals to semantics of languages

[5] This chapter is a revised and extended version of Takáč & Šefránek (2012).

with rich syntactic structure. The framework enables coexistence of subjective and intersubjective meanings, understanding of different, but mutually close, subjective meanings and a characterization of a development of intersubjective meanings on the basis of subjective meanings.

This ambition is based on a belief that there are no strict boundaries between living organisms. Consequently, there are no strict boundaries between capabilities of living organisms evolved continuously in the nature. There is a continuum of cognitive capabilities in the nature; those capabilities are implemented through cognitive processes, their results can be considered as meanings and, finally and consequently, the world of meanings is continuous, without strict boundaries. This world is inhabited by subjective meanings of cognitive agents on initial stages of their mental development (imagine a little child which denotes also pigeons by the word “dog”), by meanings evolved from those initial ones, by meanings of expressions of languages with different levels of complexity, by meanings acquired in times of the elementary school etc. New strata of meanings are placed permanently over the previous ones. This continuous world of meanings reaches up to meanings of scientific theories.

Meanings assigned to animal cognition deserve an additional explanation. The analysis of animal behaviour leads to conclusions that animals are able to reason and that they have knowledge about the external world (Šefránek, 2008). They observe results of their own actions or of actions of other agents. They distinguish success or failure of actions and learn on the basis of such observations. It can be said that they understand relevant features of the environment. This understanding can be described in terms of meanings.

Even a stronger and more general claim is justifiable. We do not assume that meanings are assigned only to language expressions. To the contrary, meanings are prior, in a sense, to language expressions. An acquisition (and also a development) of a language is possible only if some meanings are sooner acquired by the future users of the language. A little child is able to use a word correctly, to acquire a meaning of a word only after it is able to recognize, to distinguish the corresponding referent or situation in the environment.[6] Similarly, animals are able to recognize, to distinguish some important objects, their properties, situations in the environment without a use of a (human-like natural) language (Bräuer et al, 2006). As a consequence, we believe that an understanding of a stratum of some language expressions is possible only on the basis of some experience with meanings of some more elementary strata of the language. This also holds for understanding of abstract expressions. In that case an experience with abstract objects is required, e.g. we can understand the notion of the (mathematical) derivative only after some (computational and conceptual) experience with the notion of the limit of a function.

1.2 Basic features of our semantics

We emphasize cognitive nature of meanings. Objects of the real world, their properties,

[6] However, we do not deny the influence of language on further shaping of concepts.

classes of objects etc. are traditionally considered as meanings. Meanings in our semantics are embodied in a sense – they are connected to cognitions (cognitive agents), and they are our abstractions of capabilities of cognitive agents. Two important points should be explained in the context of the previous sentence. Meanings are constructed by cognitive agents (i.e., our position is constructivist; Takáč, 2009): if an agent distinguishes something and a meaning is identified with the ability to distinguish, then the meaning is a product of the agent. On the other hand – meanings are not reducible to mental or neural processes. Cognitions are connected to the external environment. Contents of cognitions are dependent on the state of the external world (this can be denoted as an externalist position). As for the "location" of meanings, we locate meanings neither in the agent, nor in the external world. Our theory of meanings constructs some abstract entities, which correspond to the relatively constant or long term patterns resulting from the interactional feedback loops between an agent and its environment and between an agent and other agents.

In this chapter, we are not going to enter debates about true ontological status of meanings; rather we conceptualize them by constructs which enable to explain some observable features of the behaviour of cognitive agents.

We sum up and motivate some important properties of our semantics. A satisfaction of those properties is important, if we want to build a realistic account of meanings used by cognitive agents.

1) First, we emphasize an *evolutive* nature of meanings. The experience of cognitive agents leads to some updates or revisions of their beliefs and notions. Notice that beliefs are in our approach meanings - we do not identify beliefs of agents with a knowledge base in a form of sentences of a language, but we view them as a cluster of meanings. Also extensions of knowledge bases and of the conceptual apparatus of an agent should be considered as an evolution of meanings because of mutual dependencies of pieces of beliefs and of concepts.

2) Further, an *approximate* nature of meanings should be taken into account. Meanings (most importantly, beliefs and also meanings of sentences) are constructions of the agents. Our opinion is that those constructions could be, and often are, improved, précised. The evolutive and approximate natures of meanings are two sides of the same coin. The second one stresses impreciseness of meanings, the first one their development in time, which may sometimes lead to more precise meanings.

3) *Fluent* nature of meanings is something different from both features mentioned above. The world of meanings contains many examples of similar meanings with small continuous differences.

4) Usually, meanings are treated as independent of knowledge. It is argued that knowledge is composed of words and their meaning is given. We believe that *meanings are tightly connected to knowledge*. Recall our opinion that knowledge bases of cognitive agents are constructed of meanings, not of words. When a knowledge base is built, a set of meanings is built. What is important, meanings of words are acquired, modified, made more precise in the context of the knowledge base. If we want to express something more subtle, words are selected in a stepwise way, their meaning is

fluently changed and accommodated in order to reach a satisfiable final or preliminary expression of our evolving idea.

5) Similarly, meanings are tightly connected to reasoning. Cognitive agents need to reason, in order to understand and create meanings.

6) Some meanings are dependent on context, viewpoint and temporary focus of a cognitive agent.

To sum up, we are aiming to build the semantics with evolving, approximate and fluent meanings, which are connected to knowledge and reasoning and dependent on a context.

1.3 Overview

The rest of our chapter is structured as follows. First, basic ideas and constructs of our semantics are described, explained and defined. We start with a conception of a situated agent ("Me") in an environment. A projection of the environment into a cluster of current percepts and a similarity function on percepts are introduced. Subsequently, situation schemata, representations of percepts, are described together with further important notions – more sophisticated similarity functions, knowledge base, event schemata and distinguishing criterion of change. A current state of "Me" is defined as a six-tuple comprising of its knowledge base, percepts, beliefs, desires, intentions and behaviour in a given time point. The last one is observable from outside; the others can be seen from the first person perspective only. After that, we introduce transformers – distinguishing criteria that express transformations of schemata. Special types of transformers, called constructors, construct detectors – a type of distinguishing criteria which represent common characteristics of categories recognized on schemata by "Me". Another type of transformers – updaters keep track of evolving schemas and distinguishing criteria. Note that with the developing semantic apparatus, further kinds of similarity functions are introduced. We introduce several types of detectors (of individuals, properties etc.). Inference rules and action rules are built over this equipment.

The subsequent part of the paper describes a way from subjective meanings (of "Me") to intersubjective meanings. That part starts with a look on an evaluation of subjective meanings with respect to observations of success or failure of actions. After that, the third person perspective (of an agent "It") is described. Only actions – the behaviour of "It", are observable. Meanings accepted and used by "It", its knowledge base with beliefs, desires and intentions can be hypothetically derived by abduction. A construct of an instrument, which represents a measure for using intersubjective meanings, is introduced. The measure is accepted by a group of agents, it is generally accessible and interpreted in a unique way. Finally, a discussion of basic design decisions, philosophical and methodological attitudes behind our approach is presented in Section Zooming out. A summary of main semantic constructs is presented in Appendix A.

2 The first-person semantics of "Me"

The goal of this section is to gradually build semantic constructs as they are seen by the cognitive agent itself. However, on the (meta-) level of presentation, we cannot completely avoid the third-person-type descriptions, as we are hoping to transfer our ideas to the reader in interpersonal communication by words with commonly established meaning. The way from subjective to interpersonally accepted meanings is proposed later in the article.

2.1 Situated agent and its environment

We already mentioned that our semantic framework is cognitive, i.e. we emphasize tight relations of meanings to cognitive agents. It also significantly overlaps with pragmatics, in the sense that meanings are related to knowledge, understanding and reasoning of a particular agent. Usefulness/correctness of meanings can be tested by pragmatic criteria in the real world/environment (external to the agent).

Imagine a cognitive agent situated in an environment *Env*. The agent is coupled with its environment via sensing and acting. The environment is dynamic in the sense it can change from moment to moment based on the agent's actions and other factors (external to the agent) including actions of other agents. We will denote a current state of the environment Env_t (where t stands for a time point).

Currently being performed actions of the agent constitute its observable behaviour Beh_t.[7] We assume that the agent has an internal view on itself – its internal state, memories, knowledge, which are not directly observable from outside.[8] This view (called "Me") is described in more detail in the following section. The agent is dynamic too, as its internal state and knowledge are changing in time (shaped by its experience).

2.2 Percepts

In any moment, the agent's perception of the environment is mediated via its senses. So, the agent views the environmental state as a collection/cluster of current percepts $P(Env_t)$. In this sense, P is a projection function (projecting the environment into the agent's internal perspective) but also a *selection* function: what exactly is projected is

---

[7] Actions usually do not last an instant but a time interval. In this article we abstract away from temporal issues and simply assume that the same action will (re)appear in the Beh_t set for all t in its time span.

[8] For the moment we put aside the question whether the agent can *consciously* access *all* its knowledge, memories, drives etc. Unconscious aspects of the agent's experiences, embodiment, etc. (if any) co-determining its decisions and behaviour can be viewed from the agent's perspective as non-deterministic aspects of its cognitive processes.

determined by the agent's embodiment and physical limits,[9] its past experiences, its current mental state and focus of attention, etc.

However, we do not ascribe to *P* much of a sense-making; this is applied to $P(Env_t)$ afterwards. $P(Env_t)$ contains rather crude (low-level) percepts forming *iconic representations* in the sense of Harnad (1990). Iconic representations allow for *discrimination*, i.e. being able to tell if the things are different/similar, and possibly how different/similar they are. We formalize this subjective discrimination ability by a similarity function sim_d. In the first approximation, sim_d operates on percepts and is able to detect perceptual similarities/differences; later we extend the agent with more sophisticated similarity functions.

2.3 Situation schemata

The similarity function enables the agent to recognize common patterns among recurring percepts and gradually extract (holistic) schematic views of their relations. In people (and probably other embodied agents too), basic schemata[10] arise directly from recurring sensorimotor experience early in development[11] (Piaget & Inhelder, 1966) and more complex ones are gradually built on top of these. Cohen et al. (1966) describe how different levels of schemata (perceptual redescriptions) can be learned based on detecting statistical contingences among perceptual streams (e.g. inferring a concept of an *object* as time-locked correlations of percepts in different sensory streams – a sort of a multimodal integration; see also Smith & Gasser, 2005). Schemata allow the agent to make sense of its current perceptions by establishing their relation to previous experiences (by recognizing similarity and evoking memories) and, more generally, integrate the new experience within the web of existing knowledge (expressed by schemata). This corresponds to Piagetian process of *assimilation* (Piaget & Inhelder, 1966).

In this sense, a sense making act σ – *signification* (Peirce, 1931-58) of the agent is a process of constructing or evoking appropriate schemata, given the current percepts $P(Env_t)$. We will denote the result of signification $\sigma(P(Env_t))$ and call it *situation schema*. Unlike percepts that are pure transductions of the external environment, a situation schema is a *representation* with the added value of interpretation of percepts (Gärdenfors, 1996). A situation schema can be formally represented by a labelled graph with percepts in vertices linked by edges expressing their mutual relations. More precisely, only some vertices correspond to percepts; other express inferred constructs. For example, if the agent recognizes percepts in multiple modalities as constituting a single object, the graph will contain a separate vertex for this object, with all its percepts linked to it by edges of an appropriate type. The type/semantics of an edge is

[9] See a convincing description of different *Umwelts* (subjective worlds) of different animals by Jacob von Uexkull (1957).

[10] These so called *image schemata* include basic spatial and topological relations, goal-directed movement etc. (Lakoff, 1987; Johnson, 1987).

[11] Some basic schemata may be innate.

represented by its label. The object vertex can further be linked to other schemata in memory, recognized/evoked as similar or related in some aspect to this object (e.g. recognizing this object as my dog). Sometimes a relation is so complex it is best expressed by a schema of its own; in that case, a (n-ary) relation is represented by a (n+1-ary) hyperedge[12] with one vertex serving as a handle/access point to another schema. So we can see that the schema can contain vertices of various types. The type of a vertex is expressed by its label. We allow multiple labels for vertices and edges; these can be interpreted as different views on the same situation. Formally we can organize labels in layers (thus creating a layered graph) or we can see the layers as separate schemata linked together (by establishing similarity/identity relations among the corresponding vertices and edges). Later we will define means for transformations among schemata.

In order to establish a relation to previous experiences, the agent needs to maintain some sort of memory. We will call the agent's long term memory its *knowledge base* KB_t. The knowledge base is a set that includes the agent's remembered situation schemata - a subset[13] of $\{ \sigma(P(Env_i)) \mid i < t \}$ (we will gradually extend the definition of KB_t with other constructs). The knowledge base also contains a set of similarity functions (without going to details, we assume that the agent gradually learns to use functions for detecting similarities/differences among schemata, derived from the most elementary sim_d that operates on percepts).

2.4 Event schemata

The world around the agent is dynamic; situations change to other situations. A change of one situation to another constitutes an *event*. Being endowed with similarity functions, the agent is able to perceive temporal changes in situations. We describe this ability by a construct of a *distinguishing criterion of change.* We formalize a distinguishing criterion of change as a function defined on pairs of the form $(\sigma(P(Env_{t-1})), \sigma(P(Env_t)))$; if the second one is a result of a change of the first, the assigned value is 1.

The agent represents distinguished events by *event schemata*. An event schema consist of two or more situation schemata linked by (hyper)edges labelled by distinguishing criteria of change. Event schemata can be constructed or evoked from memory (in case of recognition of a similarity to a past event). We will denote the act of event selection ε and its resulting event schema $\varepsilon(\sigma(P(Env_t)), KB_t)$. We will also extend the definition of the knowledge base to include event schemata

$$KB_t := KB_t \cup \{ \varepsilon(\sigma(P(Env_i)), KB_i) \mid i < t \}. \qquad (1)$$

[12] A hyperedge is an edge connecting more than two vertices. A graph that contains hyperedges is called *hypergraph*.
[13] Some situation schemata may have been forgotten.

2.5 Current state of "Me"

So far, the agent's current knowledge base is described as a bag of interlinked schemata of situations and events. However, schemata do not have a uniform status at each moment: some of them describe the interpretation of the current/recent situation/event; others are related or associated to it, yet others are "inactive" at the moment. Some are attended to or focused on, others are not. Moreover, the agent can be in the middle of executing a plan or pursuing a goal. A *goal* of an agent can be expressed as a situation schema of a desired situation. A *problem* or a question can be expressed as a situation schema (perhaps with special vertex/edge labels) too. The agent needs to distinguish what a particular schema represents in a moment – its particular *autoreflexive attitude* toward the schema. In the first approximation, we imagine the autoreflexive attitudes are represented by special labels on (elements of) schemata. Current autoreflexive attitudes temporarily give some of the schemata in the knowledge base a special status. These schemata can be further factorized to a current set of the agent's *beliefs* B_t (schemata of currently perceived situation/event), a set of *desires* D_t (schemata of the agent's needs and long-term goals), a set of *intentions* I_t (schemata of the agent's current goal, a plan to achieve this goal together with a state of its execution, and other agenda-related structures).[14] A current state of "Me" can be defined as

$$Me_t = (KB_t, P(Env_t), B_t, D_t, I_t, Beh_t), \quad (2)$$

where only the agent's overt behaviour Beh_t is observable from outside; all other structures can only be seen from the first person perspective.

2.6 Transformers

We have said that situations and events are related in various ways. Initially (while the agent's similarity function does not go much beyond crude holistic "same/different" perceptual similarity judgements), the agent's knowledge base will mostly contain holistic "snapshots" of its experiences (schemata with a few basic labels). Later, when the agent has accumulated sufficient number of exemplars, it can extract their common/invariant features, etc.[15], which results in more complex similarity functions and a richer repertoire of labels. Simpler schemata can be *refined* – transformed to more informed ones by adding new layers of labels, simplified (*zoomed out*) by removing labels, linked to other schemata by *associations*, pruned by attention shift or focusing on a particular detail (*zoomed in*), merged (*abstraction*), etc. (Šefránek, 2008).

We will formally describe the agent's ability to distinguish (and perform) these (and other) transformations on schemata with a construct of *transformer*. A transformer is a type of distinguishing criterion that expresses transformations of schemata: it has

[14] See also the BDI architecture (Rao & Georgeff, 1991).

[15] The research in machine learning and artificial neural networks has yielded many good ideas how to extract knowledge from examples by mostly uninformed statistical calculations (Haykin, 2008; Murphy, 2012).

both a declarative aspect (as a description of relations among schemata) and a procedural aspect (as a device that transforms a schema into another schema).

A special type of transformer is called *updater*. The concept of updater expresses the idea of evolutive nature of meanings: If some of the agent's meanings change in time, the agent can keep track of this change by using an updater that will take the schema of the old (original) meaning and connect it to the schema of the new (updated) meaning by a specially labelled edge. The same holds for updates of distinguishing criteria. A schema with a single node labelled by an original distinguishing criterion is linked by a specially labelled edge to another schema with a single node labelled by the new (updated) distinguishing criterion. This mechanism helps to preserve the identity of (evolving) meanings.

2.7 Detectors

By noticing recurring patterns and similarities, the agent can start grouping together situation and event schemata recognized as similar in some respect (i.e. by some similarity function). These groups of similar exemplars constitute elementary *types of situations/events*. Extracting common features of the exemplars can in turn lead to construction of more sophisticated similarity functions which can be used to factor schemata into categories.[16] Special transformers called *constructors* operate on sets of schemata (exemplars) and construct a new distinguishing criterion representing their common characteristics, called *detector*.[17] Internally, a detector consists of a schema specifying a template with features important for category membership (in some cases more or less abstract representation of a prototypical, salient or most frequent category member) and a similarity function specifying how important the particular features are and how they contribute to the overall similarity. Functionally, a detector can be formalized as a partial[18] function that operates on (fragments of) schemata and returns their degree of membership in the implicitly represented category (either as 0=no, 1=yes, or by a fuzzy value from the closed interval [0,1]).

A detector operates on schemata (or their elements – vertices and edges) and is able to distinguish not only its constituting exemplars, but also generalize to other similar schemata. Some detectors distinguish *situation types* (e.g. a traffic jam) and *event types* (e.g. a car crash), others distinguish their elements – *objects/individuals* (such as Barack Obama), *classes* of objects (dog, stone, food), *properties* of objects (red, big, hairy), *relations* between/among objects (bigger than, ancestor), *changes* (grow, faint).

[16] In the past, we have successfully formalized and implemented distinguishing criteria as similarity functions each with their own Mahalanobis metric with parameters induced from statistical characteristics of the exemplars (Takáč, 2008a).

[17] Constructors can also modify an existing detector when new exemplars arrive.

[18] The function only returns a value for some inputs; it is undefined for others, which can be interpreted as a "don't know" value.

2.8 Inference and action rules

Having defined schemata, transformers and detectors, we can revisit the signification and view it as an iterative process; for example the situation schema of a woman with a dog can initially consist of two unidentified objects (linked with their percepts), perhaps linked together by an unlabelled edge. Fragments of this situation schema will then be recognized by detectors vaguely distinguishing dogs and women. Hence, the object vertices will be appropriately labelled by or linked to the detectors. Another detector can recognize their spatial configuration, so the edge connecting the objects will be given a new label too.

This can in turn trigger further transformations on the situation schema, depending on the current context (the current state of "me"). The agent can keep track of sequences of transformations typically occurring in certain situations and extract this knowledge in the form of *rules* – schemata connecting *premises* (prerequisites – the rule's applicability conditions represented by distinguishing criteria of situation and event types) to *consequences* (represented either directly by situation and event schemata or indirectly by transformers that can be applied to the current situation/event and construct the resulting one), optionally with *justifications* (situation and event types guarding the evidence that would prevent the application of the rule in case of *default rules*; see Šefránek (2008) for more details). Some rules specify dynamics of internal transformations (so called *inference rules*), others specify the effects of overt actions on the environment (so called *action rules*, see the next section). Rules can be chained together in the form of *plans*, presumably leading to satisfaction of a goal. The agent can keep track of success/failure of a plan in the past. Remembered successful plans are called *routines*.

3 Towards intersubjective meanings

3.1 Meaning and behaviour

A first important step on the way from purely subjective meanings to intersubjective meanings is described in the following paragraphs.

Assume that a cognitive agent ("Me") equipped with subjective meanings only observes results of its own actions or of actions of other agents. "Me" distinguishes success or failure of actions and learns on the basis of such observations. "Me" evaluates its own behaviour and gets a kind of distinguishing of something what can be regarded as truth – expected results of its actions can be considered as predictions, some predictions are satisfied (true), others are not.

We will describe how such observations lead to objective meanings, more precisely, how some subjective meanings induce behaviour and how "Me" can assign truth to some schemata.

It was stated in the previous section that some transformers trigger overt behaviour. Actions are represented by complex schemata – action rules. Their consequences are transformers which assign a schema representing a resulting situation to the current situation schema. Those transformers may have for some agents a rather complex structure. They may realize a short-term mental operation – an imagination of the action, a specification or a recall of the required effects of the action and, finally, firing the action. The change specified by the transformer is an expected (predicted) result of the action and it is expected that the result complies with the specified effects of the action.

Let us describe in more detail how an action rule is selected, fired and how its result is evaluated. “Me” non-deterministically selects some desires (represented in its knowledge base by a distinguishing criterion or a schema), transforms them using some transformers onto intentions and subsequently other transformers are used in order to map intentions onto actions (members of *Beh*).

However, triggering (an attempt to do) an action is essentially a complex trial and error procedure, which comprises learning of prerequisites and effects of the action (operations on situation schemata) and evaluating success/failure of the action. We will describe it in terms of our semantics.

Assume an agent that connects an action rule with a distinguishing criterion of a required change (a goal, a required effect of the action). If the corresponding action was executed, then the premise and consequence of the corresponding action rule may be modified according to the current situation schema and the current change of the situation schema by the action.

If an action should have been executable in a situation (according to the premise of the corresponding action rule), but the attempt to execute it failed, then the agent modifies the premise of the corresponding action rule. There is a variety of possibilities how to modify it (Čertický, 2012), but we will not discuss them here.

What is important here, an evaluation of an action rule is based on a comparison of situations (the premise of the rule vs. the situation in which the action was executed; the consequence of the rule and the required effect vs. the real effect of the executed action).

The comparison is described in our semantics in terms of a similarity function. An application of this function, even if it is a subjective distinguishing criterion, enables to evaluate (subjective) meanings with respect to the results of a behaviour in the external environment and to reach a kind of understanding and of an (approximate) truth (or falsity) of prerequisites, effects and action rules, which is dependent on the external environment via the success or failure of observations.

Reasoning capabilities (some transformers, some rules) can be tuned in a similar manner.

A final remark – besides rules of the structure described above, other complex schemata, such as modalities, deontic constructions, more complex generalizations, etc., are also construable on the basis of situation or event schemata. However, we will not discuss them. As regards the truth or falsity point of view, some actions can serve as tests of the (approximate) truth.

We believe that an evaluation of a success or a failure of actions in an environment enables a stepwise more precise comparison of subjective meanings and a more precise approximation of truth.

Now, when we are equipped with a notion of an approximate truth, we can proceed to a kind of the third person perspective.

3.2 The third-person perspective

The third-person agent, observable from the viewpoint of "Me" may be represented on the basis of pure observations as $Ag_t = (Beh_t)$. We can – and will - use "It" instead of "Ag".

"Me" considers actions of other agents as events. Suppose that "Me" observes an action of an "It". The current situation and the effect of the action are observable by "Me". On that basis an abduction of action rules of "It" is possible. Similarly for an inference of its *P*(*Env*), *B, D, I,* i.e., *KB*, by "Me". Notice that the results of this inference are not in general identical to subjective meanings of the agent "It" (to emphasize this difference, we mark the inferred structures with the apostrophe ('). We will call them an *external view on subjective meanings*.

Thus, we can specify a derivable third-person agent:

$$It'_t = (KB'_t, P'(Env_t), B'_t, D'_t, I'_t, Beh_t), \tag{3}$$

also indexed by the agent if needed.

In general, an external view on distinguishing criteria and schemata of other agents may be specified in terms of distinguishing criteria and schemata of "Me". We can say that "Me" creates a "theory of mind" of other agents.

Some similarity functions enable to identify similarity of subjective meanings of one agent in two different time-points, of distinguishing criteria corresponding to different sensual inputs etc. Most importantly, they enable to compare Me's external views on subjective meanings of two different (third-person) agents. "Me" can also compare its own subjective meanings and its external view on subjective meanings of other agents.

Thus, a relation of a close neighbourhood (or of an approximate identity) of two distinguishing criteria or schemata is created for high values of a similarity function. The approximate identity specifies a chunk of distinguishing criteria or schemata and enables a step from subjective to intersubjective meanings.

3.3 Other steps towards intersubjective meanings

In this section a brief survey of some possible conditions leading to intersubjective meanings is given.

Similarity functions and their impact on creating close neighbourhood relations represent our attempt to include autoreflexive reasoning into our semantic constructions. Autoreflexive attitudes were discussed in Section Current state of "Me". It was noticed that the simpler way how to specify autoreflexive attitudes were labels. Autoreflexive reasoning implemented in terms of similarity functions and close neighbourhood chunks is a more advanced form of autoreflexive attitudes.

In the preceding section we described how this kind of autoreflexive reasoning can enable a transfer from subjective to less or more intersubjective meanings. In general, we consider autoreflexive reasoning an important step towards intersubjective meanings. It is well known that autoreflexive reasoning enables to create hypotheses about the mental states of other agents – a theory of mind (Apperly, 2010).

Consider communication and cooperation of agents (without a language capability). Again, observations of success or failure of some actions fired in a process of communication/cooperation lead to mutual tuning of meanings – rules, situation and event types, distinguishing criteria (Gärdenfors, 2004b).

Next, we note that there are physical conditions for acquiring similar meanings, i.e., agents with similar "bodies" (similar anatomic, physiologic and genetic dispositions are determined to have similar subjective meanings, if they live and act in an environment of a common type.

Finally, we mention an exceptionally effective role of a language on the way to intersubjective meanings. A detailed investigation of this topic is one of our future goals, but it should be noted that most of our past works were devoted to the distinguishing criteria semantics in a relation to a language in general (to languages with different levels of complexity) or to a language acquisition (see e.g. Šefránek, 2002; Takáč, 2006, 2008a, 2009).

Our attention is focused on a semantic treatment of verbs and sentences in order to overcome simplifications of logical or linguistic semantics. A way based on schemata of situations and events is proposed. As a consequence, we can characterize a situation based meanings of some sentences without a clear reference to some external objects.

Finally, it should be noticed that a plenty of meanings (distinguishing criteria and schemata) are introduced in terms of a language. We can speak about intersubjectivity modulo vagueness of a natural language.

3.4 Instruments

In this section a tool is introduced which models an intersubjectivity of meanings beyond the limits of natural language with an inherent vagueness. However, it should be noticed that a natural language has a potential of bootstrapping such levels of intersubjectivity which overcome a common use of the natural language.

We model intersubjective meanings (distinguishing criteria and schemata) in terms of a measure, which is generally accessible, interpretable in a unique way and accepted by a group of agents. We will call it *instrument.*

Some comments are needed. First we focus on the *acceptance by a group* of agents. Dogmata recorded in some texts with an officially codified status and interpretation may be accepted by a group of agents, but not by another group. This is not only the case of dogmata; measurements were instruments verifying truth of geometrical claims for old Egyptian experts in geometry. A proof of geometrical claims was an acceptability instrument for ancient Greeks.

A selection of an instrument may be considered a cognitive *paradigm*. Let us consider Elements by Euclid (2006). We may assume that Euclid believed that his own axiomatization of geometry is an embodiment of a pattern of human thinking, and he chose this pattern as a paradigm for a presentation of the knowledge of geometry.

Second, a general *accessibility* of an instrument is a natural condition – if an instrument should play a role of a tool of intersubjectivity for a group of agents, then an access to the instrument for each member of the group must be guaranteed.

Third, an *interpretation* of an instrument in an *unambiguous* way is an important condition, which requires a deeper analysis.

At least two levels of this condition may be distinguished. An interpretation of the instrument may be based on a *mechanical procedure*, on an *algorithm* as an extreme case, which evaluates the value of the instrument for given inputs. A simple example of such instrument is a multiplication algorithm or a cooking recipe (we will discuss examples in more detail below). However, there is also a less strict possibility. A group of agents is equipped with advanced knowledge and (reasoning) methods, which enable answer questions reliably. Distinguishing of a malign tumour by a histologist is an example of this. In an ideal case, all (good) histologists should diagnose a case of a malign tumour equivocally.[19]

Let us proceed to a more formal account of instruments. A function, which represents a distinguishing criterion equipped by an instrument, has an additional argument, which denotes the instrument. The value of the function is computed (determined) according to the instrument. The weight of an object may serve as an example. An example of a distinguishing criterion with a non-algorithmic instrument is an atlas of mushrooms.

Instead of a subjective similarity function and an induced close neighbourhood relation of distinguishing criteria, thanks to instruments we can obtain exact transformations between distinguishing criteria, e.g. from kilograms to pounds.

It is obvious that distinguishing criteria are made more precise by instruments.

Schemata with instruments require a more elaborated description. We start with an example. Imagine a situation type, which represents the multiplication operation on natural numbers. The schema may contain a ternary hyperedge assigning a result to two operands. The role of vertices (operand or result) is specified by a label.[20] In general, labels may specify different roles of vertices connected by a hyperedge in an arbitrary schema. A finite set of correct (true) instances of this schema may be generated by an instrument – a transformer associated with the well-known table.

The infinite set of all true instances may be generated, e.g. by a recursive definition of the multiplication. The table and the recursive definition play the role of instruments in our semantics. Both the table and the recursive definition are parts of the

---

[19] However, in fact this condition is rarely satisfied. With non-algorithmic instruments, there is always a possibility of alternative (mis)interpretations. In our example with a case of malign tumour, all interpretations that misdiagnose a malign tumour as benign are considered incorrect.

[20] Depending on the labels specifying which vertices have numerical values assigned, the same schema can be used for multiplication, division, or checking the truth of the corresponding statement.

knowledge base. The first one can be represented as a set of hyperedges connecting three vertices labelled by two operands and one result. The second one is discussed as follows. Our goal is to represent the following two equations by a transformer and an associated situation schema:

$$x.0 = 0 \qquad (4)$$

$$x.\mathrm{s}(y) = (x.y) + x \qquad (5)$$

The schema may contain two hyperedges: one with two occurrences of vertices labelled by 0 and one occurrence of an unlabelled vertex. This hyperedge corresponds to formula (4) and represents the base case of the recursion. The second hyperedge corresponds to formula (5) and represents the recursive case. It connects an unlabelled vertex (corresponding to *x*), then a vertex (corresponding to *s*(*y*)) connected by an edge to the access point of another schema, which assigns a predecessor to a given number, and, finally, the third vertex (result) connected by an edge back to the (access point of the) multiplication schema and by another edge to the access point of an addition schema. The transformer realizes a recursive algorithm for an arbitrary pair of natural numbers and generates a situation schema – a true instance of the schema of the situation type, e.g. an instance that contains a hyperedge with vertices labelled by 2, 3 and 6. For example, the transformer first performs pattern matching that reduces the problem to series of more elementary problems (2.2, 2.1, 2.0) and finally it halts on the case 2.0=0. On the way back, it computes the series of additions (0+2+2+2=6).

A decision about a malign tumour by a histologist was mentioned as an example of a non-algorithmic instrument. We can imagine the instrument used by a histologist as a situation schema with a vertex labelled as tumour and a set of edges with target vertices labelled by the relevant histological properties of malign tumours. Some other labels may be assigned to those vertices – they contain a description of the corresponding property in a language. Moreover, some other means of expression may be used: e.g. some properties are optional, some obligatory (this corresponds to a possibility to introduce partial properties which were discussed before). This expressivity may be added by operators labelling the corresponding edges. Sometimes also some (generalized) quantifiers applied to a set of edges might be used: for example, at least m of n properties should be present (general and existential quantifiers are special cases).

To sum-up: A *distinguishing criterion with an instrument* is a function with a parameter that specifies how to compute its value for its arguments. The parameter is called *instrument* and it is a transformer. The transformer is either an algorithm or a conventional, more or less mechanical, procedure based on an expert knowledge. In the latter case, the expert knowledge is expressed by a set of schemata associated with the transformer (as its additional arguments). A schema generated by transformer and a set of associated schemata will be called *schema with an instrument*.

Some final remarks: A precise notion of an identity of meanings can be based on transformers defined on instruments. Sometimes rather subtle tools are needed.

An optional specification of a group of agents can be added as an argument to a distinguishing criterion with an instrument.

A specification of a group of agents in a schema may serve as an example of meta-level features of schemata, e.g. a schema may be connected by an edge labelled e.g. as "owner" to a vertex labelled by an identification of a group of agents.

4 Zooming out

4.1 Our position

Building on our previous works, we have proposed a semantic framework with meanings connected to cognitive agents, rooted in their experience and separable from language, covering a wide spectrum of cognitions ranging from living organisms (animals, pre-verbal children and adult humans) to artificial agents (softbots, robots, multi-agent systems etc.) In this article, we substantially revised our previous conception of distinguishing criteria (added transformers and constructors), enriched the framework with schemata, knowledge base, belief – desire – intention structures and other constructs (for their full list, see Appendix A). The presented list of semantic constructs is open – a more detailed and more fine-grained extension of our conceptual apparatus is expected as a result of a future elaboration of our approach.

An interesting property of the proposed semantics is that it enables coexistence of subjective and intersubjective meanings. Subjective (the first person perspective) meanings are primary, and we have shown the way from them to collectively accepted (the third person perspective) meanings via observable behaviour and feedback about success/failure of actions and instruments. We have defined the notion of truth in a similar way. This is a novel and previously unpublished contribution.

Now we position our stance within the broader context of cognitive science.

Are we representationalists? We are definitely conceiving/describing cognition in terms of representations in a similar way often used in works of Gärdenfors (e.g. 1996), wherein the representation is something that *adds information* to sensory input obtained by the psychophysical transducers. However, we do not ascribe representations any ontological status; they are just descriptive tools with certain explanatory power.

Are we symbolists? We often formalize meanings in terms of functions, graphs and even rules. However, our story starts with iconic projections - percepts, similarity functions, detection of correlations and invariant patterns, which opens room for all kinds of statistical techniques including artificial neural networks. We believe that by grounding meanings in perceptual inferences, coupling them with actions and feedback of their success/failure and suggesting the way how they can be acquired and further evolved, we avoid the symbol grounding problem. Although we deliberately leave out the details of implementation, a possible implementation would probably rank among hybrid architectures.

Are we objectivists? No, we do not believe that meanings are out there in the world or that there exists a unique correct way of conceptualizing the world. Meanings in our conception are constructed by individual agents, hence have a cognitive nature. However, we commit to what Lakoff (1987) called *internal realism*, in that: we believe in existence of a real world and the link between the world and meanings via real human experience constrained by the world, and in the concept of truth based not only on internal coherence, but on coherence with our constant real experience.

Are we Platonists? As we believe that meanings are constructed, they cannot exist in a realm independent of cognitive agents. However the condition of coherence with a real experience prevents them from being completely arbitrary. Some of them reflect characteristics of external environment; others are constrained by implicit properties of "conceptual environments" constructed by cognitive agents (such as mathematical concepts "waiting to be yet discovered").

Are meanings mental or neural processes? Mental and neural processes certainly participate in the process of construction and further processing of meanings. However, due to the existence of important feedback loops with the environment via the agent's experience and due to requirements of coherence within a system of meanings, we do not identify meanings with neither neural, nor mental processes.

Relation to embodiment and dynamical systems perspective. Particular embodiment of a cognitive agent determines its meanings in an important way. Meanings grow out of direct perceptual and motor experience (processed by simple similarity functions and detection of repeating patterns and mutual contingencies in sensorimotor streams) and more complex/abstract meanings are built on top of the former. Dynamical systems can provide the hint toward "localization" of meanings. Meanings are neither in the agent, nor in the world, but are the relatively constant or long term patterns resulting from the interactional feedback loops between an agent and its environment and between an agent and other agents. Similarly, we can apply the metaphor of dynamical systems w.r.t. interactions of agents and symbolic structures (formal systems). A cognitive agent (a person) interacts also with a world of abstract entities – symbolic structures are constructed by people, but those structures form a system with fixed internal relations. If an agent intends a development of such system, she/he has to respect its laws and properties.

4.2 Instruments in science

In the rest of this part, we zoom out on cognitive science and science in general using the optics of our semantic framework, namely the instruments.

We already mentioned that a selection of an instrument can be considered a cognitive paradigm. Paradigms in science are not only characterized by their background assumptions and core beliefs, but also by their methodology or by specifying what inference/verification methods and tools are considered "correct". As we mentioned, for ancient Egyptians such a method was measurement, while for ancient Greeks it was a concept of a (theoretical) proof. Using a more recent example, most of the grand theories of personality in psychology of the 19th and early 20th century were based on case studies and on brilliant and detailed observations of small number of individuals. A now dominant paradigm in psychology often rejects such theories as speculative and demands a quantitative “proof” using its own instrument – a statistical significance test. A hypothesis is considered “true”,[21] if there is a significant effect with $p<0.01$ (or another arbitrarily specified number). Going yet one level higher, the whole western science is based on Popperian notions of experiment and falsifiability.

Instruments are not only cognitive – immaterial, but often take forms of material tools or devices. As Kováč (2000b) pointed out, the science is not revolutionized by new hypotheses, but by new methodologies enabled by new types of instruments. Similar to our senses that select from the world what is relevant for us, our devices delineate the space of our scientific inquiry. “*The devices with knowledge built in their construction determine what in the world will we notice and how we will interpret it.*” (Kováč, 2000b) Or as an old proverb says: “When you have a hammer, you see nails everywhere”. Improvement of telescopes influenced astronomy and invention of a microscope changed biology. Non-invasive brain imaging methods have given rise to neuroscience and significantly changed cognitive science, its emphasis and focus of attention (the questions of neural substrate had been previously considered irrelevant partially because there were not good enough tools for detailed study of the brain).

Science has given rise to technologies that help us to manipulate the world and achieve our goals more effectively, but the tools, devices and technologies that we use also have a huge backward impact on our culture and everyday life, considering for example mobile devices, internet, or even quick and effective information search tools, such as Google. Some scientists warn that the otherwise fruitful coexistence of science and technologies could lead to over-instrumentalization of science, which will even more increase the gap between “how much we can and how little we know” (Kováč, 2000b).

4.3 Instruments and cognitive science paradigms

In this part we will briefly describe some of the cognitive science paradigms by their instruments. A cognitive paradigm can also be characterized by models or metaphors it uses (note that they are instruments in the sense of our definition). The human mind/brain has always been compared to or metaphorised by the most sophisticated human-made artefact of current times. It was likened to hydraulic systems, clockwork, steam engine, telegraph, etc. (Daugman, 1993). Then it was likened to a computer, which gave rise to information processing psychology and cognitivism in the half of the 20th century. The characteristic beliefs of that paradigm were irrelevance of implementation/body, emotions and social context for cognition (Garnder, 1985). Cognition was described and studied in terms of representations and processes operating on them. Turing machine and other automata were adequate tools of problem formulations, even in philosophy of mind (Putnam, 1967).

Last quarter of the 20th century can be characterized by a shift from the metaphor/instrument of a machine to the metaphor of an *organism*. Advances in

[21] For analysis of typical misconceptions of statistical significance see Cohen (1994).

general systems theory (Bertalanffy, 1968), the revolutionary work of Maturana & Varela (1984), seminal book of Lakoff (1987), Nouvelle Artificial Intelligence (Brooks, 1991) all independently emphasized the role of body and environment in cognition and continuity of human cognition with its biological predecessors. Automata and algorithms as descriptive tools were complemented with dynamical systems perspective (Ward, 2001) emphasizing *interaction dynamics*. A cognitive agent is no longer passively representing an independent external world, but is actively enacting or co-creating its own meaningful subjective world (Stewart et al, 2011). In fact, an *independent* world ceased to exist, at least as an epistemic category. Intersubjective meanings are not the reflection of a God's eye view, but a result of mutual coordination in a shared environment.

4.4 Intersubjectivity and objectivity

Our definition of instrument emphasizes its true nature: it is a conventional construct accepted by a group of agents. However, the conventional aspect is constrained by a requirement of empirical confirmation or real applicability of our instruments. Our instruments are not completely arbitrary. Anyway, some alternative conventions are possible, thus conventional nature of instruments prevents to speak about purely objective character of our knowledge.

As it often happens, the group of agents forgets about intersubjective nature of their paradigm and conceives it as *the* objective reality. Can there be a universal "Truth", or just group truths verified by the tool of their conventionally accepted instruments? In spite of complex cultural strata of human cognition and their autonomous dynamics (Kováč, 2000a), we see the answer in returning back where we started emphasizing the biological roots of meanings: a final reality-check of our meanings and conceptions is success and failure in achieving our goals.

What are our goals then? As living/autopoetic units, our goal is by definition maintenance of our onticity. Less trivial option is suggested by Kováč (2000a, 2003) who emphasizes the role of emotions and phylogenetically accumulated knowledge in our bodies and says that our rational conscious cognition is just a top of the iceberg or "a monomolecular layer covering the ocean of adaptive unconsciousness" (Kováč, 2003). According to him, humans are hyperemotional beings and their ultimate goal is minimization of suffering. Now a hyperbole: the relation between scientific paradigms and minimization of suffering is often not clear at all, but enriching our instruments by affective values and bringing feelings (and perhaps aesthetics too) in science not only as an object of study, but as a research tool can be an interesting experiment. Together with Lakoff (1987), we can conclude that having in mind our biology and our multiple subjective perspectives is the proper way to outreach for intersubjectivity and get closer to truth.

Acknowledgement: We are grateful to Pavel Cmorej for his insightful comments on our semantics. This research has been supported by the grants VEGA 1/1333/12

(Knowledge representation for ambient intelligence) and VEGA 1/0439/11 (Constructivist modeling of early cognitive development in an embodied cognitive agent).

References

Adams, F.: Embodied cognition. *Phenomenology and the Cognitive Sciences,* 9(4), 619–628, 2010.

Apperly, I.: *Mindreaders: The Cognitive Basis of "Theory of Mind"*. Psychology Press, Hove, East Sussex, 2010.

Barsalou, L.W.: Grounded cognition. *Annual Review of Psychology,* 59, 617-645, 2008.

Bergen, B., Chang, N.: Embodied construction grammar in simulation-based language understanding. In Ostman, J. O. and Fried, M., eds.: *Construction grammar(s): Cognitive and cross-language dimensions*, Johns Benjamins, Amsterdam, 147–190, 2003.

Bräuer, J., et al.: Making inferences about the location of hidden food: Social dog, causal ape. *Journal of Comparative Psychology*, 120(1), 38–47, 2006.

Brooks, R. A.: Intelligence Without Representation. *Artificial Intelligence,* 47(1-3): 139–160, 1991.

Čertický, M.: Action Learning with Reactive Answer Set Programming: Preliminary Report. In Bodendorf, F., ed.: *Proceedings of The Eighth International Conference on Autonomic and Autonomous Systems (ICAS 2012),* Curran Associates, Inc., 2012.

Chemero, A.: *Radical Embodied Cognitive Science*. MIT Press, Cambridge MA, 2009.

Cohen, J.: The earth is round (p < .05). *American Psychologist*, 49, 997-1003, 1994.

Cohen, P. R., Oates, T., Atkin, M. S., Beal, C. R.: Building a Baby. In *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society,* Lawrence Erlbaum, Mahwah NJ, 518-522, 1996.

Daugman, J. G.: Brain metaphor and brain theory. In Schwartz, E. L., ed.: *Computational neuroscience*, MIT Press, Cambridge, MA, 9-18, 1993.

Euclid: *The Elements: Books I-XIII, Complete and Unabridged*. Translated by Sir Thomas Heath, Barnes & Noble, 2006.

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

Gärdenfors, P.: Cued and detached representations in animal cognition. *Behavioural* Processes 36, 263–273, 1996.

Gärdenfors, P.: *Conceptual Spaces: The Geometry of Thought*. MIT Press, Cambridge MA, 2004a.

Gärdenfors, P.: Cooperation and the Evolution of Symbolic Communication. In Oller, D.K. and Griebel, U., eds.: *Evolution of Communication Systems: A Comparative Approach*, MIT Press, Cambridge MA, 237–256, 2004b.

Gardner, H.: *The Mind's New Science: A history of the cognitive revolution*. Basic Books, New York, 1985.

Harnad, S.: The Symbol Grounding Problem. *Physica D,* 42, 335–346, 1990.

Haykin, S. O.: *Neural Networks and Learning Machines*. (3rd edition) Prentice Hall, Upper Saddle River NJ, 2008.

Johnson, M.: *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and* Reason. University of Chicago, Chicago IL, 1987.

Kelemen, J.: The agent paradigm. *Computing and Informatics,* 22, 513–519, 2003.

Kováč, L.: Fundamental Principles of Cognitive Biology. *Evolution and Cognition,* 6, 51-69, 2000a.

Kováč, L.: Ľudstvo v ohrození: potreba syntézy prírodných a kultúrnych vied. In Beňušková, Ľ., Kvasnička, V., Pospíchal, J., eds.: *Hľadanie spoločného jazyka v kognitívnych vedách*. Iris, 83-120, 2000b.

Kováč, L.: Ľudské vedomie je produktom evolučnej eskalácie emocionálneho výberu. In Kelemen, J., ed.: *Kognice a umělý život III.* Slezská univerzita, Opava, pp. 75-93, 2003.

Kripke, S.: Semantical considerations on modal logics. *Acta Philosophica Fennica,* 16, 83–94, 1963.

Lakoff, G.: *Women, Fire, and Dangerous Things*. University of Chicago Press, Chicago IL, 1987.

Langacker, R. W.: *Foundations of cognitive grammar: Theoretical prerequisites*. Vol. 1. Stanford University Press, Stanford CA, 1987.

Langacker, R. W.: *Foundations of cognitive grammar: Descriptive applications*. Vol. 2. Stanford University Press, Stanford CA, 1991.

Maturana, H., Varella, F.: *The Tree of Knowledge: The biological roots of human* understanding. Shambhala, Boston, 1984.

Montague, R.: On the nature of certain philosophical entities. *The Monist,* 53, 159–194, 1960.

Murphy, K. P.: *Machine Learning: A Probabilistic Perspective*. MIT Press, Cambridge MA, 2012.

Peirce, C. S.: *Collected writings*. Hartshorne, C.; Weiss, P. and Burks, A.W., eds., Harvard University Press, Cambridge MA, 1931-58.

Piaget, J., Inhelder, B.: *La psychologie de l'enfant*. (In French; The psychology of the child). PUF, Paris, 1966.

Putnam, H. The nature of mental states. *Readings in the Philosophy of Psychology,* 1, 222-231, 1967.

Rao, A. S., Georgeff, M. P.: Rational Agents within a BDI-Architecture. In Allen, J. et al, eds.: *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning*, Morgan Kaufmann Publishers Inc. San Francisco CA, 473–484, 1991.

Smith, L., Gasser, M.: The Development of Embodied Cognition: Six Lessons from Babies. *Artificial Life*, 11(1–2), 13–30, 2005.

Stewart, J., Gapenne, O., Di Paolo, E. A.: *Enaction*, MIT Press, Cambridge MA, 2011.

Šefránek, J.: Kognícia bez mentálnych procesov. (In Slovak; Cognition without mental processes). In Beňušková, L., Kvasnička, V., Rybár, J., eds.: *Kognitívna veda* (Cognitive science), Kaligram, Bratislava, 200–256, 2002.

Šefránek, J.: Knowledge representation for animal reasoning. In Honkela, T.; Pollä, M.; Paukkeri, M. S., Simula, O., eds.: *Proceedings of AKRR'08, the 2nd International and Interdisciplinary Conference on Adaptive Knowledge Representation and Reasoning*, 17–19th Sept. 2008. Helsinki University of Technology, Helsinki, 2008.

Takáč, M.: Cognitive semantics for dynamic environments. In Hitzler, P., Schärfe, H., Øhrstrøm, P., eds.: *Contributions to ICCS 2006 – 14th International Conference on Conceptual Structures*, 16 –21st July 2006. Aalborg University Press, Aalborg, 2006.

Takáč, M.: Autonomous Construction of Ecologically and Socially Relevant Semantics. *Cognitive Systems Research*, 9(4), 293–311, 2008a.

Takáč, M.: Developing Episodic Semantics. In Honkela, T., Pollä, M., Paukkeri, M. S., Simula, O., eds.: *Proceedings of AKRR'08, the 2nd International and Interdisciplinary Conference on Adaptive Knowledge Representation and Reasoning*, 17–19th Sept. 2008. Helsinki University of Technology, Helsinki, 2008b.

Takáč, M.: Construction of Meanings in Biological and Artificial Agents. In Trajkovski, G., Collins, S. G., eds.: *Agent-Based Societies: Social and Cultural Interactions*, IGI Global, Hershey PA, 139–157, 2009.

Takáč, M., Šefránek, J.: Semantics of distinguishing criteria: from subjective to intersubjective. *Interdisciplinary Description of Complex Systems*, 10(3). 248-269, 2012.

Talmy, L.: *Toward a Cognitive Semantics* (Vol. I, II). MIT Press, Cambridge MA, 2000.

Tarski, A.: Pojecie prawdy w jezykach nauk dedukcyjnych. (In Polish; The concept of truth in the languages of the deductive sciences). *Prace Towarzystwa Naukowego Warszawskiego, Wydzial III Nauk Matematyczno-Fizycznych* 34, 13–172, 1933.

Varela, F. J., Thompson, E. T., Rosch, E.: *The Embodied Mind: Cognitive Science and Human* Experience. MIT Press, Cambridge MA, 1992.

Von Bertalanffy, L.: *General System theory: Foundations, Development, Applications*. George Braziller, New York, 1968.

Von Uexkull, J.: A Stroll through the Worlds of Animals and Men. In C. Schiller, ed.: *Instinctive Behavior*, International Universities Press, New York, 1957.

Ward, L. M.: Dynamical *Cognitive Science*. MIT Press, Cambridge MA, 2001.

Wildgen, W.: *Process, Image, and Meaning. A Realistic Model of the Meanings of Sentences* and *Narrative Texts.* Pragmatics and Beyond, New Series (No. 31). Benjamins, Amsterdam, 29–43, 1994.

Appendix A

A.1 Summary of semantic constructs

Semantic construct	**Formalization**
Current percepts	$P(Env_t)$ – a function projecting the current state of the environment into a cluster of percepts (low-level iconic representations such as retina image etc., essentially numerical vectors/matrices).
Perceptual similarity function	sim_d – a function operating on percepts; returns their degree of similarity (0=none,1=total, or fuzzy values from [0,1]).
Schema	A basic meaning-carrying building block of our semantics. A layered labelled hypergraph; both vertices and edges can have multiple layers of labels. Some vertices are percepts, others are inferred constructs; the edges express relations. The labels are distinguishing criteria or autoreflexive attitudes (carrying type information).
Distinguishing criterion	DC – another basic meaning-carrying building block of our semantics, formalized as a function. Types: DC of change, transformers and detectors.
Signification	Sense-making function σ; it maps current percepts onto a situation schema.
Situation schema	$\sigma(P(Env_t))$ – result of signification; it represents the current situation.
Distinguishing criterion of change	A function defined on pairs $(\sigma(P(Env_{t-1})), \sigma(P(Env_t)))$; it returns 1, if the latter is the result of a change of the former.
Event schema	$\varepsilon(\sigma(P(Env_t)), KB_t)$ – two or more situation schemata linked by (hyper)edges labelled by DC of change.
Similarity functions	More sophisticated functions operating on schemata; they return their degree of similarity. We will denote the set of all the agent's current similarity functions as Sim_t.

Knowledge base	KB_t – a set of the agent's past and current situation and event schemata and similarity functions. $KB_t = \{ \sigma(P(Env_i)) \mid i \le t \} \cup \{ \varepsilon(\sigma(P(Env_i)), KB_i) \mid i \le t \} \cup Sim_t$
Autoreflexive attitude	A type of vertex/edge label carrying information about their semantic type (e.g. object vertex, schema handle, current goal, active, inactive, "related to" edge, etc.).
Beliefs	B_t – a set of schemata of the agent's currently perceived situation and event(s).
Desires	D_t – a set of schemata of the agent's current needs and long-term goals.
Intentions	I_t – a set of schemata of the agent's current goal, a plan to achieve this goal and the state of its execution.
Behaviour	Beh_t – the set of the agent's currently performed actions. The behaviour is observable by other agents.
Current state of "Me"	Me_t – the agent's current knowledge base, percepts, beliefs, desires, intentions and behaviour. $Me_t = (KB_t, P(Env_t), B_t, D_t, I_t, Beh_t)$
Transformer	A type of DC; a function that transforms schemata to other schemata, e.g. refine, zoom in/out, abstract, merge, etc.
Constructor	A type of transformer that creates/modifies detectors (usually by inducing common properties of exemplars).
Updater	A special type of transformer that keeps track of changing schemas and DC.
Detector	A type of DC; a function operating on (fragments of) schemata and returning their degree of membership in an implicitly represented category. Internally, it consists of a template schema and a similarity function. Types: individuals, classes, properties, relations, changes, situation types, event types.
Rule	A schema connecting premises (prerequisites expressed as DC of situation/event types) and justifications (situation/event types preventing the rule application) to consequences (transformers of situation/event schemata).
Inference rule	A rule with transformers realizing internal/mental operations such as change of the focus of attention, zooming in/out, etc.

Action rule	A type of a rule associated with an action (overt behaviour); the rule specifies the prerequisites and consequences of the action execution.
Goal	A desired situation - a situation schema labelled with the autoreflexive attitude “goal”.
Plan	A chain of rules supposedly leading to the fulfilment of a goal.
Routine	A plan successful in the past.
The other agent – observable “It” in a time point t	$It_t = (Beh_t)$ – observed behaviour in a time point t.
Abducible agent “It” (possibly in a time point t)	$(KB'_t, P'(Env_t), B'_t, D'_t, I'_t)$ – all components marked with ‘ are constructed by abduction based on the Me’s own knowledge (theory of mind).
Complete view of another agent	$It'_t = (KB'_t, P'(Env_t), B'_t, D'_t, I'_t, Beh_t)$.
Close neighbourhood relation	A relation between distinguishing criteria or schemata determined by high values of a corresponding similarity function.
Distinguishing criterion with an instrument	A function with an instrument parameter; the parameter specifies a transformer able to compute the value of the function.
Schema with an instrument	A schema generated by a transformer (an algorithm or a conventional, rather mechanical procedure, based on an expert knowledge) and a set of associated schemata.