A Cortex-Inspired Neural-Symbolic Network for Knowledge Representation

Florian Röhrbein, Julian Eggert, Edgar Körner
Honda Research Institute Europe GmbH
Carl-Legien-Strasse 30, 63071 Offenbach am Main, Germany
Florian.Roehrbein@honda-ri.de

Abstract

Semantic systems for the representation of declarative knowledge are usually unconnected to neurobiological mechanisms in the brain. In this paper we report on efforts to bridge this gap by proposing a neural-symbolic network based on processing principles of the cortical column. We show how a locally controlled activation spread on conceptual nodes leads to bottom-up and top-down processing streams which allow for feature inheritance, context effects and the generation of predictions.

1 Introduction

We aim at building a biological motivated relational representation for objects and events, which can be used for basic perceptual tasks like categorization or generation of hypotheses and expectations. For that purpose we developed a graphical network structure which may be classified as a unified system according to the taxonomy introduced by Lallement *et al.* [1995]. Concepts are represented using network nodes in a distributed manner with their constituting parts and properties represented at different nodes. These nodes are uniform units meaning that the same type of unit is used throughout the network irrespective of the represented content. They have simple activation functions which lead to a spread of energy across the network. We allow for several basic link types which represent different semantic relations like "has property" or "is composed of". The network examples shown in Figures 1 and 2 cover all semantic relations used so far and serve as examples throughout the text. Each node can be seen as organized in a columnar way, with each of the standard semantic network links being related to a distinct columnar subsystem located in different cortical layers. Another key feature targeted here is the interplay between bottom-up and top-down processing, see e.g. [Ahissar and Hochstein, 2002]. It is well known, that most everyday activities heavily rely on the interaction of data driven bottom-up processing and category driven top-down information flow. This is especially true for perceptual tasks, for which high-level effects like current context, emotional state or generated expectations have been reported. With our network structure we aim at a joint representation for perception, cognition and action as it has been proposed e.g. by Barsalou [1999]. As a consequence, perceptual states in the brain are assumed not to be transduced into arbitrary amodal symbols, but instead a subset of them is extracted and stored in memory to function as symbols, and association areas then partially reactivate sensory or motor areas in a top-down manner. Consequently, our relational network avoids any strict separation between perceptual, motor and cognitive representations.

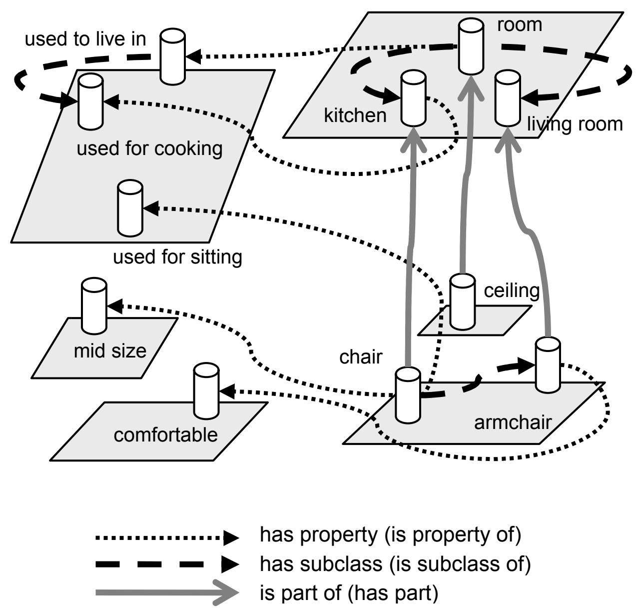

Figure 1: Small network showing all basic facets of knowledge the current system is dealing with including non-trivial dependencies between column-like nodes. All links can be understood as being reciprocal, but for clarity only one direction is shown (in the legend of the arrows the semantic relations corresponding to the opposite direction is indicated).

2 The Cortical Column as Neural-Symbolic Integrator

The cortical column is well known as the basic computational unit in the brain and has been addressed by several researchers with multicellular models to unravel the functional role of the six-layered cortical architecture [Raizada and Grossberg, 2003; Lücke and von der Malsburg, 2004; Kupper *et al*., 2006]. The system described here does not target at a biologically detailed modeling of the single cortical column. Instead we concentrate on a network build out of columnar-like nodes. These cortical columns are typically sectioned into subsystems which comprise different horizontal layers and thereby provide different links for forward, backward and lateral processing. Here we refer to a schema described in [Körner *et al*., 1997] which assumes six distinct systems: Subsystem A1 receives input from lower areas, subsystems A2 and B2 project to higher areas, thus establishing together a bottom-up processing stream (for the difference between A2 and B2 see below), whereas C2 projects to lower areas, which is received in cortical layer I. (since there are no neurons in this layer it is not called a subsystem). The two remaining systems are for lateral processing (B1), which comprises many different cell types and may be subdivided further, and a system for sequential information (C1), which is not used in the work reported here (see also 4.2). One important aspect of cortical columns is that they allow for a smooth transition between signal-type (variant signal) and symbol-type (invariant signal) representations by providing a mechanism to split the ascending signal arriving in A1 into (at least) two components (A2 and B2) which project to different columns on the next hierarchical level. This is illustrated in Figure 2, where variant representations (e.g. certain instances of lips and teeth or even larger combinations of such parts) are passed upwards the hierarchy by each node. If an invariant representation is available (which only makes sense if there are at least two signal representations for the concept in question), this can be used instead. Nodes which consist of symbolically represented parts form super-classes of nodes which contain corresponding parts in a signal representation.

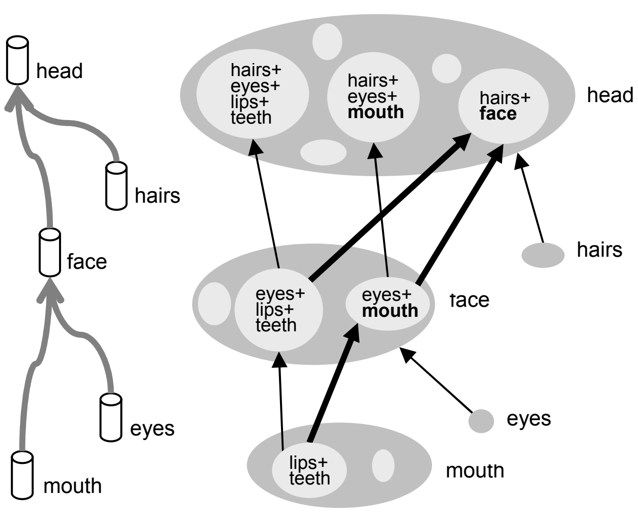

Figure 2: A partonomic hierarchy with five nodes (left) is extended by splitting three nodes ("mouth", "face", "head") into multiple representations (right). Thin arrows are of type "is signal component of", thick ones read "is symbol component of" (see text).

3 Two Interweaved Bodies of Knowledge

Unlike typical semantic networks which allow for a variety of links, we must get along with very few basic links which fit to the constraints posed by the cortical column. The basic dimensions used here are associated with two bodies of knowledge which are of outstanding interest for most cognitive tasks: knowledge about hierarchical relationships and ontological knowledge about properties and subclass relations. Hierarchies are used all over the neocortex as core organization principle to deal with the nested structure of the surrounding world. For example, the visual area TE is assumed to code for object features, which are then combined in perirhinal cortex to form feature-conjunctions [Buckley and Gaffan, 2006]. Along this dimension of knowledge chunks the notions of bottom-up and top-down processing apply. Expressed is knowledge about hierarchical relationships usually in meronymies and holonymies ("is part of"), but also in relations like "is located in" or in the temporal domain ("happens during" etc). Ontological knowledge is expressed in hyponyms and hypernyms ("has superclass", "is instance of" etc.) and is especially useful for feature inheritance. Unlike projects like WordNet or Cyc we put an emphasis on behaviorally relevant concepts rather than on detailed linguistic word meanings. This knowledge is used here on every level of the chunking hierarchy when it is useful in terms of coding efficiency. Interestingly, signal-type representations can be interpreted as subclasses of corresponding symbolic representations.

Other attempts of finding basic dimensions which could span semantic networks come to partially overlapping results, e.g. [Sagerer and Niemann, 1997] propose a three-dimensional hierarchy for scene understanding with the semantic relations "part", "specialization" and "concretization". While information about holonyms and hyponyms are doubtless essential (and are also covered here), the third proposed dimension (concretization) is quite weakly defined as a connection between different levels of abstraction, e.g. between both "locomotion" and "object" and between "object" and "3D-body".

4 Network Constituents

4.1 Nodes

The domain knowledge is represented in a graph structure with nodes representing concepts on various levels and links representing a selected set of relations, which hold between them (see Figure 1). There is only one type of node in the network, which is the representational entity of all concepts of the domain. These concepts are usually associated with different representational levels and consequently a node in our network can be the representation of a sensory measurement, the representation of an instance of a concept or the representation of a category. Since the corresponding biological entity is assumed to be the cortical column, the complexity of the nodes proposed here is beyond those of neural networks and graphical representations like Petri nets or state machines, but clearly below the capabilities of e.g.

multi-agent systems. In this paper, “node” and “column” are used interchangeably. The advantage of focusing on columns instead of neurons here is that we have both, one single unit which is the same for the representation of sensory, cognitive and motor concepts, and the possibility to use various “labeled lines” to deal with semantic relations. Our goal is to get new nodes learned by the system, but so far all nodes and also all relations are manually build or drawn from public domain knowledge bases (see 5.3)

4.2 Links

There are seven different (directed and labeled) types of links in the system reported here. They connect two different nodes with one another modeling inter-columnar connections. Three link types are used to build the chunking hierarchy:

- has component – link
- is signal component of – link
- is symbol component of – link

The “has component” link originates in system C2 and projects to layer I of nodes on a lower level and is thus used for top-down processing. The two “is component of” links stem from different cortical layers (A2 for the signal and B2 for the symbol representation) but terminate both in input layer A1. Together they serve for bottom-up information flow, and just differ in the granularity of transmitted information. For the ontological knowledge four link types are used:

- has property – link
- is property of – link
- has subclass – link
- is subclass of – link

They all form connections within the B1 subsystem (numbered a,b,c,d). Links denoting subclass relationships are assumed to connect columns within one level (e.g. within one cortical area), whereas property links are rather inter-area connections, since they connect conceptual representations with more perceptually based ones. A node with its columnar organization and with all links used is depicted in Figure 3.

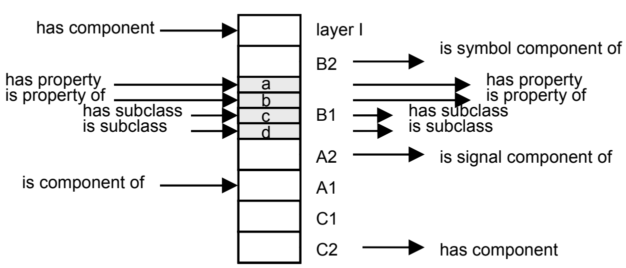

Figure 3: Set of link types available for connecting two nodes.

The links span three basic dimensions (property, subclass and component) with reciprocal links resulting in six possible inputs. Note, that there is one additional output due to the partitioning into signal / symbol-type representation. All network links proposed here differ in two important respects to common semantic network links: First, we only use a very restricted set of basic link types, which are somehow biologically justified, i.e. they can be associated with a specific cell type or a neuronal population within a columnar layer. Second, these links do not vary from node to node, but are common to all nodes. Not all links, of cause, are used by every node, but the point here is that there are no links which are available only for certain nodes. The motivation for this homogenous layout is that the basic structure of the biological column is independent of the cortical site.

4.3 Activation Spread

The activation spread, i.e. all activity that originates from a fixed node results from intra-columnar connectivity patterns. Internally each node has an activity vector with one entry for each subsystem. The production rules for the activities $a_{subsystem}(j)$ of a node j are defined in the following with the abbreviations B1a (has property), B1b (is property of), B1c (has subclass), B1d (is subclass of). Activation from higher cortical areas is passed mainly via subsystem C2 to lower areas, supported by parts of B1. To start with rather simple rules, the activation of the correspondingly connected nodes i is summed, without weighting and thresholds:

$$a_{C2}(j) = \sum_i \{a_{B1a}(i) + a_{B1d}(i) + a_{C2}(i)\}$$

The same activation is propagated to the involved subsystems yielding $a_{B1a}(j)$ and $a_{B1d}(j)$. To model the bottom-up stream, activation has to be propagated via the A2 and B2 systems, depending on the activity of the top-down stream:

$$\begin{aligned} &\textit{if} \quad a_{C2}(j) = 0 : \\ &a_{B2}(j) = 0 \quad \textit{and} \\ &a_{A2}(j) = \sum_i \{a_{A2}(i) + a_{B1b}(i) + a_{B1c}(i) + a_{B2}(i)\} \end{aligned}$$

If there is already activation a_{C2} from top-down at this node, the activation is passed to a_{B2} and not to a_{A2}. The activation again is supported by parts of B1 and also passed to them, whereas the subsystems are complementary to those involved above for a_{C2} (here B1b and B1c are used). Note, that in all cases the activity vector remains unchanged, unless the incoming activity changes (there is no automatic fading away).

5 Results

To illustrate the system behavior, first we apply the proposed mechanisms to the example networks introduced in Figures 1 and 2 and then extend to large common sense databases which were fed into the system.

5.1 Hierarchical Processing

Input to the system can be provided by setting the input activity vector of one or several nodes to values larger than zero. One typical situation in a bottom-up scenario is that some perceptual information is available (e.g. about the size of an object) and the systems' task is to suggest possible objects, which could have given rise to the perceived information. To demonstrate this within the small network, we set the variable a_{A1}(mid size) to some arbitrary value and observe how the activation spreads (Table 1, first row).

single input	t=1	t=2	t=3
↑: mid size	chair	armchair kitchen	living room
↑: comfortable	armchair	living room	
↓: living room	room armchair	used to live in ceiling comfortable chair	used for sitting mid size
↓: kitchen	used for cooking room chair	used to live in ceiling used for sitting mid size	

Table 1: Activation spread originating from different bottom-up or top-down inputs. Nodes which get additionally activated after each time step are shown in the corresponding columns.

After two time steps (i.e. two node transitions) all objects in the knowledgebase which have the property "mid size" are reached: Obviously "chair", which has a direct connection, but also "armchair" gets activated due to the fact that the latter is a subclass of the former. This is interesting insofar as it can be interpreted as inheritance mechanism: "armchair" has also the property of being mid size and gets the same activation as if "mid size" was directly connected to the node. Additionally, concepts get activated which somehow contain the selected objects, here certain rooms (kitchen, living room) and these may provide important contextual information (see Table 2). If in the same way a_{A1}(comfortable) is used as input (Table 1, second row), only "armchair" and "living room" get activated. This is also a desired behavior, since "comfortable" is specific to "armchair" and should not be inherited to "chair".

Next we try how the system operates on top-down input, which could have arisen from some preceding experience (e.g. "I'm in the living room"). The list of activated nodes which results from input on a_{layerI}(living room) and a_{layerI}(kitchen) are shown in the lower part of Table 1. Two interesting behaviors can be seen from there. First, inheritance now works in the opposite direction: An activation of a specific room activates objects and properties which are specific for that room (armchair, comfortable for "living room", but "place for cooking" for kitchen) and also those that are common to all rooms like "place to live in" and "ceiling". Note, we could also set e.g. a_{A1}(kitchen) to some value in order to trigger a bottom-up flow starting from there, but due to the small size of the network this would not lead to any further activations. Finally a combination of forward and backward activation should be considered for three different inputs (rows in Table 2). To ease a comparison all nodes now are grouped according to their source of activation (bottom-up, top-down or both) as a result of the two inputs (first column), which can be interpreted as measurement (BU input) and current hypothesis (TD input).

mixed input	only BU	only TD	BU and TD
↑: comfortable ↓: living room		room used to live in ceiling chair used for sitting mid size	armchair
↑: mid size ↓: living room	kitchen	room used to live in ceiling comfortable used for sitting	chair armchair
↑: mid size ↓: kitchen	arm chair living room	room used to live in ceiling used for cooking used for sitting	chair

Table 2: Interaction of bottom-up (BU) and top-down (TD) processing. Final list of activated nodes after three time steps is shown, input nodes are only listed in the left column.

Nodes which receive both inputs (last column) represent concepts where measurement and hypothesis fit together and lay on a path between the two inputs. On the other hand, nodes with bottom-up activation only represent alternative hypotheses (e.g. kitchen vs. living room) or pieces of evidence which cannot be explained by the current hypothesis. These nodes ("residuals") will play an important role when it comes to learning new representations autonomously. The remaining group of nodes which got only top-down activation also offers an interesting interpretation: They point to objects which can be expected in the scene (e.g. the ceiling in all three cases) or to properties which are probable to be measured (e.g. something of mid size in the first example). These nodes can thus be used for prediction and may be suited for guiding attentional mechanisms.

5.2 Representational Switch

Closely related to the predictive behavior within the hierarchical processing is the switch from signal-type to symbol-type representations. In the results, no distinction was drawn so far between these two. Here we consider the example of Figure 2 and have a closer look on how the activation changes over time (Figure 4): Node A passes the signal representation to C only when no feedback signal is available (as in the beginning of the sequence). As soon as the feedback signal has moved top-down and reached A, the symbol representation is passed instead. Since the only information that is propagated through the network are activation values, the node A has to project to a different node than C to express a different kind of information. Here D takes this role in that it contains a symbolic representation of A. Other nodes which generate this symbolic representation are not depicted here. No symbolic representation is generated in node B (there is no “symbol component of”-relation targeting at any node), so the signal representation is always passed to the next higher level (in the example nodes C, D).

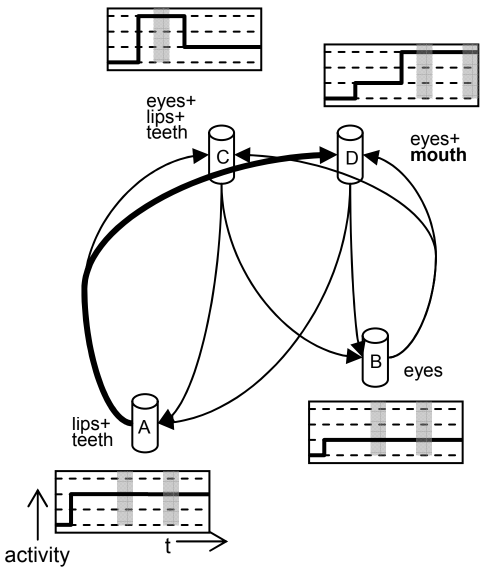

Figure 4: Network configuration with signal- and symbol-type information linked with partonomic relations. Upward links read either “is signal component of” or “is symbol component of” (bold), downward links are of type “has component” (see Figure 2). The activity plots attached to each node show the time course of both the bottom-up spread (black) and feedback signal (grey).

5.3 Experiences with Large Knowledge Bases

As a database for relational knowledge we made use of the corpus collected in the Open Mind Common Sense project (OMCS, http://commonsense.media.mit.edu). The OMCS database seems to be the largest freely-available database of commonsense knowledge and comprises about 1.6 million assertions. So far we only use a fraction of these: First, since OMCS is a World Wide Web based collaborative project with many contributors, for quality reasons we only selected a subset of about 200.000 assertions.

Out of these all assertions were selected which could be matched to one of the basic links described above. For example to insert the assertion “LocationOf – steering wheel – car” from OMCS the nodes “steering wheel” and “car” were generated (if not already existing) and connected via “is symbol component of” and “has component” links. This led to a knowledge base of about 8.400 assertions (see Table 3).

As a next step we compared our system with ConceptNet [Liu and Singh, 2004] which makes use of the same database and also claims to make context-oriented inferences. Therefore we selected the same concepts for the ConceptNet and our system as input nodes and compared the resulting activation. While ConceptNet addresses several additional tasks that go beyond our system, it is clearly outperformed by ours on the basis of the set of nodes which got activated for categorization. To illustrate the differences let us assume concepts *red* and *edible* are activated. In ConceptNet this leads via the “Guess Concept” function to a ranked list of nodes which comprise all concepts that are linked to these properties. Most activation is denoted to the concepts *tomato* and *apple*, since they fulfill both properties. Unfortunately, this works only for properties directly connected with the concepts and not for those connected to some superclass. Tomato and apple for example share the superclass “fruit” which has the property of being nutritious, but an activation of nutritious leads only to an activation of fruit (and “milk”). So there is no feature inheritance in ConceptNet, at least in the concept guessing task. In the OMCS database *tomato* and *apple* are also related to various other concepts, but ConceptNet does not use these relations here, so there is no possibility to trigger concepts through the activation of parts or by assuming a certain location.

In order to check how the proposed network scales up further and since we also encountered some problems with the OMCS database (still partially inconsistent entries, unmapped synonyms etc.), we connected the system with other databases like MILO (http://www.ontologyportal.org), Learner (http://learner.isi.edu), SUMO (http://suo.ieee.org). SUMO is proposed as standard upper ontology by IEEE P1600.1 and comprises about 1100 concepts. It is written in a simplified version of the knowledge interchange format KIF and from this we extracted all assertions designating *instance of* and *subclass of* relations, which amounts to about 1500 relations. Since the concepts of an upper ontology are high-level abstractions, a mid-level ontology is needed to bridge the gap to detailed domain ontologies. Therefore we have chosen MILO, because it shares concepts already defined in SUMO and consequently can easily be interfaced with the upper ontology. MILO is written in the same first order logic language as SUMO, so the same procedure for knowledge extraction was applied. This resulted in nearly 1900 concepts and over 1800 assertions about *instance of* and *subclass of* relationships. Learner finally provides a collection about the everyday world over objects with special emphasis on partonomic relations (overview in Table 3).

extracted relation	amount of assertions	knowledge source
is a	1548 1836 1302	SUMO MILO OMCS
property of	853	OMCS
part of	9973 632	Learner OMCS
location of	5567	OMCS
TOTAL	21711	

Table 3: Number of links used for testing and comparisons in our large-scale network. An assertion always includes two nodes and the appropriate relation. Shown are the numbers of extracted assertions for the different databases.

As result, the proposed schema scales up nicely also with quite large knowledge bases. The inheritance mechanisms observed in the toy example also worked with longer cascades of subclass relationships and chains of property-links. The reason for this is that the activation is passed through within each of the B1 subsystems (indicated by the four straight arrows in Figure 4). We did not observe problems due to cycles or multiple paths (as already contained in Figure 1), except for cases with inconsistent data. Sometimes it was desirable to constrain the top-down spread further in a way that it stopped, if no bottom-up activation is available. This was done easily by modifying the intra-columnar connection rules and there seems also to be neurobiological evidence for a gating role of subsystem A1 on subsystem C2.

6 Related Work and Outlook

In this report we demonstrated the current status of our neural-symbolic network which combines ideas from classical semantic networks with recent findings of the neocortical wiring. By using uniform columnar-like nodes as representational entities we obtained interesting results including feature inheritance, context influence and prediction with a small set of basic semantic relations applied to large common sense databases. Related ideas on how hierarchical representations are used especially for the prediction of sequences have been put forward by Hawkins in [Hawkins and Blakeslee, 2004]. The six-layered cortical organization there plays a central role and is assumed to be also the key for cognition, but so far no modeling results seem to be published. In contrast, van der Velde and Kamps [2006] propose a concrete architecture for dealing with the nested nature of linguistic structures. They stress the role of bottom-up and top-down streams for feature binding, but only provide vague reference to the cortical column. In a multi-agent scenario Bach [2006] proposes “quads” as representational building blocks with links for both chunking and sequence information, which we already associated with corresponding columnar subsystems. Contrary to our locally controlled gating mechanism, they propose “activator neurons” to make the activation spread selective for certain semantic relations. Biologically this seems quite unrealistic, because these nodes need to be connected with every concept node in the network.

Current work concentrates on the inclusion of sequential information and a refinement of the activation schema with weighted links towards a Bayesian framework. In order to increase the expressiveness without losing the generality of the proposed nodes, further semantic relations will be included by representing them as nodes, connected with the basic links described herein.

References

[Ahissar and Hochstein, 2002] Merav Ahissar and Shaul Hochstein. View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, 36:791-804, 2002.

[Bach, 2006] Joscha Bach. MicroPsi: A cognitive modeling toolkit coming of age. In *Proceedings of 7th ICCM, International Conference on Cognitive Modeling*, Trieste, Italy, 2006. Edizioni Goliardiche.

[Barsalou, 1999] Lawrence Barsalou. Perceptual symbol systems. *Behavioral and Brain Sciences*, 22:577–660, 1999.

[Buckley and Gaffan, 2006] Mark J. Buckley and David Gaffan. Perirhinal cortical contributions to object perception. *Trends in Cognitive Sciences*, 10(3):100-107, 2006.

[Hawkins and Blakeslee, 2004] Jeff Hawkins and Sandra Blakeslee. *On Intelligence*. Times Books, New York, 2004.

[Körner *et al.*, 1997] Edgar Körner, Hiroshi Tsujino, and Tomohiko Masutani. A cortical-type modular neural network for hypothetical reasoning. *Neural Networks*, 10:791-814, 1997.

[Kupper *et al.*, 2006] Rüdiger Kupper, Andreas Knoblauch, Marc-Oliver Gewaltig, Ulrike Körner, and Edgar Körner. Simulations of signal flow in a functional model of the cortical column. *Neurocomputing* (accepted).

[Lallement *et al.*, 1995] Yannick Lallement, Mélanie Hilario, and Frédéric Alexandre. Neurosymbolic Integration: Cognitive Grounds and Computational Strategies. In M. DeGlas, Z. Pawlak, (eds). *World Conference on the Fundamentals of A.I.,* Paris, 1995.

[Liu and Singh, 2004] Hugo Liu and Push Singh. ConceptNet: A Practical Commonsense Reasoning Toolkit. *BT Technology Journal*. 22(4):211-226, 2004.

[Lücke and von der Malsburg, 2004] Jörg Lücke and Christoph von der Malsburg. Rapid Processing and Unsupervised Learning in a Model of the Cortical Macrocolumn. *Neural Computation*, 16:501-533, 2004.

[Raizada and Grossberg, 2003] Rajeev D.S. Raizada and Stephen Grossberg. Towards a theory of the laminar architecture of cerebral cortex: Computational clues from the visual system. *Cerebral Cortex*, 13:100-113, 2003.

[Sagerer and Niemann, 1997] Gerhard Sagerer and Heinrich Niemann. Semantic Networks for Understanding Scenes. *Advances in Computer Vision and Machine Intelligence*. Plenum Press, New York and London, 1997.

[Tsunoda *et al.*, 2001] Kazushige Tsunoda, Yukako Yamane, Makoto Nishizaki, and Manabu Tanifuji. Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nature Neuroscience*, 4:832-838, 2001.

[van der Velde and Kamps, 2006] Frank van der Velde and Marc de Kamps. Neural blackboard architectures of combinatorial structures in cognition. *Behavioral and Brain Sciences* (in press).