A Biologically Motivated and Computationally Efficient Natural Language Processor

João Luís Garcia Rosa

Centro de Ciências Exatas, Ambientais e de Tecnologias
Pontifícia Universidade Católica de Campinas - PUC-Campinas
Rodovia D. Pedro I, km. 136 - Caixa Postal 317
13012-970 - Campinas - SP - Brazil
joaoluis@puc-campinas.edu.br

Abstract. Conventional artificial neural network models lack many physiological properties of the neuron. Current learning algorithms are more concerned to computational performance than to biological credibility. Regarding a natural language processing application, the thematic role assignment – semantic relations between words in a sentence –, the purpose of the proposed system is to compare two different connectionist modules for the same application: (1) the usual simple recurrent network using backpropagation learning algorithm with (2) a biologically inspired module, which employs a bi-directional architecture and learning algorithm more adjusted to physiological attributes of the cerebral cortex. Identical sets of sentences are used to train the modules. After training, the achieved output data show that the physiologically plausible module displays higher accuracy for expectable thematic roles than the traditional one.

1 Introduction

Several connectionist natural language processing systems often employ recurrent architectures instead of feedforward networks. These systems with "reentrancy" are expected to be more adequate to deal with the temporal extension of natural language sentences, and, at the same time, they seem to be physiologically more realistic [1]. Other biological features are being taken into account in order to achieve new models that restore the artificial neural systems first concerns. Connectionist models based on neuroscience are about to be considered the next generation of artificial neural networks, inasmuch as nowadays models are far from biology, mainly for mathematical simplicity reasons [2].

In this paper, it is compared two distinct connectionist modules of a system about the thematic role assignment in natural language sentences: a conventional simple recurrent network employing the backpropagation learning algorithm (TRP-*BP*) with a bi-directional architecture using a biologically plausible learning algorithm, adapted from the Generalized Recirculation algorithm [3] (TRP-*GR*). Through the same set of test sentences, it is shown that, for the same training set, the neurophysiological module reflects better the thematic relationships taught to the system.

R. Monroy et al. (Eds.): MICAI 2004, LNAI 2972, pp. 390-399, 2004.

2 Thematic Roles

Linguistic theory [4] refers to the roles words usually have in relation to the predicate (often the verb) as thematic roles, so that the verb *break*, for instance, in one possible reading of sentence (1), assigns the thematic roles AGENT, PATIENT, and INSTRUMENT, because the subject *man* is supposed to be deliberately responsible for the action of breaking (the "agent"), the object *vase* is the "patient" affected by the action, and the complement *stone* is the "instrument" used for such action.

The man broke the vase with the stone . (1)

But the thematic structure can change for some verbs. So, in sentence (2), there is a different thematic grid ([CAUSE, PATIENT]) assigned by the same verb *break*, since the subject *ball* causes the breaking, but in an involuntary way.

The ball broke the vase . (2)

Verbs presenting two or more thematic grids depending on the sentence they take place, like the verb *break*, are named here as *thematically ambiguous* verbs. In a componential perspective, it is possible to have a representation for verbs independently of the sentence in which they occur. Considering sentences (1) and (2) again, it seems that the nouns employed as subjects make the distinction between AGENT and CAUSE. In other words, thematic roles must be elements with semantic content [5].

One of the reasons that the thematic assignment is chosen for a connectionist natural language processing application is because of its componential feature. Details can be found in [6].

2.1 Word Representation

In the system presented, word representation is adapted from the classical distributed semantic microfeature representation [7]. Twenty three-valued logic semantic microfeature units account for each verb and noun. Table 1 and table 2 display the semantic features for verbs and nouns, respectively. See also the microfeatures for two different readings of the thematically ambiguous verb *break* on table 3 [8].

It is important to notice here that the microfeatures for verbs are chosen in order to contemplate the semantic issues considered relevant in a thematic frame. The microfeatures outside this context are not meaningful. They only make sense in a system where the specification of semantic relationships between the words in a sentence plays a leading role [6].

Table 1. The ten semantic microfeature dimensions for verbs. For thematically unambiguous verbs, only one feature in each dimension is *on*

“positive” feature	**“negative” feature**
control of action	*no control of action*
direct process triggering	*indirect process triggering*
direction to source	*direction to goal*
impacting process	*no impacting process*
change of state	*no change of state*
psychological state	*no psychological state*
objective	*no objective*
effective action	*no effective action*
high intensity of action	*low intensity of action*
interest on process	*no interest on process*

Table 2. The seven semantic microfeature dimensions for nouns, separated in rows. Only one value in each dimension is *on* for each unambiguous noun (adapted from [7])

<table>
<tr><td colspan="3">human</td><td colspan="3">non-human</td></tr>
<tr><td colspan="3">soft</td><td colspan="3">hard</td></tr>
<tr><td colspan="2">small</td><td colspan="2">medium</td><td colspan="2">large</td></tr>
<tr><td colspan="2">1-D/compact</td><td colspan="2">2-D</td><td colspan="2">3-D</td></tr>
<tr><td colspan="3">pointed</td><td colspan="3">rounded</td></tr>
<tr><td colspan="3">fragile/breakable</td><td colspan="3">unbreakable</td></tr>
<tr><td>value</td><td>furniture</td><td>food</td><td>toy</td><td>tool/utensil</td><td>animate</td></tr>
</table>

Table 3. The semantic microfeatures for the thematically ambiguous verb *break*, with the default reading and two alternative readings (*break1* and *break2*). The “?” sign represents ambiguity [8]

microfeature	***break***	***break1***	***break2***
control of action	?	yes	no
process triggering	?	direct	indirect
direction	goal	goal	goal
impacting process	yes	yes	yes
change of state	yes	yes	yes
psychological state	no	no	no
objective	?	yes	no
effective action	yes	yes	yes
intensity of action	high	high	high
interest on process	?	yes	no

3 TRP-*BP*

The Thematic Role Processor (TRP) is a connectionist system designed to process the thematic roles of natural language sentences, based on its symbolic-connectionist hybrid version [9]. For each input sentence, TRP gives as output, its thematic grid. TRP is deployed in two modules with completely different approaches: *BP* and *GR*. TRP-*BP* learns through backpropagation algorithm and employs an architecture representing a four-layer simple recurrent neural network with forty input units (A), fifteen hidden units (B), fifteen context units (D), and ten output units (C), one for each of the ten thematic roles: AGENT, PATIENT, EXPERIENCER, THEME, SOURCE, GOAL, BENEFICIARY, CAUSE, INSTRUMENT, and VALUE (figure 1).

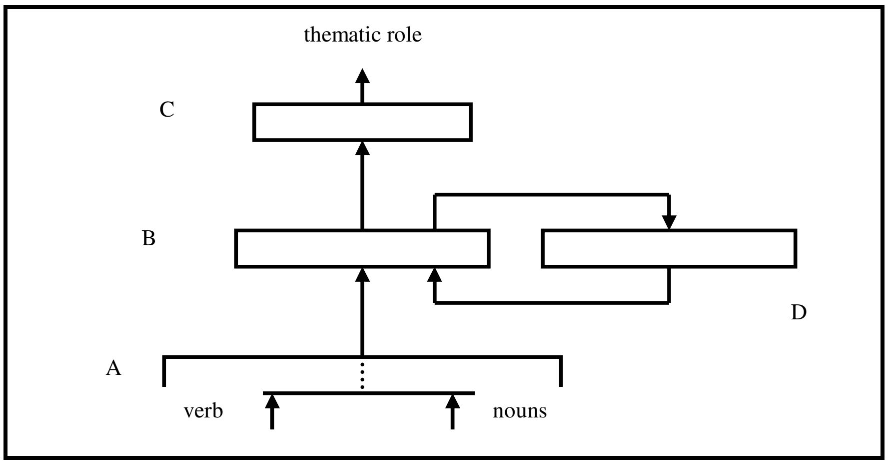

Fig. 1. The four-layer simple recurrent connectionist architecture of TRP-*BP*. To the input layer *A* the words, represented by their distributed semantic microfeatures, are entered sequentially at their specific slots according to their syntactic category: *verb* or *nouns* (subject, object, or complement). At the output layer *C*, a *thematic role* (AGENT, PATIENT, EXPERIENCER, THEME, SOURCE, GOAL, BENEFICIARY, CAUSE, INSTRUMENT, or VALUE), is displayed as soon as a word is entered in layer *A*. The context layer *D* represents the memory of the network, to which the hidden layer *B* is copied after each training step [10]

The input layer is divided in a twenty-unit slot for the verb and another twenty-unit slot for nouns. Words are presented in terms of their semantic microfeatures, one at a time, at their specific slots, until the whole sentence is completely entered. This way, besides semantics, included as part of the distributed representation employed, all kinds of natural languages with regard to word order (verb-subject-object -VSO -, as well as SVO) could be considered, since a predicate-arguments relation is established. At output layer C, thematic roles are highlighted as soon as they are assigned. For instance, when the subject of a sentence is presented, no thematic role shows up, because it is unknown which will be the main verb, the predicate that assigns such role. When the verb appears, immediately the network displays the thematic role assigned to the subject presented previously. For the other words, the correspondent thematic roles are displayed at the output, one at a time, for every input word.

3.1 The Biological Implausibility of Backpropagation

The backpropagation algorithm is largely employed nowadays as the most computationally efficient connectionist supervised learning algorithm. But backpropagation is argued to be biologically implausible [11]. The reason is that it is based on the error back propagation, that is, while the stimulus propagates forwardly, the error (difference between the actual and the desired outputs) propagates backwardly. It seems that in the cerebral cortex, the stimulus that is generated when a neuron fires, crosses the axon towards its end in order to make a synapse onto another neuron input (called dendrite). Supposing that backpropagation occurs in the brain, the error must have to propagate back from the dendrite of the post-synaptic neuron to the axon and then to the dendrite of the pre-synaptic neuron. It sounds unrealistic and improbable. Researchers believe that the synaptic “weights” have to be modified in order to make learning possible, but certainly not in this way. It is expected that the weight change uses only local information available in the synapse where it occurs. That is the reason why backpropagation seems to be so biologically implausible [8].

4 TRP-*GR*

The module TRP-*GR* (*GR* for *Generalized Recirculation*) consists of a bi-directional connectionist architecture, with three layers (A units in input layer, B units in hidden layer, and C units in output layer) and lateral inhibition occurring at the output level (figure 2). The input and the output operate in the same way as TRP-*BP*.

4.1 The Learning Procedure

The learning procedure of TRP-*GR*, also employed in [12] and [8], is inspired by the Recirculation [13] and GeneRec algorithms [3], and uses the two phases notion (*minus* and *plus* phases). Firstly, the inputs x_i are presented to the input layer. In the minus phase, there is a propagation of these stimuli to the output through the hidden layer (bottom-up propagation). There is also a propagation of the previous actual output o_k (t-1) back to the hidden layer (top-down propagation). Then, the hidden minus activation h_j^- is generated (sum of the bottom-up and top-down propagations – through the sigmoid activation function, represented by σ in equation 3). Finally, the current real output $o_k(t)$ is generated through the propagation of the hidden minus activation to the output layer (equation 4). The indexes i, j, and k refer to input, hidden, and output units, respectively.

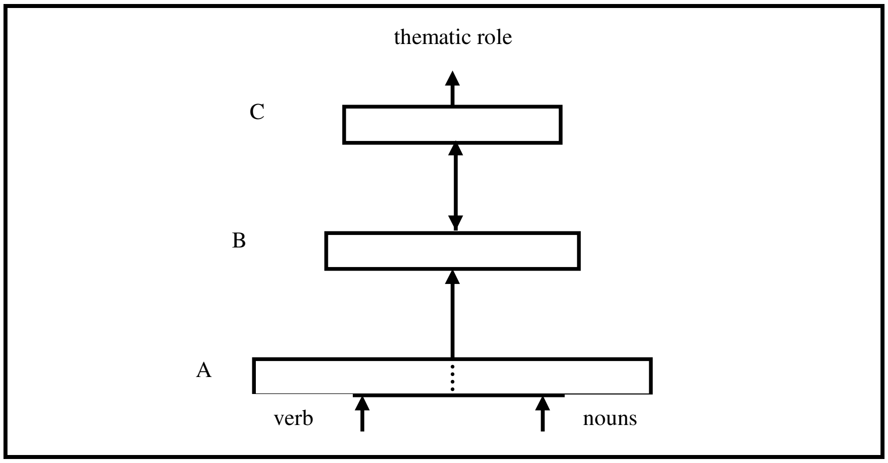

Fig. 2. The three-layer bi-directional connectionist architecture of TRP-*GR*. To the input layer *A* the words, represented by their distributed semantic microfeatures, are entered sequentially at their specific slots according to their syntactic category: *verb* or *nouns* (subject, object, or complement). At the output layer *C*, a *thematic role* (AGENT, PATIENT, EXPERIENCER, THEME, SOURCE, GOAL, BENEFICIARY, CAUSE, INSTRUMENT, or VALUE), is displayed as soon as a word is entered in layer *A*. This architecture is similar to the TRP-*BP* (figure 1), except that there is no layer D and the connections between layers B and C are bi-directional

$$h_j^- = \sigma(\sum_{i=0}^{A} w_{ij}.x_i + \sum_{k=1}^{C} w_{jk}.o_k(t-1)) \tag{3}$$

$$o_k(t) = \sigma(\sum_{j=1}^{B} w_{jk}.h_j^-) \tag{4}$$

In the plus phase, there is a propagation from the input x_i to the hidden layer (bottom-up). After this, there is the propagation of the desired output y_k to the hidden layer (top-down). Then the hidden plus activation h_j^+ is generated, summing these two propagations (equation 5).

$$h_j^+ = \sigma(\sum_{i=0}^{A} w_{ij}.x_i + \sum_{k=1}^{C} w_{jk}.y_k) \tag{5}$$

In order to make learning possible, the synaptic weights w are updated, based on x_i, h_j^-, h_j^+, $o_k(t)$, and y_k, in the way shown in equations 6 and 7. Notice the presence of the learning rate (η), considered an important variable during the experiments [14].

$$\Delta w_{jk} = \eta.(y_k - o_k(t)).h_j^- \tag{6}$$

$$\Delta w_{ij} = \eta.(h_j^+ - h_j^-).x_i \tag{7}$$

5 Comparing TRP-*GR* with TRP-*BP*

Nowadays, neural network models are considered biologically impoverished, although computationally efficient. It has been proved that neurophysiologically based systems can be as computationally efficient as current connectionist systems, or even better [15]. This paper demonstrates that a connectionist system, with architecture and learning procedures based on neuroscience features, therefore biologically plausible, are also computationally efficient, more efficient than conventional systems, at least regarding a particular natural language processing application.

5.1 Training

A sentence generator supplies 364 different training sentences, presented one word at a time, according to semantic and syntactic constraints, employing a lexicon consisting of 30 nouns and 13 verbs, including thematically ambiguous verbs. It is important to emphasize here that the same training sentences are generated for both modules. After about 100,000 training cycles, which corresponds to an average output error[1] of 10^{-3}, the system is able to display, with a high degree of certainty, the thematic grid for an input sentence.

5.2 Set of Test Sentences

In order to compare TRP-*GR* with TRP-*BP*, 16 test sentences, different from training sentences, were generated by the sentence generator for both modules. Four of them are shown in figures 3 to 6, with their outputs representing thematic roles. These sentences reveal the better computational performance of the biologically plausible module TRP-*GR*, at least regarding the sentences belonging to the test set (*GR* is 11.61% more efficient than *BP*). Alternatively to the sentences generated automatically, the user can enter by hand the sentence to be tested.

On figure 3, one can see outputs of the system for the sentence *the boy fears the man*. The current word entered is in bold, while the previous words already entered are listed sequentially in parentheses. The first output line is for *BP* and the second for *GR*. The closer the output is to 1.0 the more precise is the thematic role prediction. Notice that in the *GR* module, the outputs are more accurate, at least regarding the expected thematic role for each word. Recall that the first word entered (the subject

[1] The average output error is the difference between "actual" and "desired" outputs, and it is obtained from the *average squared error energy formula* [14] for each set of different sentences presented to the network.

boy) has no thematic role displayed – that is the reason it does not appear in the figure – at least until the verb shows up. To the subject (*boy*) is assigned the thematic role EXPERIENCER, because *fear* asks for an experiencer subject it is a psychological verb.

Sentence: *the boy fears the man*

Word presented to the system: (boy) **fear**

	agent	patie	**exper**	theme	sourc	goal	benef	cause	instr	value
BP	0.001	0.004	**0.966**	0.000	0.007	0.015	0.020	0.016	0.000	0.000
GR	0.000	0.000	**0.998**	0.031	0.000	0.000	0.000	0.000	0.000	0.000

Word presented to the system: (boy-fear) **man**

	agent	patie	exper	**theme**	sourc	goal	benef	cause	instr	value
BP	0.000	0.024	0.054	**0.101**	0.044	0.049	0.034	0.001	0.003	0.001
GR	0.000	0.021	0.058	**0.168**	0.136	0.047	0.015	0.000	0.000	0.000

Fig. 3. Outputs of the system for the sentence *the boy fears the man*. Both modules arrived at the expected thematic grid [EXPERIENCER, THEME], but module *GR* values are closer to 1.0

Figure 4 displays outputs for the sentence *the girl bought a ball for a hundred dollars*. Notice again that in the *GR* module, the outputs are more precise.

Sentence: *the girl bought a ball for a hundred dollars*

Word presented to the system: (girl) **buy**

	agent	patie	exper	theme	sourc	goal	benef	cause	instr	value
BP	**0.984**	0.000	0.015	0.049	0.004	0.005	0.003	0.004	0.016	0.000
GR	**1.000**	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000

Word presented to the system: (girl-buy) **ball**

	agent	patie	exper	**theme**	sourc	goal	benef	cause	instr	value
BP	0.011	0.036	0.010	**0.412**	0.014	0.010	0.019	0.004	0.077	0.016
GR	0.000	0.038	0.000	**0.877**	0.000	0.000	0.000	0.000	0.055	0.001

Word presented to the system: (girl-buy-ball) **hundred**

	agent	patie	exper	theme	sourc	goal	benef	cause	instr	**value**
BP	0.001	0.029	0.002	0.146	0.006	0.005	0.006	0.002	0.080	**0.928**
GR	0.000	0.000	0.000	0.046	0.000	0.000	0.000	0.000	0.002	**0.998**

Fig. 4. Outputs of the system for the sentence *the girl bought a ball for a hundred dollars*. Both modules arrived at the expected thematic grid [AGENT, THEME, VALUE], although in module *GR* the displayed values are closer to 1.0

Figure 5 shows the outputs of the system for the sentence *the man hit the doll*, with the thematically ambiguous verb *hit*, entered as its default reading, since it is unknown for the system which *hit* is intended. Notice that, in this case, the thematic role assigned to the subject must be AGENT instead of CAUSE, because the noun *man* has features associated with being capable of controlling the action, for instance, *human* and *animate* (see table 2). As one can see, only the *GR* module gave a suitable prediction for the subject thematic role.

Sentence: *the man hit the doll*

Word presented to the system: (man) **hit**

	agent	patie	exper	theme	sourc	goal	benef	cause	instr	value
BP	0.531	0.000	0.001	0.012	0.004	0.004	0.003	**0.768**	0.003	0.000
GR	**0.542**	0.000	0.000	0.004	0.000	0.000	0.000	0.482	0.000	0.000

Word presented to the system: (man-hit) **doll**

	agent	patie	exper	**theme**	sourc	goal	benef	cause	instr	value
BP	0.020	0.082	0.001	**1.000**	0.001	0.000	0.004	0.001	0.000	0.000
GR	0.000	0.079	0.000	**0.881**	0.000	0.000	0.000	0.000	0.000	0.000

Fig. 5. Outputs of the system for the sentence *the man hit the doll*. Module *GR* arrived at the expected thematic grid [AGENT, THEME], while module *BP* arrived at [CAUSE, THEME]

At last, figure 6 shows outputs for the sentence *the hammer broke the vase*, which employs another thematically ambiguous verb (*break*). To the subject of this sentence (*hammer*) is assigned the thematic role CAUSE instead of AGENT, because *hammer* causes the breaking, but it is not responsible for this action. Its semantic features include *non-human* and *tool/utensil*, incompatible to *control of action*, feature expected to be associated to the verb that assigns the thematic role AGENT. For the object, module *BP* worked better, assigning PATIENT to *vase*, instead of THEME.

Sentence: *the hammer broke the vase*

Word presented to the system: (hammer) **break**

	agent	patie	exper	theme	sourc	goal	benef	**cause**	instr	value
BP	0.486	0.000	0.001	0.006	0.004	0.005	0.003	**0.764**	0.004	0.000
GR	0.093	0.000	0.000	0.005	0.000	0.000	0.000	**0.912**	0.000	0.000

Word presented to the system: (hammer-break) **vase**

	agent	**patie**	exper	theme	sourc	goal	benef	cause	instr	value
BP	0.001	**0.325**	0.011	0.313	0.004	0.004	0.017	0.001	0.077	0.003
GR	0.000	0.333	0.001	**0.536**	0.000	0.000	0.000	0.000	0.038	0.000

Fig. 6. Outputs for the sentence *the hammer broke the vase*. Both modules arrived at the expected thematic role CAUSE for the subject. Concerning the object, while in *BP* there is a slight preference for PATIENT, in *GR* there is an unexpected assignment of THEME

6 Conclusion

The modules TRP-*BP* and TRP-*GR*, of the proposed system, are connectionist approaches to natural language processing, regarding the thematic role relationships between words of a sentence. The aim of this paper is to show that a biologically motivated connectionist system, with a bi-directional architecture and learning algorithm that uses only local information to update its synaptic weights, is able not only to take care of this natural language processing problem, but also to be more computationally efficient than the conventional backpropagation learning procedure through a simple recurrent connectionist architecture. This is confirmed by the outcomes for a same set of test sentences presented to both modules of the system.

References

1. Edelman, G. M., Tononi, G.: A Universe of Consciousness - How Matter Becomes Imagination. Basic Books (2000)
2. Rosa, J. L. G.: An Artificial Neural Network Model Based on Neuroscience: Looking Closely at the Brain. In: Kůrková, V., Steele, N. C., Neruda, R., Kárný, M. (eds.): Artificial Neural Nets and Genetic Algorithms – Proc. of the Intl. Conf. in Prague, Czech Republic – ICANNGA-2001. April 22-25, Springer-Verlag (2001) 138-141
3. O'Reilly, R. C.: Biologically Plausible Error-driven Learning using Local Activation Differences: The Generalized Recirculation Algorithm. Neural Comp.8:5(1996)895-938
4. Haegeman, L.: Introduction to Government and Binding Theory. Basil Blackwell Inc., Cambridge, MA (1991)
5. Dowty, D.: On the Semantic Content of the Notion of 'Thematic Role'. In Chierchia, G. *et al.* (eds.): Properties, Types and Meaning. Dordrecht, Kluwer (1989)
6. Rosa, J. L. G., Françozo, E.: Linguistic Relations Encoding in a Symbolic-Connectionist Hybrid Natural Language Processor. In Monard, M. C., Sichman, J. S. (eds.): Advances in Artificial Intelligence – Proc. of the IBERAMIA-SBIA 2000, Atibaia, São Paulo, Brazil, November 19-22. Lecture Notes in Artificial Intelligence, Vol. 1952. Springer-Verlag, Berlin (2000) 259-268
7. McClelland, J. L., Kawamoto, A. H.: Mechanisms of Sentence Processing: Assigning Roles to Constituents of Sentences. In McClelland, J. L., Rumelhart, D. E. (eds.): Parallel Distributed Processing, Vol. 2. A Bradford Book, The MIT Press (1986)
8. Rosa, J. L. G.: A Biologically Inspired Connectionist System for Natural Language Processing. In Proc. of the 2002 VII Brazilian Symposium on Neural Networks (SBRN 2002). 11-14 November. Recife, Brazil. IEEE Computer Society Press (2002) 243-248
9. Rosa, J. L. G., Françozo, E.: Hybrid Thematic Role Processor: Symbolic Linguistic Relations Revised by Connectionist Learning. In Proc. of IJCAI'99-16th. Intl. Joint Conf. on Artificial Intelligence, Vol. 2, Stockholm, Sweden, 31 July-6 August(1999) 852-857
10. Elman, J. L.: Finding Structure in Time. Cognitive Science 14 (1990) 179-211
11. Crick, F. H.: The Recent Excitement about Neural Networks. Nature 337(1989)129-132
12. Rosa, J. L. G.: A Biologically Motivated Connectionist System for Predicting the Next Word in Natural Language Sentences. In Proc. of the 2002 IEEE Intl. Conf. on Systems, Man, and Cybernetics. 06-09 October. Hammamet, Tunisia (2002)
13. Hinton, G. E., McClelland, J. L.: Learning Representations by Recirculation. In Anderson, D. Z. (ed.): Neural Information Processing Systems, 1987. New York: American Institute of Physics (1988) 358-366
14. Haykin, S.: Neural Networks - A Comprehensive Foundation. 2nd edn. Prentice Hall, Upper Saddle River, New Jersey (1999)
15. O'Reilly, R. C., Munakata, Y.: Computational Explorations in Cognitive Neuroscience - Understanding the Mind by Simulating the Brain. A Bradford Book, The MIT Press, Cambridge, MA (2000)