Proceedings of the 2001 IEEE International Conference on Systems, Man, and Cybernetics - SMC2001
October 7-10, 2001, Tucson, Arizona, USA, 488-493.

HTRP II: Learning thematic relations from semantically sound sentences

JOÃO LUÍS GARCIA ROSA

Instituto de Informática – PUC-Campinas
Rodovia D. Pedro I, km. 136 – 13086-900 – Campinas – SP – Brasil – joaol@ii.puc-campinas.br

Abstract

The system HTRP – Hybrid Thematic Role Processor – is a symbolic-connectionist hybrid system, combining the advantages of symbolic approaches with the advantages of connectionism, in order to process the thematic roles, the semantic relations between words in a sentence. However, HTRP has some limitations: the sentences must be broken into verb-noun pairs to be presented to the network. This makes it impossible for the system to deal with instances in which constraints are operative not only between the verb and one of its arguments (nouns), but also between two arguments of the same verb. Another possible drawback is training with negative examples (semantically unsound sentences). Although many researchers point out that negative inputs are necessary for a system to learn a grammar, several authors believe that, under certain circumstances, a network is able to learn in absence of negative examples. From a psycholinguistic standpoint, especially regarding language acquisition, explicit negative evidence is hardly to be expected as part of the cognitive environment. In this paper, new versions of HTRP are proposed (HTRP II) to account for the whole sentence as input with no negative examples provided during training.

Keywords
Natural language processing, neural networks, machine learning.

1 Introduction

The Natural Language Processing system called HTRP (*Hybrid Thematic Role Processor*) [12] is a symbolic-connectionist hybrid system, in which one can introduce symbolic knowledge, based on a production rule set for thematic roles, as weights of a connectionist network. After training, the system is able to reveal the thematic grid of semantically sound sentences.

HTRP was deployed in two versions: RIW – *random initial weight version* – without initial symbolic knowledge, and BIW – *biased initial weight version* – in which initial symbolic knowledge is inserted into the connectionist network as connection weights. From both versions, it is possible to extract final symbolic rules after training.

Now, second versions of HTRP are proposed. Unlike its previous versions, HTRP II receives as input the whole sentence instead of verb-noun pairs. In addition, it does not contain semantically unsound sentences in its training set. The motivation for such versions (RIW-II and BIW-II) comes from the necessity of clarifying some psycholinguistic issues, concerning essentially language acquisition. In order to achieve a complete and sound thematic processing, the words in a sentence must be related to each other, and not only to the verb. That is the reason for the presentation of whole sentences as input. Furthermore, children when learn language, are unlikely to have semantic anomalous sentences as examples. Consequently, HTRP II was designed to contain only semantically sound whole sentences in its training set.

2 Thematic roles

The Government and Binding linguistic theory [5] states that thematic roles – the semantic relations between words in a sentence – are in the lexicon, so a specific verb has a single thematic grid, the structure containing the thematic roles of a sentence. This is a “slot and filler” lexicalist view. For instance, the verb *kill* would have an AGENT (*i*) and a PATIENT (*j*), no matter in which sentence it occurs, like in *Michael$_i$* killed *Peter$_j$*. There are verbs, however, which have different thematic grids in different sentences, for instance the verb *hit* in sentences (1) and (2).

(1) The woman hit the girl
(2) The ball hit the girl

In the sentences (1) and (2), although the same verb is employed, their thematic grids are different. In sentence (1), the thematic grid is [AGENT, PATIENT] and in sentence (2), [CAUSE, PATIENT]. The reason is that *the woman*, in the intended reading of sentence (1), is supposed to have the *control of action*, that is, the intention of hitting. The same does not occur in sentence (2). *The ball* is not willing of hitting anybody. A verb that assigns two different thematic grids in different sentences is called *thematically* *ambiguous*. How to solve the thematic ambiguity of the verb *hit* in a lexicalist view?

In a non-lexicalist view (componential), one could have a representation for thematically ambiguous verbs, like *hit*, that would allow them to function as predicates in several sentence types. Taking sentences (1) and (2) again, it seems that the distinction between AGENT and CAUSE has something to do with the nouns that are assigned such roles. Thus, since only an animate noun is supposed to be an AGENT, some kind of semantic analysis is necessary in order to distinguish between different thematic assignments. In other words, thematic roles must be elements with semantic content [2]. Consequently, the words, which can fill each of the slots for a given thematic grid, should share a common semantic core.

2.1 Microfeatural representations

In HTRP II, word representation is the same as in HTRP; that is, it is adapted from the classical semantic microfeature representations used by Waltz and Pollack [16] and McClelland and Kawamoto [9]. Twenty three-valued logic semantic microfeature units account for each noun and verb. The schema in table 1 displays the semantic features for verbs. Table 2 shows the microfeatures for nouns.

control of action	no control of action
direct process triggering	indirect 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 1. The ten semantic microfeature dimensions for verbs. The left column represents the positive weights for BIW-II.

It is important to notice here that the verb microfeatures are chosen in order to encompass the semantic issues considered relevant in a thematic frame. The microfeatures outside this thematic context are not meaningful. They only make sense in a system like HTRP II, where the specification of semantic relationships between the words in a sentence plays a leading role.

When the user enters a sentence with a thematically ambiguous verb like *hit* into HTRP II, the system does not know which reading of the verb is intended. The network input is the "average" of both readings. Consequently, some of the microfeatures will be undetermined. HTRP II, as its previous versions, will arrive at the missing values for the intended reading of the verb.

<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 2. The seven semantic microfeature dimensions for nouns, separated in rows. Only one value in each dimension is *on* for each noun (adapted from [9]).

3 The connectionist architecture

The previous versions of HTRP system include a connectionist architecture representing eleven independent three-layer artificial neural networks, one for each thematic role and one for the error output. Each network has forty input units, two hidden units, and one output unit. The input units are responsible for the representation of two words of a sentence, the verb and one noun. The first hidden unit represents the conjunction of all verb microfeatures, and the second the conjunction of all noun microfeatures. The output unit represents the conjunction of these two microfeature sets.

Unlike HTRP, HTRP II employs an architecture representing only one three-layer neural network with eighty input units, twenty hidden units, and ten output units, one for each of the ten thematic roles: AGENT (A), PATIENT (P), EXPERIENCER (E), THEME (T), SOURCE (S), GOAL (G), BENEFICIARY (B), CAUSE (C), INSTRUMENT (I), and VALUE (V). To the eighty-unit input layer are presented four words of a sentence (a subject, a verb, an object, and a complement). For the sake of simplicity, in the version of HTRP II *with* initial knowledge (BIW-II), the insertion and extraction of symbolic rules are related to a specific configuration with two hidden units linked to each of the ten thematic roles. That is, for the thematic role AGENT, for instance, only the first two hidden units are considered (figure 3).

This architecture does not apply to the thematic roles PATIENT and EXPERIENCER, since they may be either subjects or objects in sentences belonging to the training set. In (3), PATIENT is assigned to the subject, but in (4) it is assigned to the object.

(3) The vase broke

(4) The man broke the vase

In (5), EXPERIENCER is the subject of *love*, while in (6), it is the object of *frighten*.

(5) The man loves the woman
(6) The wolf frightened the girl

In these cases, instead of twenty inputs to the hidden unit, there are forty: twenty for the subject and twenty for the object.

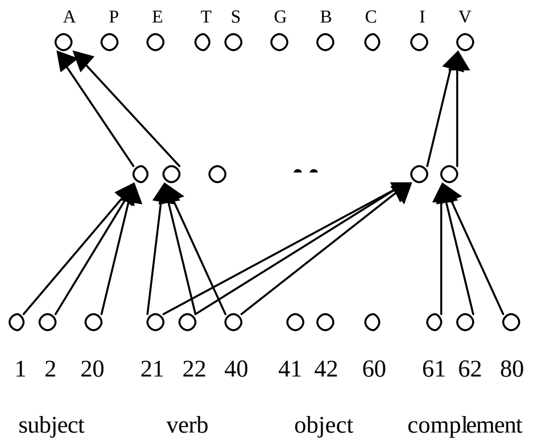

Figure 3. The connectionist architecture of HTRP II.

In the thematic theory, the words in a sentence should be related to each other, in order to arrive at the correct thematic grid. Of course, there are sentences in which one or more components are absent, like (7), where there is no object or complement, while in (8) a complete sentence is presented.

(7) The window broke
(8) The woman delivered the curtain to the boy

Sentences with the same verb can be either semantically sound or unsound. For instance, in the semantically sound sentences (9) and (10), the thematic grids would be [AGENT, PATIENT, INSTRUMENT] and [CAUSE, PATIENT], respectively, while the sentence (11) is clearly a semantically anomalous sentence, since it is unexpected that *the ball* can use a *hammer* to break an object. In this case, HTRP-II displays neither AGENT nor CAUSE for *ball*. How the system distinguishes between these different sentences with the same verb *break*? In this case it is necessary for the system to relate the subject of the sentence (*man* in sentence 9 and *ball* in sentences 10 and 11) not only with the verb *break*, but also with the complement (*hammer* in sentences 9 and 11) to conclude that only an animate noun like *man* could use a *hammer* as an instrument of breaking. Unlike HTRP, this kind of distinction is possible in HTRP II, because it is presented with the whole sentence as input.

(9) The man broke the vase with the hammer
(10) The ball broke the vase
(11) *The ball broke the vase with the hammer

4 The starting large approach

Rohde and Plaut [11] and Elman [3] discuss the importance of starting “small” or starting “large” regarding language acquisition. Elman argues that the learning is successful only in cases when sentences are getting more complex gradually or in cases when limited memory is given initially to the network. Rohde and Plaut, on the other hand, argue that starting “small” is not important: the starting of simplified inputs or limited memory is not necessary for a network to learn language, providing that semantic and syntactic constraints are introduced in the training set. That is the case of HTRP II. Semantic, and also syntactic, constraints are introduced when sentences are generated during the training step. Only semantically and syntactically well formed sentences are input to the network.

Both Elman’s and Rohde and Plaut’s systems employ the Elman network, a recurrent neural network, in order to account for sentences presented in separate words, one by one. For that reason, the system must have memory, which is achieved by the extra layer in recurrent networks. In HTRP II recurrent networks are unnecessary, since whole sentences are presented as inputs, with all the words at the same time. Consequently, HTRP II uses a simple feed-forward connectionist architecture.

Lawrence *et al*. [8] invoke the necessity of having negative examples in the training set for the network learn the correct grammatical relations between the words in sentences. But this consideration is not plausible in a language acquisition standpoint, since it is not reasonable suppose that a child receives negative examples, or semantically unsound sentences, when he or she is listening to his or her parents. Elman [3] and Rohde and Plaut [11] share this position. So, for the sake of cognitive coherence, HTRP II is implemented without the error output, and of course, without semantically unsound sentences in its training set, present in its previous versions.

5 Hybrid systems

Neural networks perform very well in many domains. But, their critics emphasize that they lack transparency, that is, one does not know how they work, how they develop internal representations. And it is known that the training step often takes too long. An answer to such criticism is the so-called Knowledge-Based Neural Networks, or Hybrid Symbolic-Connectionist Systems. In these systems 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. Since the version BIW-II of HTRP II has initial knowledge, it takes less time to train it. After training, symbolic knowledge will be available through the extraction of a 'new' (revised) symbolic theory. Now, it is possible to say how the network is working, since their connection weights represent symbolic knowledge that can be taken out of the connectionist architecture.

In a symbolic-connectionist hybrid approach, symbolic rules are 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. The symbolic knowledge generated by the net can be extracted, in HTRP II, in a way comparable to initial symbolic knowledge insertion in BIW-II. It had been proved that the set of rules and the network, from which it is extracted, are very equivalent [1].

Although many researchers believe that symbolic and connectionist systems are so different that they are irreconcilable, others emphasize that the integration of both is not only possible but also crucial for the systems understand cognition behind the computational implementations. Honavar and Uhr [7] examine the two sides of this controversial issue.

6 The system HTRP II

Firstly, initial symbolic knowledge concerning thematic roles is inserted as connection weights into HTRP II architecture. Then the system begins to learn, through presentations of semantically sound sentence-thematic grid pairs. After training, symbolic rules can be extracted from the network, revising the initial thematic symbolic theory.

6.1 Initial symbolic knowledge

The initial symbolic rules for the semantic microfeatures of the verb implemented as connection weights in HTRP II are based on a thematic theory [5, 2, 10]. They are *if-then* rules:

- If (*control of action*) and (*direct process triggering*) and (*impacting process*) and (*objective*) and (*interest on process*) then AGENT
- If (*impacting process*) and (*effective action*) and (*high intensity of action*) then PATIENT
- If (*direction to source*) and (*no change of state*) and (*no objective*) and (*no effective action*) and (*low intensity of action*) and (*no interest on process*) then EXPERIENCER
- If (*no change of state*) and (*low intensity of action*) then THEME
- If (*direct process triggering*) and (*direction to source*) and (*no change of state*) and (*effective action*) and (*interest on process*) then SOURCE
- If (*control of action*) and (*direction to goal*) and (*no change of state*) and (*effective action*) and (*interest on process*) then GOAL
- If (*control of action*) and (*direct process triggering*) and (*no change of state*) and (*psychological state*) and (*effective action*) then BENEFICIARY
- If (*no control of action*) and (*indirect process triggering*) and (*direction to goal*) and (*no objective*) and (*no interest on process*) then CAUSE
- If (*control of action*) and (*direct process triggering*) and (*impacting process*) and (*no psychological state*) and (*objective*) and (*effective action*) and (*high intensity of action*) and (*interest on process*) then INSTRUMENT
- If (*control of action*) and (*direct process triggering*) and (*no change of state*) and (*effective action*) and (*interest on process*) then VALUE

6.2 Learning

As well as in HTRP, in HTRP II the elementary processors are classical perceptron-like units, and the algorithm used is the supervised *backpropagation* [13]. A sentence generator supplies the training sentences, according to semantic and syntactic constraints, absolutely necessary for a system to learn without negative examples [11]. In the previous HTRP, after 2,186 training cycles in RIW, which corresponds to an average output error[1] of 10^{-5}, the system is able to judge, with a high degree of certainty, if a sentence is meaningful or not, and, if it is, which its thematic grid is. In BIW, the learning is reached within 1,625 cycles, for the same average error, as expected, because the learning time is decreased for systems trained with initial knowledge.

Regarding the new versions (HTRP II), for the same average output error, the ability to learn thematic relations is achieved within 2,262 training

[1] The average output error is the difference between "actual" output and "desired" output, and it is obtained from the *average squared error energy* formula [6] for each set of different sentences presented to the network.

cycles in RIW-II, and within 1,781 cycles in BIW-II. Recall that in HTRP II, the whole sentence is fed in an eighty input-unit network, and no negative examples are provided.

6.3 Extracted rules

The constrained architecture presented in figure 3 is used for rule extraction, which consists in reversing the process of initial rule insertion. The net weights are assessed and a weighted antecedent is obtained, corresponding to the connection weight. The symbolic knowledge thus extracted from the present connectionist architecture corresponds to the network learning and generalization capacities. As a consequence, the network is able to "revise" the initial symbolic rules. The rule extraction from the network, after training, is based on [4, 14, 15].

Table 4 compares the connection weights extracted from the network in HTRP (BIW) and HTRP II (BIW-II), regarding the thematic role AGENT, between the input and the hidden layers ("hidden weights"). As one can see, there are no 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, for instance, regarding *effective action* (*ef* and *ne*), in HTRP (upper table, in *italics*) the difference is so small (0.1) as insignificant. But in HTRP II (lower table), this difference (1.4) should be taken into account. Then, the AGENT appears mostly in sentences that show *effective action* of the verb. This happens because the training set of HTRP is different from HTRP II, since in the latter it had to be modified to eliminate the semantically unsound sentences.

In relation to the initial rules inserted into the network as connection weights before training, *control of action*, *direct process triggering*, *objective*, and *interest on process* were confirmed by the connectionist learning. Even *impacting process* was confirmed not so emphatically as the previous ones. Beyond these, *change of state*, *no psychological state*, and *effective action* were also highlighted. So, regarding the thematic role AGENT, new features were discovered.

HTRP II found out interesting novelties about the other thematic roles. For PATIENT, *high intensity of action* appeared as the most relevant feature, and *no psychological state*, which was not given as initial knowledge to the system, revealed importance. For EXPERIENCER, every initial feature associated was confirmed, with exception of *low intensity of action*. The system, otherwise, showed that the verb, which assigns the thematic role EXPERIENCER, has *high intensity of action*. The reason why this occurred can be attributed to the existence in the training set, of sentences with the verb *frighten*, a verb usually related to *high intensity of action*. *Frighten* has in its thematic grid, an EXPERIENCER as object. HTRP II presented a very interesting result correcting the initial microfeature input to the network.

AGE	*ca*	*nc*	*dt*	*it*	*ds*	*dg*	*im*	*ni*	*cs*	*ns*
H-I	**0.9**	-0.8	**1.2**	-1.2	-0.9	**0.8**	**0.5**	-0.4	**0.4**	-0.5
AGE	*ps*	*np*	*ob*	*no*	*ef*	*ne*	*hi*	*li*	*ip*	*nm*
H-I	-0.2	**0.1**	**1.2**	-1.2	*-0.1*	*0.0*	**0.2**	-0.3	**1.2**	-1.2

AGE	*ca*	*nc*	*dt*	*it*	*ds*	*dg*	*im*	*ni*	*cs*	*ns*
H-II	**0.3**	-1.3	**0.9**	-1.8	-0.9	**-0.2**	**-0.1**	-0.9	**0.2**	-1.3
AGE	*ps*	*np*	*ob*	*no*	*ef*	*ne*	*hi*	*li*	*ip*	*nm*
H-II	-1.3	**0.1**	**0.9**	-1.8	**0.1**	-1.3	**-0.2**	-0.9	**0.9**	-1.8

Table 4. A comparison between hidden weights (weights between input and hidden layers) for verbs for the thematic role AGENT in HTRP (*H-I*) and HTRP-II (*H-II*). The values in bold represent the greater values in each dimension. 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.*

For THEME, on the contrary, *low intensity of action* was highly confirmed. *Direction to goal* appeared also as an important feature associated to THEME. In relation to SOURCE, as expected, *direction to source* appeared with high value. And, for GOAL, *direction to goal* confirmed the initial value. Beyond this feature, *no psychological state* was also highlighted. For BENEFICIARY, *psychological state* presented a very high value. For CAUSE, INSTRUMENT, and VALUE no surprises appeared.

AGE	*hu*	*nh*	*so*	*ha*	*sm*	*me*	*la*	*1d*	*2d*	*3d*
H-II	**2.1**	-0.5	**1.6**	0.1	-0.2	0.7	1.2	-0.0	0.1	**1.6**
AGE	*po*	*ro*	*fr*	*un*	*va*	*fu*	*fo*	*to*	*tu*	*an*
H-II	0.0	**1.6**	-0.0	**1.7**	0.0	0.0	0.0	-0.2	0.0	**1.8**

Table 5. The hidden weights for BIW-II for nouns for the thematic role AGENT in HTRP-II. The values in bold represent the greater significant values in each dimension. Abbreviations: *hu* = *human*; *nh* = *no human*; *so* = *soft; ha* = *hard*; *sm* = *small*; *me* = medium; *la* = *large*; *1d* = *1-D/compact*; *2d* = *2-D*; *3d* = *3-D*; *po* = *pointed*; *ro* = *rounded*; *fr* = *fragile/breakable*; *un* = *unbreakable*; *va* = *value*; *fu* = *furniture*; *fo* = *food*; *to* = *toy*; *tu* = *tool/utensil*; *an* = *animate.*

From table 5, where the weights for nouns are displayed, one could notice that the AGENT is mostly *human*, *soft*, *3-D*, *rounded*, *unbreakable*, and *animate*. This is very interesting, since the connectionist architecture revealed what kind of AGENT it discovered, without any initial explicit knowledge supplied by a symbolic theory, in the case of nouns.

7 Conclusions

HTRP II, as its previous versions, is a symbolic-connectionist hybrid approach to natural language processing. In this approach, the advantages of symbolic systems are combined with the advantages of connectionism to yield a more discriminating thematic role processing.

Unlike McClelland and Kawamoto's system [9], in HTRP II a single network accounts for each sentence; thus generalizing over *both* nouns and verbs. In fact, this is crucial in dealing with thematic roles, for they are but the generalization of semantic relationships between verbs and nouns.

Two are the main differences between HTRP and HTRP II. First, the previous versions employ eleven independent neural networks (configuration: 40-2-1, that is, forty units in the input layer, two units in the hidden layer, and one unit in the output layer), one for each thematic role and one for the error output. The proposed HTRP II uses only one connectionist architecture (configuration: 80-20-10) to account for the ten thematic roles. In this way, the subject could influence the complement in the thematic role assignment. The other difference concerns the way the system learns: in HTRP, the training set contains 28% of negative sentences, that is, for each seven sentences presented to the network, two are semantically anomalous. In HTRP II, only semantically acceptable sentences are presented, that is, it learns in absence of negative inputs. This way, HTRP II can account for an effective thematic processing and simulate a cognitively plausible model regarding language acquisition.

Acknowledgement: I thank Edson Françozo for his collaboration on previous drafts of this paper.

References

[1] A. S. d'Avila Garcez, K. Broda, and D. M. Gabbay. Symbolic knowledge extraction from trained neural networks: A sound approach. *Artificial Intelligence* 125, 155-207, 2001.

[2] D. Dowty. On the Semantic Content of the Notion of 'Thematic Role'. In *Properties, Types and Meaning*. Edited by G. Chierchia, B. H. Partee, and R. Turner. Dordrecht, Kluwer, 1989.

[3] J. L. Elman. Learning and Development in Neural Networks: the Importance of Starting Small. *Cognition* 48, 71-99, 1993.

[4] L. M. Fu. Knowledge-Based Connectionism for Revising Domain Theories. *IEEE Trans. Systems, Man, and Cybern.*, Vol. 23, No.1, 173-182, 1993.

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

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

[7] V. Honavar and L. Uhr. Integrating Symbol Processing Systems and Connectionist Networks. In S. Goonatilake and S. Khebbal (eds.) *Intelligent Hybrid Systems*. John Wiley & Sons Ltd., 1995.

[8] S. Lawrence, C. L. Giles, and S. Fong. Natural Language Grammatical Inference with Recurrent Neural Networks. *IEEE Trans. on Knowledge and Data Engineering*, Vol. 12, No. 1, 126-140, 2000.

[9] J. L. McClelland and A. H. Kawamoto. Mechanisms of Sentence Processing: Assigning Roles to Constituents of Sentences. In *Parallel Distributed Processing, Vol. 2*. Ed. by J. McClelland and D. Rumelhart. MIT Press, 1986.

[10] K. McRae, T. R. Ferretti, and L. Amyote. Thematic Roles as Verb-specific Concepts. *Lang. and Cognitive Processes*, 12 (2/3), 137-176, 1997.

[11] D. L. T. Rohde and D. C. Plaut. Language Acquisition in the Absence of Explicit Negative Evidence: How Important is Starting Small? *Cognition* 72, 67-109, 1999.

[12] J. L. G. Rosa and E. Françozo. 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, 852-857, 1999.

[13] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning Internal Representations by Error Propagation. In Rumelhart and McClelland (Eds.), *Parallel Distributed Processing, Vol. 1*. A Bradford Book, MIT Press, 1986.

[14] R. Setiono and H. Liu. Symbolic Representation of Neural Networks. *IEEE Computer*, Vol. 29, No. 3, 71-77, 1996.

[15] G. Towell and J. W. Shavlik. Extracting Refined Rules from Knowledge-Based Neural Networks. *Machine Learning*, 13, 71-101, 1993.

[16] D. L. Waltz and J. B. Pollack. Massively Parallel Parsing: A Strongly Interactive Model of Natural Language Interpretations. *Cognitive Science* 9, 51-74, 1985.