Proceedings of the 2003 IEEE International Conference on Systems, Man, and Cybernetics – IEEE-SMC'03. 05-08 October 2003. Washington, District of Columbia, United States of America, pp. 2845-2850.

A Biologically Plausible and Computationally Efficient Architecture and Algorithm for a Connectionist Natural Language Processor*

João Luís Garcia Rosa
Centro de Ciências Exatas, Ambientais e de Tecnologias - PUC-Campinas
Mestrado em Informática – UniSantos
Brasil
joaoluis@puc-campinas.edu.br

Abstract – *Nowadays, most connectionist models are more oriented to computational efficiency instead of neurophysiological inspiration. Classical learning algorithms, like the largely employed backpropagation, is argued to be biologically implausible. This paper aims to prove that a biologically inspired connectionist architecture and algorithm is not only capable of dealing with a high level cognitive task, like a natural language processing application, but also be more computationally efficient. It is presented a comparison between a standard simple recurrent network using backpropagation with a physiologically inspired system. Symbolic data, extracted from connectionist architectures, show that the physiologically plausible model displays more expectable semantic features about thematic relations between words than the conventional one.*

Keywords: neural networks, natural language processing, machine learning.

1 Introduction

Since a natural language processing task has a temporal extension, it is expected that it would be more adequately treated with a recurrent, instead of a feedforward connectionist network. In addition, these systems with "reentrancy" seem to be more physiologically realistic [4]. However, the backpropagation algorithm, which is largely employed nowadays as the most computationally efficient connectionist supervised learning algorithm, is argued to be biologically implausible [1]. Models based on neuroscience are about to be considered the next generation of artificial neural networks, since current models are biologically impoverished, mainly for mathematical simplicity reasons [12].

In this paper, it is compared two distinct connectionist models: a conventional simple recurrent network with backpropagation learning algorithm and a bi-directional architecture with a biologically plausible learning algorithm, adapted from the Generalized Recirculation algorithm [10], both concerning the same application: the thematic role assignment in natural language sentences. Thematic roles are semantic relationships between words in a sentence, like AGENT, THEME, INSTRUMENT, etc. [6]. Through the symbolic data extracted from the connectionist architecture (it had been already proved that the data set and the network, from which it is extracted, are very equivalent [2]), it is shown that, for the same number of training cycles and the same training set, the biologically plausible version reflects better the thematic relationships taught to the system.

2 Symbolic-connectionist hybrid systems

The critics of neural networks accentuate that they lack transparency, that is, one does not know how they work, how they develop internal representations. And that training often takes too long. A solution to such stricture is the hybrid symbolic-connectionist approach. In this method one can combine symbolic approach benefits, like expressive power of the general logical implications, ease of knowledge representation, and understanding through logical inference, with connectionism advantages, like learning, generalization, and fault tolerance [13].

In a symbolic-connectionist hybrid approach, symbolic knowledge is inserted in a connectionist architecture as connection weights. The network is submitted to a training period, like conventional connectionist systems. After training, the symbolic theory, which gave initial knowledge to the network, is revised by the connectionist learning. This way, it is possible to overcome the drawbacks presented previously: since the system has initial knowledge (weights are not random anymore; they are assigned symbolic data) it takes less time to learn; and because of the nodes now naming concepts, the weights linked to them does make sense. The

Proceedings of the 2003 IEEE International Conference on Systems, Man, and Cybernetics – IEEE-SMC'03. 05-08 October 2003. Washington, District of Columbia, United States of America, pp. 2845-2850.

symbolic knowledge generated by the net can be extracted. The symbolic data set and the network, from which it is extracted, are supposed to be very equivalent [2].

3 Thematic roles

Thematic roles are semantic functions assigned to words in a sentence, by a predicate, usually the verb [6]. So, the verb *frighten*, for instance, in one possible reading of sentence (1), assigns the thematic roles AGENT and EXPERIENCER, because *man* is supposed to be deliberately responsible for the action of frightening (the "agent"), and *girl* is the "experiencer" affected by the action.

The man frightened the girl (1)

But there are verbs that can assign more than one thematic grid, depending on the sentence they occur. For instance, in sentence (2), there is a different thematic grid ([CAUSE, EXPERIENCER]) assigned by the same verb *frighten*, since *ball* causes the frightening, but in an involuntary way.

The ball frightened the girl (2)

Verbs with more than one thematic grid are called *thematically ambiguous* verbs. 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 [3].

3.1 Word representation

The representation chosen for words in the presented system are based on the classical semantic microfeature distributed representation [9]. Ten dimensions with two units each account for each noun and verb. Table 1 shows the semantic features for verbs. Table 2 displays the microfeatures for verb *frighten* [14].

Table 1. The semantic microfeature dimensions for verbs according to a thematic frame

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 semantic microfeatures for verb *frighten*, with the default reading and two alternative readings (*frighten1* and *frighten2*). The "?" sign represents thematic ambiguity [14]

microfeature	*frighten*	*frighten1*	*frighten2*
control of action	?	yes	no
process triggering	?	direct	indirect
direction	goal	goal	goal
impacting process	yes	yes	yes
change of state	no	no	no
psychological state	yes	yes	yes
objective	?	yes	no
effective action	no	no	no
intensity of action	low	low	low
interest on process	?	yes	no

Since the aim of the presented system is to deal with the thematic relationships between words in a sentence, the microfeatures chosen for the verbs attempt to contemplate the semantic issues considered relevant in a thematic frame. The microfeatures outside this context are not meaningful [13].

4 The conventional HTRP-*BP*

The system HTRP – Hybrid Thematic Role Processor – is a symbolic-connectionist hybrid system designed to process the thematic roles of natural language sentences [15]. Symbolic rules about thematic roles are inserted as initial knowledge of the system. After a connectionist training, a revised symbolic theory is extracted. For each input sentence, HTRP gives as output, its thematic grid. Now, it is proposed two new versions with completely different approaches for HTRP: *BP* and *GR*. HTRP-*BP* learns through backpropagation algorithm and employs a simple recurrent architecture representing a four-layer neural network with eighty input units, twenty hidden units, twenty context units, and ten output units, one for each of the ten thematic roles: AGENT, PATIENT, EXPERIENCER, THEME, SOURCE, GOAL, BENEFICIARY, CAUSE, INSTRUMENT, and VALUE. The fourth layer – the context layer – with the same seize of the hidden layer, is designed to store the previous internal representations. To the context layer is copied, after each training step, the hidden layer. And this layer is responsible for a second input stimulus to the network (figure 1).

The words, represented by their semantic microfeatures, are presented at input layer, sequentially, one at a time, at their specific slots, depending on their syntactic categories, until the whole sentence is completely entered. This way, besides semantics, included as part of the distributed representation employed, syntactic constraints are also considered. At output layer, thematic

Proceedings of the 2003 IEEE International Conference on Systems, Man, and Cybernetics – IEEE-SMC'03. 05-08 October 2003. Washington, District of Columbia, United States of America, pp. 2845-2850.

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 roles. 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.

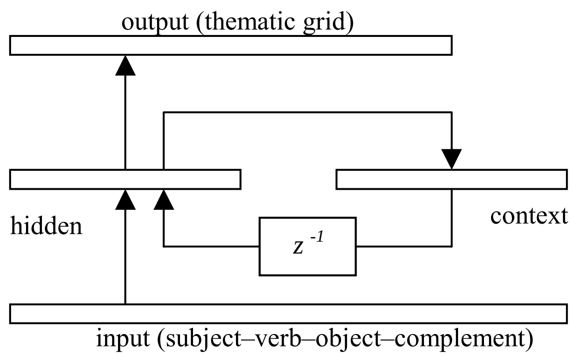


Figure 1. HTRP-*BP* architecture with four layers: input, hidden, output, and an extra (context) layer in a partially recurrent Elman network [5]. The network input receives one word at a time in the localized slot concerning the specific syntactic category

4.1 Backpropagation algorithm is biologically implausible

The backpropagation learning procedure, largely used nowadays as the most computationally efficient connectionist supervised algorithm, is argued to be neurophysiologically implausible [1]. This is mainly because of the error back propagation – the way the algorithm propagates backwardly the difference between real and desired outputs, in a manner against the belief the biological neural tissue does. While the stimulus propagates forwardly, there is a back propagation of the error signal, through the network layers. It sounds biologically implausible because the synaptic weight change in the cerebral cortex seems to happen by employing only available information local in the synapse. 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). Suppose 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 [14].

5 The biologically inspired HTRP-*GR*

The system HTRP-*GR* (*GR* stands 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 layer (figure 2). The input words and the output thematic roles operate in the same way as HTRP-*BP*.

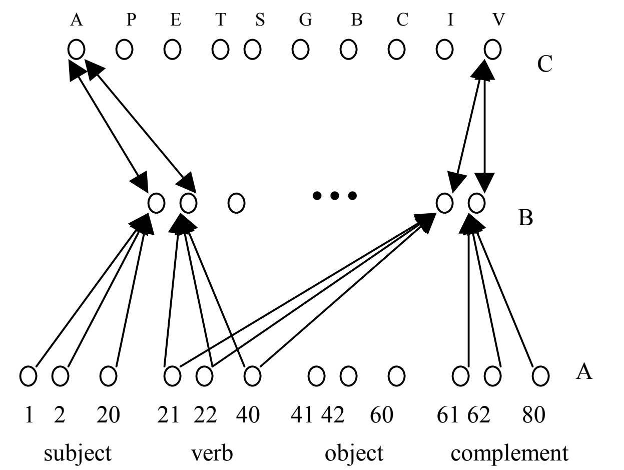


Figure 2. The three-layer bi-directional connectionist architecture of HTRP. The labels of the output units (in layer C) represent the ten thematic roles (AGENT, PATIENT, EXPERIENCER, THEME, SOURCE, GOAL, BENEFICIARY, CAUSE, INSTRUMENT, and VALUE). To the input layer A the words, represented by their distributed microfeatures, are entered sequentially at their specific slot according to their syntactic category (subject, verb, object, or complement)

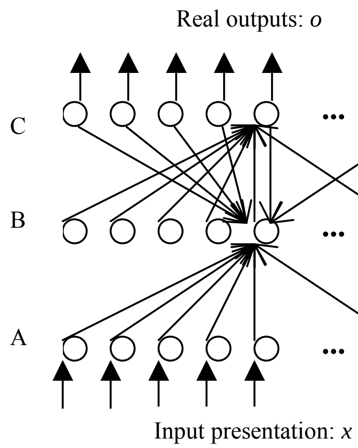
5.1 The learning procedure

The learning algorithm of HTRP-*GR* is inspired by the Recirculation [8] and GeneRec algorithms [10] (figure 3). In the minus phase, the semantic microfeature representation of the first word of a sentence is presented to the input layer. Then, there is a propagation of these stimuli x to the output through the hidden layer (bottom-up propagation). There is also a propagation of the previous actual output o, which is initially empty, back to the hidden layer (top-down propagation). Then, a hidden minus activation is generated (sum of the bottom-up and top-down propagations – through the sigmoid logistic activation function). Finally, the current real output o is generated through the propagation of the hidden minus activation to the output layer.

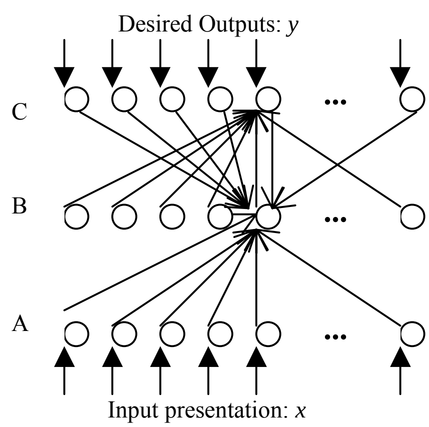
In the plus phase, there is a propagation from the input x to the hidden layer (bottom-up). After this, there is the propagation of the desired output y to the hidden layer (top-down). Then a hidden plus activation is generated, summing these two propagations. For the other words,

Proceedings of the 2003 IEEE International Conference on Systems, Man, and Cybernetics – IEEE-SMC'03. 05-08 October 2003. Washington, District of Columbia, United States of America, pp. 2845-2850.

presented one at a time, the same procedure (minus phase first, then plus phase) is repeated.



minus phase



plus phase

Figure 3. The two phases of the GeneRec algorithm [10]. In the minus phase, when input *x* is presented to layer A, there is propagation of these stimuli to the hidden layer. Then, a hidden minus signal is generated based on inputs and previous output stimuli. Then, these hidden signals propagate to the output layer C, and an actual output *o* is obtained. In the plus phase, inputs *x* are presented to layer A again; there is propagation to hidden layer. After this, desired outputs *y* are presented to the output layer and propagated back to the hidden layer, and a hidden plus signal is generated, based on inputs and on desired outputs. Recall that the architecture is bi-directional, so it is possible for the stimuli to propagate either forwardly or backwardly

In order to make learning possible, the synaptic weights are updated, considering only the local information made available by the synapse. The learning rate used in the algorithm, was considered an important variable during the experiments [7].

6 Comparing HTRP-*BP* with HTRP-*GR*

Because current connectionist models lack many properties of the biological neuron, they are considered neurophysiologically impoverished. The developers' option has been for computational efficiency rather than biological credibility. For about ten years, researchers have shown that biologically plausible systems can be as efficient as conventional models, even better [11]. This paper demonstrates that a connectionist system, with biologically plausible architecture and learning procedures, is more computationally efficient than biological implausible systems, at least regarding a particular natural language processing application.

Firstly, initial symbolic knowledge concerning thematic roles is inserted as connection weights into HTRP-*BP* and *GR* architectures (table 3). These values reflect the expected features for verbs for each thematic role, according to the linguistic theory. Then the system begins to learn, in a supervised way, through presentations of semantically sound sentence-thematic grid pairs. After training, symbolic data can be extracted from the network, revising the initial thematic symbolic theory.

Table 3. Initial symbolic knowledge inserted into the network as connection weights in HTRP-*BP* and HTRP-*GR*. Abbreviations: *y* = *yes*; *n* = *no*; *d* = *direct*; *i* = *indirect*; *s* = *source*; *g* = *goal*; *l* = *low*; *h* = *high*

THEMATIC ROLE \ *microfeatures*	*control of action*	*process triggering*	*direction*	*impacting process*	*change of state*	*psychological state*	*objective*	*effective action*	*intensity of action*	*interest on process*
AGENT	*y*	*d*		*y*			*y*			*y*
PATIENT				*y*				*y*	*h*	
EXPERIENCER			*s*		*n*		*n*	*n*	*l*	*n*
THEME					*n*				*l*	
SOURCE		*d*	*s*		*n*			*y*		*y*
GOAL	*y*		*g*		*n*			*y*		*y*
BENEFICIARY	*y*	*d*			*n*	*y*		*y*		
CAUSE	*n*	*i*	*g*				*n*			*n*
INSTRUMENT	*y*	*d*		*y*		*n*	*y*	*y*	*h*	*y*
VALUE	*y*	*d*			*n*			*y*		*y*

6.1 Training

A sentence generator, through a given lexicon, generates syntactic and semantic sound sentences in order to train the system. It is crucial to the comparison of the two versions that the same set of sentences is generated for

Proceedings of the 2003 IEEE International Conference on Systems, Man, and Cybernetics – IEEE-SMC'03. 05-08 October 2003. Washington, District of Columbia, United States of America, pp. 2845-2850.

both versions. After about 1,000 epochs, reaching an average output error of 10^{-3} (according to the *average squared error energy* formula [7]), the system is able to display the learned thematic grid for an input sentence.

6.2 Extracted symbolic data

For data extraction, the network connections are assessed and real numbers are obtained, corresponding to connection weights. The symbolic knowledge thus extracted from the connectionist architecture corresponds to the network learning and generalization capacities. As a consequence, the network is able to "revise" the initial symbolic theory [13].

Table 4 compares the connection weights extracted from the network in HTRP-*BP* and HTRP-*GR* (shadowy), regarding five thematic roles: AGENT, PATIENT, EXPERIENCER, THEME, and CAUSE, between the input and the hidden layers. As one can see, there are many significant differences between both versions. The numbers in bold show the "winner" microfeature inside each dimension. To arrive at this greater value, one should consider the difference between the two features inside a dimension.

In order to make understandable the numbers that appear on table 4, one thing is important to clarify. The dimensions are composed of two real numbers (see table 1). Inside each dimension (for instance, *ca – control of action* and *nc – no control of action* are values of the same dimension), what really matters is the difference between the values, that is, for AGENT, *control of action* has 1.7 in *BP* version and 4.3 in *GR* version (shadowy). Notice also that the values inside a dimension have, at most cases, opposite signs.

The AGENT is someone who is supposed to have *control of the action*, the process should have *direct triggering*, be *impacting*, shows *no psychological state*, be *effective* and the subject should have *interest on process*. One can notice that these features are more highlighted in the biologically plausible system (*GR* version – shadowy on table 4), because the weight differences inside each dimension are more distinguishable, for instance, control of action is more highlighted in *GR* (difference = 4.3) than in *BP* version (difference = 1.7). Other features appear with significant differences: *direct triggering* (5.5 in *GR* and 2.0 in *BP*); *impacting process* (1.5 in *GR* and 0.6 in *BP*); *no psychological state* (1.6 in *GR* and 0.5 in *BP*); *objective* (3.6 in *GR* and 1.4 in *BP*); *effective action* (1.9 in *GR* and 0.5 in *BP*), and *interest on process* (5.5 in *GR* and 2.0 in *BP*).

Table 4. Weights between input and hidden layers for verbs for several thematic roles in HTRP-*BP* and HTRP-*GR* (shadowy). The values in bold represent the significant values in each dimension. Abbreviations: AGE = AGENT; PAT = PATIENT; EXP = EXPERIENCER; THE = THEME; CAU = CAUSE; *ca = control of action*; *nc = no control of action*; *dt = direct process triggering*; *it = indirect process triggering*; *ds = direction to source*; *dg = direction to goal*; *im = impacting process*; *ni = no impacting process*; *cs = change of state*; *ns = no change of state*; *ps = psychological state*; *np = no psychological state*; *ob = objective*; *no = no objective*; *ef = effective action*; *ne = no effective action*; *hi = high intensity of action*; *li = low intensity of action*; *ip = interest on process*; *nm = no interest on process*

AGE	*ca*	*nc*	*dt*	*it*	*ds*	*dg*	*im*	*ni*	*cs*	*ns*
BP	**1.0**	-0.7	**1.1**	-0.9	-0.0	0.1	0.4	-0.2	-0.1	0.1
AGE	*ps*	*np*	*ob*	*no*	*ef*	*ne*	*hi*	*li*	*ip*	*nm*
BP	-0.2	0.3	**0.8**	-0.6	0.3	-0.2	-0.2	0.3	**1.1**	-0.9
AGE	*ca*	*nc*	*dt*	*it*	*ds*	*dg*	*im*	*ni*	*cs*	*ns*
GR	**2.1**	-2.2	**2.7**	-2.8	-0.5	0.3	**0.7**	-0.8	-0.4	0.2
AGE	*ps*	*np*	*ob*	*no*	*ef*	*ne*	*hi*	*li*	*ip*	*nm*
GR	-0.9	**0.7**	**1.8**	-1.8	**0.8**	-1.1	-0.7	**0.5**	**2.7**	-2.8
PAT	*ca*	*nc*	*dt*	*it*	*ds*	*dg*	*im*	*ni*	*cs*	*ns*
BP	-0.1	0.2	0.0	0.0	-0.7	**0.9**	**0.5**	-0.3	0.1	-0.1
PAT	*ps*	*np*	*ob*	*no*	*ef*	*ne*	*hi*	*li*	*ip*	*nm*
BP	-0.2	0.4	0.0	0.0	**0.9**	-0.7	**0.9**	-0.7	0.0	0.0
PAT	*ca*	*nc*	*dt*	*it*	*ds*	*dg*	*im*	*ni*	*cs*	*ns*
GR	**1.1**	0.1	**1.5**	-0.4	-1.6	**2.9**	**2.2**	-0.9	0.4	0.8
PAT	*ps*	*np*	*ob*	*no*	*ef*	*ne*	*hi*	*li*	*ip*	*nm*
GR	-0.9	**2.2**	**1.3**	-0.2	**3.9**	-2.6	**2.6**	-1.3	**1.5**	-0.4
EXP	*ca*	*nc*	*dt*	*it*	*ds*	*dg*	*im*	*ni*	*cs*	*ns*
BP	-0.0	-0.1	-0.3	0.2	**1.0**	-1.0	-0.6	**0.5**	-0.0	0.4
EXP	*ps*	*np*	*ob*	*no*	*ef*	*ne*	*hi*	*li*	*ip*	*nm*
BP	**0.8**	-0.5	-0.2	0.1	-1.2	**1.1**	-1.0	**0.9**	-0.3	0.2
EXP	*ca*	*nc*	*dt*	*it*	*ds*	*dg*	*im*	*ni*	*cs*	*ns*
GR	-2.1	**2.1**	-2.7	**2.7**	**0.5**	-0.6	-0.8	**0.8**	**0.5**	-0.2
EXP	*ps*	*np*	*ob*	*no*	*ef*	*ne*	*hi*	*li*	*ip*	*nm*
GR	**1.0**	-0.6	-1.8	**1.7**	-1.0	**1.0**	**0.6**	-0.7	-2.7	**2.7**
THE	*ca*	*nc*	*dt*	*it*	*ds*	*dg*	*im*	*ni*	*cs*	*ns*
BP	0.2	-0.1	0.1	-0.0	**0.8**	-0.7	-0.1	0.2	-0.0	**0.6**
THE	*ps*	*np*	*ob*	*no*	*ef*	*ne*	*hi*	*li*	*ip*	*nm*
BP	0.4	-0.3	0.2	-0.1	-0.8	**0.8**	-0.7	**1.2**	0.1	-0.0
THE	*ca*	*nc*	*dt*	*it*	*ds*	*dg*	*im*	*ni*	*cs*	*ns*
GR	**2.8**	-1.5	**2.7**	-1.5	**3.1**	-1.9	0.1	**1.1**	0.5	1.2
THE	*ps*	*np*	*ob*	*no*	*ef*	*ne*	*hi*	*li*	*ip*	*nm*
GR	0.9	0.3	**2.6**	-1.4	-1.5	**2.8**	-1.9	**3.6**	**2.7**	-1.5
CAU	*ca*	*nc*	*dt*	*it*	*ds*	*dg*	*im*	*ni*	*cs*	*ns*
BP	-0.6	**1.0**	-0.9	**1.3**	0.1	0.3	-0.1	0.4	0.1	0.2
CAU	*ps*	*np*	*ob*	*no*	*ef*	*ne*	*hi*	*li*	*ip*	*nm*
BP	0.4	-0.1	-0.7	**1.1**	0.0	0.3	0.4	-0.2	-0.9	**1.3**
CAU	*ca*	*nc*	*dt*	*it*	*ds*	*dg*	*im*	*ni*	*cs*	*ns*
GR	-3.1	**1.8**	-2.8	**1.4**	-2.9	**1.6**	-0.3	-1.2	-0.2	-1.3
CAU	*ps*	*np*	*ob*	*no*	*ef*	*ne*	*hi*	*li*	*ip*	*nm*
GR	-1.2	-0.3	-2.2	**0.9**	**1.1**	-2.6	**2.1**	-3.6	-2.8	**1.4**

For the other thematic roles, similar results can be observed. This outcome is very representative since it shows that a connectionist system with a bi-directional architecture and an algorithm which are supposed to be more biologically realistic reveals that the "symbolic" data extracted from the connectionist architecture confirm, in a more consistent way, the semantic features expected for each thematic role.

7 Conclusion

HTRP-*BP* and HTRP-*GR* are symbolic-connectionist hybrid approaches to natural language processing. In these approaches, the advantages of symbolic systems are combined with the benefits of connectionism to yield a more discriminating thematic role processing.

This paper aims to show that a biologically plausible symbolic-connectionist hybrid system, consisting of a bi-directional architecture and a learning algorithm that uses only local information to update its weights, is able not only to take care of a 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 symbolic data extracted from the connectionist architecture, reflecting the semantic features expected for ten thematic roles taught to the system.

References

[1] F. H. C. Crick, "The Recent Excitement about Neural Networks", *Nature* 337, pp. 129-132, 1989.

[2] A. S. d'Avila Garcez, K. Broda, and D. M. Gabbay, "Symbolic Knowledge Extraction from Trained Neural Networks: A Sound Approach", *Artificial Intelligence* 125, pp. 155-207, 2001.

[3] D. Dowty, "On the Semantic Content of the Notion of 'Thematic Role'", in G. Chierchia, B. H. Partee, and R. Turner (Eds.), *Properties, Types and Meaning*, Dordrecht, Kluwer, 1989.

[4] G. M. Edelman, and G. Tononi, *A Universe of Consciousness – How Matter Becomes Imagination*, Basic Books, 2000.

[5] J. L. Elman, "Finding Structure in Time", *Cognitive Science* 14, pp. 179-211, 1990.

[6] L. Haegeman, *Introduction to Government and Binding Theory*, Blackwell, 1991.

[7] S. Haykin, *Neural Networks - A Comprehensive Foundation*, 2nd edition, Prentice Hall, Upper Saddle River, New Jersey, 1999.

[8] G. E. Hinton, and J. L. McClelland, "Learning Representations by Recirculation", in D. Z. Anderson (Ed.), *Neural Information Processing Systems*, 1987. New York: American Institute of Physics, 1988, pp. 358-366.

[9] J. L. McClelland and A. H. Kawamoto, "Mechanisms of Sentence Processing: Assigning Roles to Constituents of Sentences", in J. L. McClelland and D. E. Rumelhart (Eds.), *Parallel Distributed Processing*, Vol. 2, A Bradford Book, MIT Press, 1986.

[10] R. C. O'Reilly, "Biologically Plausible Error-driven Learning using Local Activation Differences: The Generalized Recirculation Algorithm", *Neural Computation*, 8:5, pp. 895-938, 1996.

[11] R. C. O'Reilly, and Y. Munakata, *Computational Explorations in Cognitive Neuroscience – Understanding the Mind by Simulating the Brain*, A Bradford Book, The MIT Press, Cambridge, Massachusetts, 2000.

[12] J. L. G. Rosa, "An Artificial Neural Network Model Based on Neuroscience: Looking Closely at the Brain", in V. Kůrková, N. C. Steele, R. Neruda, and M. Kárný (Eds.), *Artificial Neural Nets and Genetic Algorithms - Proceedings of the International Conference in Prague*, Czech Republic, 2001 – ICANNGA-2001. April 22-25, Springer-Verlag, pp. 138-141, 2001.

[13] J. L. G. Rosa, "HTRP II: Learning Thematic Relations from Semantically Sound Sentences", in *Proceedings of the 2001 IEEE International Conference on Systems, Man, and Cybernetics – IEEE-SMC2001*, October 7-10, 2001, Tucson, Arizona, USA, pp. 488-493, 2001.

[14] J. L. G. Rosa, "A Biologically Inspired Connectionist System for Natural Language Processing", in Proceedings of the 2002 VII Brazilian Symposium on Neural Networks (SBRN 2002). 11-14 November 2002. Recife, Brazil. IEEE Computer Society Press.

[15] J. L. G. Rosa and E. Françozo, "Hybrid Thematic Role Processor: Symbolic Linguistic Relations Revised by Connectionist Learning", in *Proceedings of IJCAI'99 – Sixteenth International Joint Conference on Artificial Intelligence*, Volume 2, Stockholm, Sweden, 31 July-6 August, pp. 852-857. Morgan Kaufmann, 1999.