Linguistic Relations Encoding in a Symbolic-Connectionist Hybrid Natural Language Processor

João Luís Garcia Rosa[1] and Edson Françozo[2]

[1] Instituto de Informática, PUC-Campinas, Rod. D. Pedro I, km. 136 - Caixa Postal 317
13086-900 - Campinas - SP - Brazil
joaol@ii.puc-campinas.br
[2] LAFAPE, Instituto de Estudos da Linguagem, Unicamp - Caixa Postal 6045
13081-970 - Campinas - SP - Brazil
edson@iel.unicamp.br

Abstract. In recent years, the Natural Language Processing scene has witnessed the steady growth of interest in connectionist modeling. The main appeal of such an approach is that one does not have to determine the grammar rules in advance: the learning abilities displayed by such systems take care of input regularities. Better and faster learning can be obtained through the implementation of a symbolic-connectionist hybrid system. Such system combines the advantages of symbolic approaches, by introducing symbolic rules as network connection weights, with the advantages of connectionism. In a hybrid system called HTRP, words within a sentence are represented by means of semantic features. The features for the verbs are arranged along certain semantic dimensions, and are mutually exclusive within each dimension. One may infer that this happens because of the semantic features encoded in the network inputs.

1 Introduction

The attribution of *thematic roles* is the way linguists refer to (and theorize about) some of the semantic relations between a predicate (usually the verb) and its arguments [5]. The structure that contains all thematic roles of a sentence is called a *thematic grid*. For instance, in sentence (1)

The man gave a ball to the girl , **(1)**

there are the following thematic roles: AGENT for *the man*, THEME for *a ball*, and BENEFICIARY for *the girl*.

In a system called HTRP [9], designed to reveal the thematic grid of semantically sound sentences, individual words are represented by means of semantic features. Verbs, specially, are represented as a three-valued array of semantic microfeatures ([14], [8]) which are based on relevant features in a thematic frame [2]. Semantic microfeatures are arranged in sub-arrays along certain semantic dimensions. For verbs, each semantic dimension encompasses two elements – e.g., *control of action* and *no control of action* – and, for thematically unambiguous verbs, only one of such elements is *on*. For thematically ambiguous ones, an intermediate value is applied for

M. C. Monard and J. S. Sichman (Eds.): IBERAMIA-SBIA 2000, LNAI 1952, pp. 259-268, 2000.

the dimensions about which there is uncertainty. Thematic ambiguity here means that the same verb can reveal two different thematic grids, depending on the composition of the sentence in which it occurs. The point here is that the network learns how to represent the features within dimensions in each verb sub-array as *complementary* (one has positive sign and the other, negative), after the training step. This outcome is interesting from a cognitive perspective, since complementarity within semantic dimensions in HTRP is the key for the representation of thematic assignments.

2 The HTRP System

HTRP (for *Hybrid Thematic Role Processor*) consists of a connectionist architecture and a set of symbolic rules for thematic roles. HTRP has two versions: the first, called RIW (for *Random Initial Weight version*), is trained *without* initial knowledge, i.e., with random initial connection weights; the second, called BIW (for *Biased Initial Weight version*), is trained after initial symbolic knowledge has been fed into the system as network connection weights. This knowledge is represented as *if-then* rules based on a thematic role theory ([5], [1], [7], [2]). After training, symbolic rules are extracted from the network in the same way symbolic knowledge is input, i.e., as connection weights. One can thus say that connectionist learning revised the initial symbolic theory.

The network in HTRP has three layers. The input layer is presented with the semantic microfeatures of the words making up a sentence. The hidden layer groups the microfeatures of the verb and the microfeatures of one noun for each thematic role. The output layer gives the thematic grid (for sentence (1), the grid would be [AGENT, THEME, BENEFICIARY]). For each of the ten thematic roles implemented in HTRP, the architecture is defined as in figure 1.

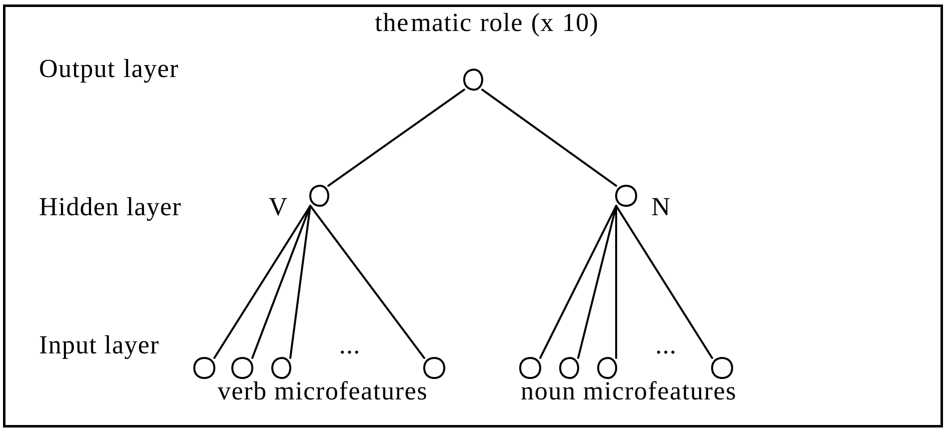

Fig. 1. Three-layer architecture used in HTRP. Verb and noun microfeatures are grouped in order to activate one thematic role

2.1 Symbolic Rules as Network Weights

In the network architecture, a link between nodes A and C with connection weight w_{AC}, and another link between B and C with connection weight w_{BC}, generates the symbolic rule (2):

$$(w_{AC} * A) + (w_{BC} * B) \rightarrow C . \quad (2)$$

For instance, for RIW the "hidden rule" between the input layer and the hidden layer, extracted after training for the thematic role AGENT, is the following:

If for verb (-0.6 * *control of action*) + (-1.0 * *direct process triggering*) + (-0.1 * *direction to goal*) + (-0.9 * *impacting process*) + (-1.1 * *change of state*) + (-0.1 * *no psychological state*) + (-2.2 * *objective*) + (-0.6 * *effective*) + (0.2 * *high intensity*) + (-0.8 * *interest on process*) *Then* V . (**3**)

Notice that the if-then rule (3) has weighted antecedents, and it is implemented in an *and* way, that is, for a unit to be on, all its inputs taken together should add up to a value which is high enough to activate it [4]. Recall that the consequent "V" of the rule is related to the hidden unit for the verb, as shown in figure 1, and the antecedents refer to semantic microfeatures for the verb (see table 2).

2.2 The Architecture of the Network

The connectionist architecture of HTRP is built from elementary processors representing eleven independent connectionist networks, one for each thematic role and one for error output. Except for the error output, each one of these networks has 40 input units, 20 for the verb and 20 for the noun, 2 hidden units (V and N) and 1 output unit. The input units are responsible for the representation of two words of the sentence: the verb and a noun. Since each sentence in HTRP has, at most, three nouns (arguments) and a verb (predicate), each sentence uses at most three connectionist networks, in order to arrive at a thematic role grid. The first hidden unit (V) represents the conjunction of all verb microfeatures and the second (N), the conjunction of all noun microfeatures. The output unit combines these two hidden units to represent one thematic role (see figure 1).

2.2.1 The Elementary Processor

The elementary processor employed in HTRP is the classical *perceptron* [10]. The perceptron schema is depicted in figure 2, where *x* represents the input, *w* represents the connection weight associated to that input, and *sum* is given by $\sum^{n}_{1} w_i x_i$. The activation function used is the sigmoid; that is, the output of the elementary processor is given by

$$\text{output} = 1 / (1 + e^{-\text{sum}}) . \quad (4)$$

2.2.2 The Error Output

An error output is implemented in order to account for sentences such as

The stone bought the man . (**5**)

Barred metaphor, (5) is clearly anomalous and will cause HTRP to activate its error output. It has already been argued that learning grammar is impossible without negative examples, and the error output grants HTRP with such property [6].

As for architecture, the error output, which also has two hidden units, differs from the other networks at the input layer. It has 80 units (20 for the verb and 60 for nouns), instead of 40, since it is unknown which nouns, in conjunction with the verb, activate the error output.

2.3 How the System Works

After the introduction of the initial symbolic rules as connection weights, the networks begins to learn the input sentences during 3,000 cycles of activation (the initial symbolic knowledge for thematic roles can be seen in table 1). A sentence generator generates the input sentences. As soon as the training is over, symbolic rules can be obtained from the connectionist architecture by running an extraction procedure ([3], [12], [13]).

3 Verb Microfeatures and Complementarity in HTRP

The representations used by HTRP are based on McClelland and Kawamoto's [8] and Waltz and Pollack's [14] notion of semantic microfeature. For the verb, the representation is mainly derived from Franchi and Cançado [2]. Twenty binary semantic microfeature units take care of each noun or verb. For verbs, pairs of microfeatures are grouped together into ten different sub-arrays – the semantic dimensions (see table 2).

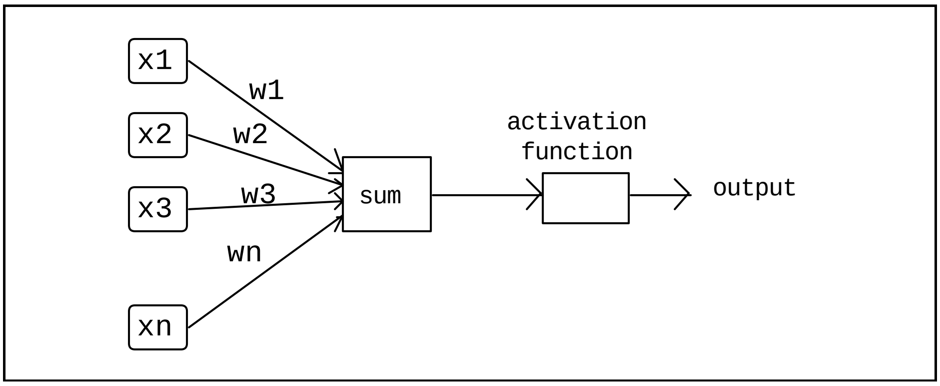

Fig. 2. The perceptron, where x_i represents the input *i*, w_i represents the connection weight associated to the input *i*, and *sum* is given by $\sum^n_1 w_i x_i$. The *activation function* used is the sigmoid; that is, the *output* is given by equation (4)

Table 1. Initial symbolic knowledge fed into the network as connection weights in BIW, for each one of the ten thematic roles. The '-' stands for a very small greater-than-zero value

THEMATIC ROLE / *microfeatures*	*control*	*process triggering*	*direction*	*impacting process*	*change of state*	*psychological state*	*objective*	*effective action*	*intensity of action*	*interest on process*
AGENT	yes	direct	-	yes	-	-	yes	-	-	yes
BENEFICIARY	yes	direct	-	-	no	yes	-	yes	-	-
CAUSE	no	indirect	goal	-	-	-	no	-	-	no
EXPERIENCER	-	-	source	-	no	-	no	no	low	no
GOAL	yes	-	goal	-	no	-	-	yes	-	yes
INSTRUMENT	yes	direct	-	yes	-	no	yes	yes	high	yes
PATIENT	-	-	-	yes	-	-	-	yes	high	-
SOURCE	-	direct	source	-	no	-	-	yes	-	yes
THEME	-	-	-	-	no	-	-	-	low	-
VALUE	yes	direct	-	-	no	-	-	yes	-	yes

Table 2. Twenty semantic microfeatures for verbs grouped in ten semantic dimensions

dimension	*positive weight*	*negative weight*
D1	control of action	no control of action
D2	direct process triggering	indirect process triggering
D3	direction to source	direction to goal
D4	impacting process	no impacting process
D5	change of state	no change of state
D6	psychological state	no psychological state
D7	objective action	no objective action
D8	effective action	no effective action
D9	high intensity of action	low intensity of action
D10	interest on process	no interest on process

3.1 The Complementarity of the Semantic Microfeatures

For thematically unambiguous words, in each of the semantic dimensions in the verb representation, the microfeatures are mutually exclusive – one feature is *on* and the other is *off* (see the microfeatures of each verb in table 3). The network is trained with several different sentences on a supervised error backpropagation procedure [11]. The expected outputs for each verb (thematic grids) are given in table 4. After learning is over, the system is able to categorize on the basis of the complementarity of the microfeatures for most of the semantic dimensions. For instance, *control of action* assumes two opposite values: 'yes' or 'no'; *process triggering* can either be 'direct' or 'indirect', and so on. In figure 3, the "hidden symbolic rule" extracted for the thematic role AGENT in BIW is shown (full line represents connection with positive weight and dotted line, negative weight). Notice that, except for *effective action*, all the other items are complementary. In this case, one can conclude that the system took *effective action* to be irrelevant, at least as far as the training sentences are concerned.

It is important to notice that this is not a trivial result. For all inputs, during the learning step, in one cycle the *on* value in a specific semantic microfeature is presented to the input layer of the network and in another cycle, the *off* value is presented. But this does not imply that positive or negative weights, respectively, are necessarily arrived at by the network. If it were so, *effective action* should have followed the regularity displayed by the other nine dimensions. The fact that it did not behave as expected shows that the architecture developed here is able to *discover* the complementarity of its inputs, based mainly on their encoded semantic features.

4 The Internal Representation of Sentences in HTRP

In order to allow for a better understanding of the claim concerning complementarity, this section examines the behavior of the network regarding the input sentences and their internal representations.

Table 3. HTRP verb microfeatures. For thematically ambiguous verbs there are two possible readings, for instance, *break1* and *break2.* In this case, the "?" stands for unknown value for the default reading

verb / micro-feature	*control of action*	*process triggering*	*direction*	*impacting process*	*change of state*	*psychological state*	*objective action*	*effective action*	*intensity of action*	*interest on process*
break	?	?	goal	yes	yes	no	?	yes	high	?
break1	no	indirect	goal	yes	yes	no	no	yes	high	no
break2	yes	direct	goal	yes	yes	no	yes	yes	high	yes
buy	yes	direct	source	yes	no	no	yes	yes	low	yes
buy1	yes	direct	source	yes	no	no	yes	yes	low	yes
buy2	yes	direct	source	yes	no	no	yes	yes	low	yes
deliver	yes	direct	goal	yes	no	no	yes	yes	low	yes
fear	no	indirect	source	yes	no	yes	no	no	low	no
frighten	?	?	goal	yes	no	yes	?	no	low	?
frighten1	no	indirect	goal	yes	no	yes	no	no	low	no
frighten2	yes	direct	goal	yes	no	yes	yes	no	low	yes
give	yes	direct	goal	yes	no	yes	yes	yes	low	yes
hit	?	?	goal	yes	no	no	?	yes	high	?
hit1	no	indirect	goal	yes	no	no	no	yes	high	no
hit2	yes	direct	goal	yes	no	no	yes	yes	high	yes
love	?	indirect	source	no	no	yes	no	no	low	no
love1	no	indirect	source	no	no	yes	no	no	low	no
love2	yes	indirect	source	no	no	yes	no	no	low	no

Table 4. The “thematic grids” for each verb of HTRP training sentences

verb / THEMATIC ROLE	*break1*	*break2*	*buy1*	*buy2*	*deliver*	*fear*	*frighten1*	*frighten2*	*give*	*hit1*	*hit2*	*love1*	*love2*
AGENT		*	*	*	*			*	*		*		
BENEFICIARY									*				
CAUSE	*						*			*			
EXPERIENCER						*						*	*
GOAL					*								
INSTRUMENT		*									*		
PATIENT	*	*								*	*		
SOURCE			*										
THEME			*	*	*	*	*	*	*			*	*
VALUE				*									

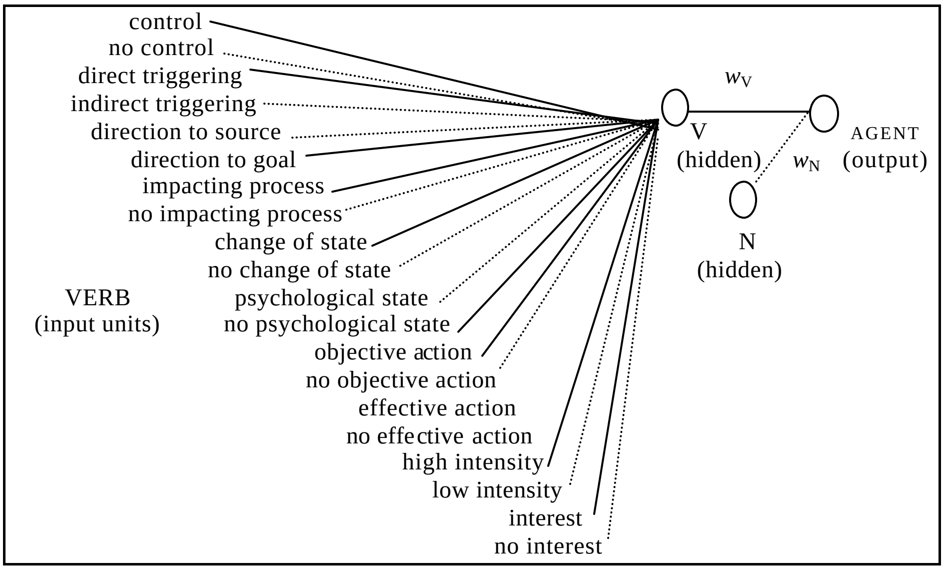

Fig. 3. “Hidden symbolic rule” extracted from the network for the thematic role AGENT in BIW, after 3,000 cycles of activation (full line represents connection with positive weight and dotted line, negative weight). Notice that this figure is also showing the output unit for AGENT, which has two inputs from the hidden layer: V for verb and N for noun. w_V and w_N are the connection weights between V and the output unit and between N and the output unit, respectively. The inputs for N are not relevant here

Table 5 presents initial and after-training hidden verb weights for the thematic role AGENT. The semantic dimensions (D*i*) are represented as pairs of elements (a sub-array), e.g., D1 = *control of action* (*ca*) and *no control of action* (*nc*). The initial hidden weights are the values obtained from symbolic knowledge in BIW for some dimensions before training.

4.1 The Signs of the Dimensions

From the hidden weights for RIW and for BIW, one can notice that the majority of dimensions are complementary in the sense that they have opposite signs. *Hidden weights*, in table 5, are the connection weights between the input and the hidden layers (see table 6 for the output weights, that is, the weights between hidden and output units). For instance, $ca < 0$ and $nc > 0$ for RIW, and $ca > 0$ and $nc < 0$ for BIW. In this case, the weight between hidden unit and output unit for the verb (shown, in figure 3, as w_V) is negative in RIW and positive in BIW, as can be seen in table 6. This means that *control of action* is a feature to be associated to AGENT, since $ca < 0$ and $w_V < 0$ in RIW (negative signs cancel out), and $ca > 0$ and $w_V > 0$ in BIW. Notice that the signs depend on the weight of the connection between the hidden unit and the output unit (w_V), which is a demonstration that, for HTRP, thematic roles assignment is achieved on the basis of the dynamic relationships between verb and nouns.

4.2 The Sentence Generator

HTRP employs a sentence generator for the training step. Instead of entering the sentences by hand, they are generated automatically by a seven-frame set for each one of the thirteen verbs (eight different verbs and five alternative readings). Each frame set includes two semantically anomalous sentences. As an example, see the frames and their thematic grids for the two readings of the thematically ambiguous verb *frighten* in table 7. The generator replaces the categories present in frames by the words for each category given in table 8.

Table 5. A comparison between initial and after-training hidden weights (weights between input and hidden layers) for verbs for the thematic role AGENT. Abbreviations: *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 action*; *no* = *no objective action*; *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*. The '-' stands for a not significant value

thematic role: AGENT	*ca*	*nc*	*dt*	*it*	*ds*	*dg*	*im*	*ni*	*cs*	*ns*
BIW initial	0.2	-	0.2	-	-	-	0.2	-	-	-
RIW after-training	-0.6	2.8	-1.0	1.4	1.1	-0.1	-0.9	0.1	-1.1	-
BIW after-training	0.9	-0.8	1.2	-1.2	-0.9	0.8	0.5	-0.4	0.4	-0.5
thematic role: AGENT	*ps*	*np*	*ob*	*no*	*ef*	*ne*	*hi*	*li*	*ip*	*nm*
BIW initial	-	-	0.2	-	-	-	-	-	0.2	-
RIW after-training	0.1	-0.1	-2.2	-	-0.6	0.6	0.2	-	-0.8	2.0
BIW after-training	-0.2	0.1	1.2	-1.2	-0.1	0.0	0.2	-0.3	1.2	-1.2

Table 6. Initial and after-training output weights (weights between the hidden layer and the output layer) for verbs (w_V) and nouns (w_N) for the thematic role AGENT

thematic role: AGENT	w_V	w_N
BIW initial output weight	0.5	0.5
RIW after-training output weight	-7.3	6.9
BIW after-training output weight	7.1	-7.1

Table 7. The frames of the sentence generator for the two readings of verb *frighten* and their thematic grids in HTRP

	frame for frighten1	*thematic grid*
1	the object frightens the human	[CAUSE, THEME]
2	the predator frightens the prey	[CAUSE, THEME]
3	the thing frightens the animal	[CAUSE, THEME]
4	the value frightens the object	*error*
5	the animal frightens the human	[CAUSE, THEME]
6	the object frightens the human	[CAUSE, THEME]
7	the value frightens the object	*error*
	frame for frighten2	*thematic grid*
1	the human frightens the human	[AGENT, THEME]
2	the human frightens the animal	[AGENT, THEME]
3	the human frightens the human	[AGENT, THEME]
4	the value frightens the value	*error*
5	the human frightens the human	[AGENT, THEME]
6	the human frightens the animal	[AGENT, THEME]
7	the value frightens the value	*error*

Table 8. The categories for the frames in the sentence generator (table 7)

category	*word 1*	*word 2*	*word 3*	*word 4*
animal	chicken	dog	wolf	monkey
human	man	girl	boy	woman
object	ball	jack	doll	dish
predator	wolf	dog	wolf	dog
prey	chicken	monkey	chicken	monkey
thing	doll	chicken	mango	vase
value	ten	hundred	thousand	ten

5 Conclusion

In connectionist Natural Language Processing systems, the words belonging to a sentence must be represented in such a way as to keep the meaning of the words and, at the same time, to be useful for the network to develop significant internal representations.

Even without initial prompting (in RIW), HTRP is able to classify and categorize the intended mutually exclusive microfeatures within a semantic dimension, and subsequently to adjust the weights connecting hidden units to output units in order to correctly reveal the thematic assignment for each pair verb-noun in a sentence. This is attributed to the fact that the network architecture, with no initial biasing, induces the connection weights related to pairs of semantic features to be taken as complementary, in quite the same way as the version with initial symbolic knowledge does (see initial knowledge of HTRP in table 1). That is, some sort of internal representation of implications has been developed for thematic roles, which are not introduced as input to the network.

References

1. Dowty, D.: On the Semantic Content of the Notion of 'Thematic Role', in G. Chierchia, B. H. Partee, and R. Turner (Eds.), Properties, Types, and Meaning, vol. 2, Semantic Issues, Dordrecht: Kluwer (1989) 69-129
2. Franchi, C. and Cançado, M.: Thematic Hierarchy (in Portuguese), Unpublished paper, Unicamp/USP, UFMG, Brazil (1998)
3. Fu, L. M.: Knowledge-Based Connectionism for Revising Domain Theories, IEEE Transactions on Systems, Man, and Cybernetics, Vol. 23, No.1 (1993) 173-182
4. Garcez, A. S. d'A., Zaverucha, G., and Carvalho, L. A. V.: Connectionist Inductive Learning and Logic Programming System, System Engineering and Computing Program – COPPE/UFRJ, Rio de Janeiro, Brazil, Technical Report (1997)
5. Haegeman, L.: Introduction to Government and Binding Theory. Cambridge: Blackwell (1991)
6. Lawrence, S., Giles, C. L., and Fong, S.: Natural Language Grammatical Inference with Recurrent Neural Networks. IEEE Transactions on Knowledge and Data Engineering, Vol. 12, No. 1 (2000) 126-140
7. McRae, K., Ferretti, T. R., and Amyote, L.: Thematic Roles as Verb-specific Concepts, Language and Cognitive Processes, 12 (2/3) (1997) 137-176
8. McClelland, J. L. and Kawamoto, A. H.: Mechanisms of Sentence Processing: Assigning Roles to Constituents of Sentences. In J. L. McClelland, D. E. Rumelhart (Eds.), Parallel Distributed Processing, Volume 2. A Bradford Book, The MIT Press (1986)
9. Rosa, J. L. G. and Françozo, E.: Hybrid Thematic Role Processor: Symbolic Linguistic Relations Revised by Connectionist Learning. Proceedings of IJCAI'99 – Sixteenth International Joint Conference on Artificial Intelligence, Volume 2, Stockholm, Sweden, 31 July-6 August, Morgan Kaufmann (1999) 852-857
10. Rosenblatt, F.: The Perceptron: A Perceiving and Recognizing Automaton, Report 85-460-1, Project PARA, Cornell Aeronautical Laboratory, Ithaca, New York (1957)
11. Rumelhart, D. E., Hinton, G. E., and Williams, R. J.: Learning Internal Representations by Error Propagation, in D. E. Rumelhart and J. L. McClelland (Eds.), Parallel Distributed Processing - Volume 1: Foundations, A Bradford Book, The MIT Press (1986)
12. Setiono, R. and Liu, H.: Symbolic Representation of Neural Networks, IEEE Computer, Vol. 29, No. 3 (1996) 71-77
13. Towell, G. G. and Shavlik, J. W.: Extracting Refined Rules from Knowledge-based Neural Networks, Machine Learning, 13 (1993) 71-101
14. Waltz, D. L. and Pollack, J. B.: Massively Parallel Parsing: A Strongly Interactive Model of Natural Language Interpretations. Cognitive Science 9 (1985) 51-74