Hybrid Thematic Role Processor: Symbolic Linguistic Relations Revised by Connectionist Learning

Joao Luis Garcia Rosa
Institute* de Informatica - PUC-Campinas
Rodovia D. Pedro I, km. 136 - C P. 317
13020-904 - Campinas - SP - Brasil
joaol @ zeus.puceamp.br

Edson Franeozo
LAFAPE - BEL - Unicamp
Caixa Postal 6045
13081-970 - Campinas - SP - Brasil
edson@ieLunieamp.br

Abstract

In linguistics, the semantic relations between words in a sentence are accounted for, *inter alia,* as the assignment of *thematic roles,* e.g. AGENT, INSTRUMENT, etc. As in predicate logic, simple linguistic expressions are decomposed into one predicate (often the verb) and its arguments. The predicate assigns thematic roles to the arguments, so that each sentence has a *thematic grid,* a structure with all thematic roles assigned by the predicate. In order to reveal the thematic grid of a sentence, a system called HTRP *(Hybrid Thematic Role Processor)* is proposed, in which the connectionist architecture has, as input, a featural representation of the words of a sentence, and, as output, its thematic grid. Both a random initial weight version (RIW) and a biased initial weight version (BIW) are proposed to account for systems *without* and *with* initial knowledge, respectively. In BIW, initial connection weights reflect symbolic rules for thematic roles. For both versions, after supervised training, a set of final symbolic rules is extracted, which is consistently correlated to linguistic - symbolic - knowledge. In the case of BIW, this amounts to a revision of the initial rules. In RIW, symbolic rules seem to be induced from the connectionist architecture and training.

1 Introduction

In sentences such as

(.1) The man broke the window with the stone,

one can intuitively find an AGENT *(the man),* a PATIENT *(the window),* and an INSTRUMENT *(the stone).* Linguistic theory [Haegeman, 1991] refers to the roles words usually have in relation to a verb as *thematic roles,* so that one can say that *break* has a thematic structure with the following roles [AGENT, PATIENT, INSTRUMENT], in this sentence. But linguistic theory also assumes that this structure can change, depending on the sentence. So, for the sentence

(2) The stone broke the vase,

there is a different thematic structure, since *stone* is CAUSE (the one that causes the action) and *vase* is PATIENT. The difference between (1) and (2) is that although the same verb is employed *(break),* no AGENT or INSTRUMENT is expressed in (2); thus, the thematic structure for (2) - [CAUSE, PATIENT] - is different from the thematic structure for (1).

The theoretical approach to thematic roles in linguistics is symbolic. As in predicate logic, the linguistic expressions are decomposed into a central predicate (often the verb) and a number of arguments that complete its meaning [Raposo, 1992]. The predicate assigns thematic roles to the arguments, and each sentence has a *thematic grid,* i. e., a structure with all thematic roles assigned to the sentence arguments by the predicate.

A Natural Language Processing system, called HTRP (which stands for *Hybrid Thematic Role Processor),* is proposed to identify the thematic grid of a semantically sound input sentence. Two versions are deployed: the first, *without* initial knowledge, and the second, *with* initial knowledge. The first version specifies an ordinary connectionist architecture with initial random connection weights, henceforth called RIW *(random initial weight version).* In the second version, henceforth called BIW *(biased initial weight version),* a set of biased initial network connection weights is introduced to represent symbolic rules for ten thematic roles. In both versions, after supervised training, a set of final symbolic rules is extracted, which is consistently correlated to linguistic (symbolic) knowledge. In the case of BIW, this amounts to a revision of the initial rules. In RIW, symbolic rules seem to be induced from the connectionist architecture.

2 Thematic Roles

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 [Dowty, 1989]. One could then imagine that the words, which can fill each of the slots for a given thematic grid, share a common semantic core. Assuming this is regular, one could try to capture such regularity (a) by describing each word in terms of its semantic features, and (b) by generalizing over all such descriptions for each thematic slot.

Semantic feature generalization is the hallmark of McClelland and Kawamoto's [1986] pioneering proposal, and of much subsequent work. In a system called CPPro [Rosa, 1997], a connectionist architecture based on an adaptation of McClelland and Kawamoto's [1986] model is proposed. The words are represented by arrays of semantic microfeatures, formed by subsets accounting for aspects of word meaning, like *human* and *non-human,* where only one value in each subset is active. For the verb, these arrays are arranged on the basis of thematic relationships between the verb and the other words of a sentence, thus mapping thematic roles onto semantic features. The aim of CPPro is to explore the idea of microfeature representation in order to build an architecture able to analyze and to learn the correct thematic relationship attributions of the words in a sentence. Its output reflects judgements of semantic acceptability of a sentence.

In HTRP, the output is constituted by the thematic grid of a sentence, composed of up to ten thematic roles: AGENT, EXPERIENCER, CAUSE, PATIENT, THEME, SOURCE, GOAL, BENEFICIARY, VALUE, and INSTRUMENT. For HTRP, some intuitive thematic role definitions are adopted, as follows. AGENT is the argument having the control of the action expressed by the predicate. EXPERIENCER is a participant who does not have the control of an action expressing a psychological state. CAUSE is the argument that initiates the action expressed by the predicate without controlling it. PATIENT is the participant affected directly by the action of the predicate, usually changing state. THEME is the participant affected indirectly by the action of the predicate, without changing state. The other role labels are self-explanatory.

In HTRP, only sentences with up to three arguments are taken care of. Thus, the argument structure [Haegeman, 1991] of the sentences is as follows:

verb;	1	2	3
	argl	arg2	arg3

where *argl, argl* and *arg3* are the arguments of the verb, to which the predicate (the verb) assigns thematic roles. A limited set of verbs is chosen for the present implementation of HTRP: *break, buy, deliver, fear, frighten, give, hit,* and *love.*

2.1 Verb Representation

The representation of the verb in HTRP is strongly based on Franchi and Can§ado [1998]. They use a non-lexicalist representation; that is, the thematic role assignment compositionally depends on the whole sentence. For instance, taking the verb *break,* (5) and (6) are the thematic grids for (3) and (4) respectively:

(3) Mary broke the vase with a hammer.
(4) The stone broke the vase.

(5) [AGENT, PATIENT, INSTRUMENT]
(6) [CAUSE, PATIENT].

To explain the difference, one can resort again to the notion that thematic roles are elements with semantic content. In this case, it seems that sometimes (e.g. in sentence (3)) control of action is required by the verb *break* in relation to *argl,* while no such control is required in sentence (4). Thus, one could say that *control of action* is a feature to be associated with the verb.

The same is true for the verb *frighten,* regarding a different feature: *direct process triggering.*

(7) Mary frightened Paul with a scream.
(8) The tests frightened Paul.

In (7) control of action is part of the game, while in (8) direct process triggering assumes a central role.

Thus, a small set of features can be associated with the verb, in the same manner that nouns are associated with a set of (different) features [Waltz and Pollack, 1985; McClelland and Kawamoto, 1986; Rosa, 1997].

The compositional features associated with the verb change according to the sentence in which the verb is used. So, it is inadequate to say that a specific verb has a single thematic grid, because this will depend on the whole sentence in which the verb occurs. In sum, a non-lexicalist approach is preferable.

3 The Connectionist Architecture

HTRP system uses a connectionist architecture representing eleven independent artificial neural networks, one for each thematic role and one for the error output [Lawrence *et al,* 1999]. The elementary processors are classical perceptron-like units, and each net has 40 input units, 2 hidden units, and one output unit. The input units are responsible for the representation of two words of a sentence, the verb and one of the nouns. Since each HTRP sentence has, at most, three nouns beyond the verb, each sentence works with at most three neural networks, in order to activate a grid of up to three thematic roles. The first hidden unit (V) represents the conjunction of all the verb microfeatures, and the second (N), the conjunction of all the noun microfeatures. The output unit represents the conjunction of these two microfeature sets (see figure 1). The error output, which has also two hidden units and one output unit, differs at the input layer, which in this case has 80 units, because it is unknown which nouns, in conjunction with the verb, activate the error output.

3.1 The Error Output

Lawrence *et aL* [1999] propose a recurrent neural network to classify English sentences as grammatical or un~

grammatical, exhibiting the same [illegible] supplied by linguistic theory. The network is not divided into innate and learned knowledge. Instead, positive and negative examples are used to discriminate between grammatically acceptable and unacceptable sentences.

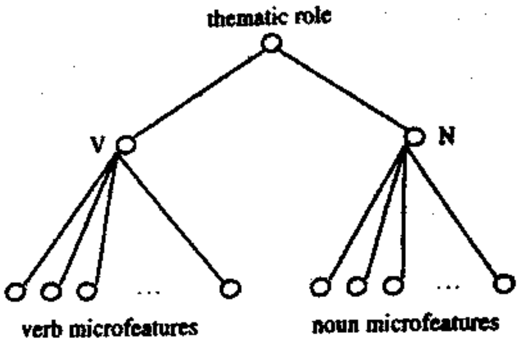

Figure 1. The connectionist architecture for one them

In HTRP, an error output network is implemented, in order to account for this. For a semantically unacceptable input like:

(9) The stone bought the man

the *system* activates the error output. So, before generat-*ng* the thematic grid for a sentence, HTRP tests the semantic acceptability of such sentence, so that the system >nly reveals the thematic grid for semantically well

4 The Hybrid Approach

Since its inception, Artificial Intelligence (AI) is torn *between two opposing fields: the symbolic* paradigm, *based on logic, and the connectionist* paradigm, based on the *propagation of the activity of elementary processors.*

Artificial neural networks do not have the expressive power of general logical representations, since they are not adequate for manipulation of high level symbols [Fodoi and Pylyshyn, 19881 They are usually preferred in a number of situations (such as pattern recognition) because they are able to generalize over the inputs, they are fault tolerant, and *they exhibit the ability to learn from* experience.

But neural networks have a disadvantage: often because of lack of transparency it is hard to understand how they build their inner representations. For instance, it is not easy to ascertain the meaning of the connections and their weights or the configuration of the hidden layers as regards a certain input-output pair.

But, the so-called knowledge-based neural networks, which bring the opposing AI paradigms into closer contact, allow for symbolic knowledge to be introduced in as well as to be extracted from neural networks - that is called *hybrid approach.*

The extraction of symbolic knowledge from trained

between [illegible] sentations and [illegible] knowledge repre- what the actual [illegible] to understand [illegible] Shavlik, 1994]. Additionally, a [illegible] learning time can be obtained by training [illegible] initial knowledge [Omlin and Giles, 1996]. Also, the symbolic knowledge can be input into neural networks and then refined after training.

In the hybrid approach adopted here, the symbolic knowledge is represented through connection weights between neural network processing units. For instance, a *fuzzy* logical rule, with weighted antecedents A and B, and consequent C,

(10) $((w_{AC} * A) + (w_{BC} * B)) \rightarrow C$

can be represented by a connectionist schema, as shown in figure 2. The rule is *fuzzy-like*, because w_{AC} and w_{BC} (connection weights) are not binary values but real numbers. *Also, it* simulates an *and* unit, such that *only the presence of both inputs A and B* causes unit C *to be activated.*

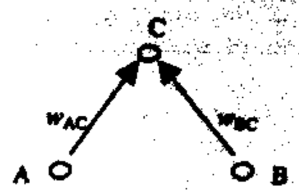

Figure 2. A schema for the rule $((w_{AC} * A) + (w_{BC} * B)) \rightarrow C$.

The symbolic knowledge generated by the *net can be* extracted, in both versions of HTRP, in a way comparable *to initial* symbolic knowledge implementation in BIW, *using the above structure.*

4.1 Microfeatural *Representations*

Word representation in HTRP *is adapted from the semantic microfeature representations used by Waltz* and *Pollack [1985] and McClelland and Kawamoto [19861 for the noun. For the verb, the representation is mainly based on Franchi and Cangado [1998], Nouns and verbs* are accounted for by twenty binary semantic microfeature units each. The following general schema represents the nouns:

- *human - non-human*
- *soft - hard*
- *small - medium - large*
- *1-D/compact - 2-D - 3-D*
- *pointed - rounded*
- *fragile/breakable - unbreakable*
- *value - furniture - food - toy - tool/utensil - animate*

For each of these subsets, only one feature is active, and all the others are inactive. For instance, *man* is *human, soft, large, 3-D, rounded, unbreakable*, and *animate*; *stone* is *non-human, hard, small, 3-D, pointed, unbreakable*, and *tool/utensil*.

The system also includes ambiguous nouns, so that some of its microfeatures are undetermined. In such cases, the system will arrive at the missing values for the intended reading, because it is fault tolerant.

The following schema represents the verbs:

- arg 1 has *cmtrol of action* - on *control*
- *direct process triggering - indirect triggering*
- *direction to source* (argl) - *direction to goal* (arg2)
- *impacting process* for arg2 - *no impacting process*
- *change of state* of arg2 - no *change of state*
- *psychological state* - *no psychological state*
- argl has *objective* - no *objective*
- *effective action* - no *effective action*
- *high intensity* of the process - tow *intensity*
- argl has *interest on process* - no *interest on process*

Again, for each of these subsets, one feature is active, and the other is inactive. For instance, in the sense of sentence (1) above, for *break* the following features are active: *control of action, direct process triggering, direction to goal, impacting process, change of state, no psychological state, objective, effective action, high intensity,* and *interest on process.* In the sense of sentence (2), the following features are active: *no control of action, indirect process triggering, direction to goal, impacting process, change of state, no psychological state, no objective, effective action, high intensity,* and *no interest on process.* As one can see, two different readings for the same verb *break.*

But when the user enters the verb *break* into HTRP, the system does not know which *break* is intended. And, the network input is the "average" of the two readings of *break.* Again, some of the microfeatures will be undetermined. And again, the system will arrive at the missing values for the intended reading of *break.*

4.2 Initial Symbolic Rules

The HTRP "thematic rules" inspired by Haegeman [1991J and McRae *et al.* [1997] for 13 types of verbs (8 different verbs and 5 alternative readings) were also implemented. The rules are *if-then* rules (logical implications), and they are implemented as an *and* gate, i. e., if an input is absent, the unit should not be activated. Unlike classical logic, each element in the antecedent part of the rules is weighted in a fuzzy way, by the connection weight of the respective element in the network. Then, for a unit to be active, all its inputs together should be such that their sum is enough to activate the unit (see figure 2).

For each thematic role there are two 'hidden' rales whose *antecedents map* the u*nits belonging to the input* layer and whose *consequents map hidden units — one for the* verb, and the otter for the noun (see figure *1).* Bar instance, for the thematic role AGENT IN BIW, there is no initial rule for the noun (N), because any noun can in principle be an AGENT. The system, after learning, will decide which nouns could be AGENTS. But for the verb (V), the rule is:

If for verb (0.2 *control of action)* + (0.2 *direct process triggering)* + (0.2 *impacting process)* + (0.2 *objective)* + (0.2 *interest on process)*
Then V
//(0.5 V) + (0.5 N) *then* thematic role = AGENT.

4.3 The Learning Step

The training sentences are generated by a sentence generator, alternating verbs and nouns. Both semantically sound *and* ill-formed sentences are generated. For BIW, learning begins after the introduction of initial symbolic rules as connection weights of the network. The algorithm used is the supervised *backpropagation* [Rumelhart *et al.,* 1986]. After 3,000 training cycles, 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.

One interesting consequence of learning is that the system is able to categorize on the basis of the complementarity of the verb microfeatures for most subsets. Consider the system without initial knowledge (RIW); in this case, the initial connection weights for each subset of microfeatures are random. Since, during training the sentences exhibited mutually exclusive values within each subset of microfeatures, the final connection weights are found to be complementary in the sense that their respective values are of opposite signals. That is, the network incorporates the complementarity of microfeatures in virtue of its architecture and experience.

4.4 Final Rules

Rule extraction consists in reversing the process of initial rule insertion, in BIW. That is, the net weights are assessed and a weighted antecedent is obtained, corresponding to the connection weight. This rule is fuzzy because it allows for weighted antecedents in the production rule. 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 fuzzy rule extraction from the network, after training, for both versions of HTRP is based on Fu [1993], Setiono and Liu [1996], and Towell and Shavlik[1993].

grammatical, exhibiting the same discriminatory power supplied by linguistic theory. The network is not divided into innate and learned knowledge. Instead, positive and negative examples are used to discriminate between grammatically acceptable and unacceptable sentences.

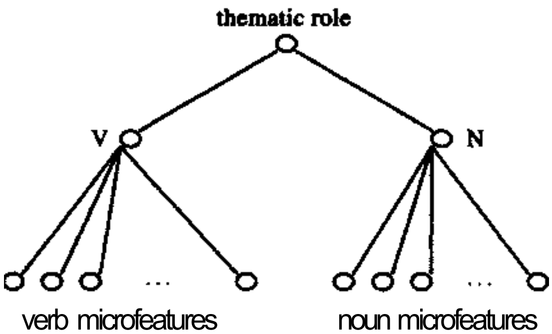

Figure 1. The connectionist architecture for one thematic role.

In HTRP, an error output network is implemented, in order to account for this. For a semantically unacceptable input like:

(9) The stone bought the man

the system activates the error output. So, before generating the thematic grid for a sentence, HTRP tests the semantic acceptability of such sentence, so that the system only reveals the thematic grid for semantically well formed sentences.

4 The Hybrid Approach

Since its inception, Artificial Intelligence (AI) is torn between two opposing fields: the *symbolic* paradigm, based on logic, and the *connectionist* paradigm, based on the propagation of the activity of elementary processors.

Artificial neural networks do not have the expressive power of general logical representations, since they are not adequate for manipulation of high level symbols [Fodor and Pylyshyn, 1988]. They are usually preferred in a number of situations (such as pattern recognition) because they are able to generalize over the inputs, they are fault tolerant, and they exhibit the ability to learn from experience.

But neural networks have a disadvantage: often because of lack of transparency it is hard to understand how they build their inner representations. For instance, it is not easy to ascertain the meaning of the connections and their weights or the configuration of the hidden layers as regards a certain input-output pair.

But, the so-called knowledge-based neural networks, which bring the opposing AI paradigms into closer contact, allow for symbolic knowledge to be introduced in as well as to be extracted from neural networks - that is called *hybrid approach.*

The extraction of symbolic knowledge from trained neural networks permits the exchange of information between connectionist and symbolic knowledge representations and has been of great interest to understand what the neural network actually does [Shavlik, 1994]. Additionally, a significant decrease in learning time can be obtained by training networks with initial knowledge [Omlin and Giles, 1996]. Also, the symbolic knowledge can be input into neural networks and then refined after training.

In the hybrid approach adopted here, the symbolic knowledge is represented through connection weights between neural network processing units. For instance, a *fuzzy* logical rule, with weighted antecedents A and B, and consequent C,

(10) $((w_{AC} * A) + (w_{BC} * B)) \rightarrow C$

can be represented by a connectionist schema, as shown in figure 2. The rule is *fuzzy-like,* because w_{AC} and w_{Bc} (connection weights) are not binary values but real numbers. Also, it simulates an *and* unit, such that only the presence of both inputs A and B causes unit C to be activated.

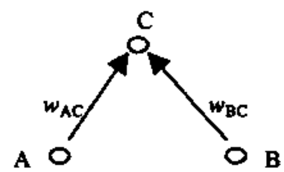

Figure 2. A schema for the rule $((w_{AC} * A) + (w_{BC} * B)) \rightarrow C$.

The symbolic knowledge generated by the net can be extracted, in both versions of HTRP, in a way comparable to initial symbolic knowledge implementation in BIW, using the above structure.

4.1 Microfeatural Representations

Word representation in HTRP is adapted from the semantic microfeature representations used by Waltz and Pollack [1985] and McClelland and Kawamoto [1986], for the noun. For the verb, the representation is mainly based on Franchi and Cancado [1998]. Nouns and verbs are accounted for by twenty binary semantic microfeature units each. The following general schema represents the nouns:

- *human - non-human*
- *soft - hard*
- *small - medium - large*
- *1-D/compact - 2-D - 3-D*
- *pointed - rounded*
- *fragile/breakable - unbreakable*
- *value - furniture - food - toy - tool/utensil - animate*

For each of these subsets, only one feature is active, and all the others are inactive. For instance, *man* is *human, soft, large, 3-D, rounded, unbreakable,* and *animate', stone* is *non-human, hard, small, 3-D, pointed, unbreakable,* and *tool/utensil.*

The system also includes ambiguous nouns, so that some of its microfeatures are undetermined. In such cases, the system will arrive at the missing values for the intended reading, because it is fault tolerant.

The following schema represents the verbs:

- argl has *control of action* - *no control*
- *direct process triggering* - *indirect triggering*
- *direction to source* (argl) - *direction to goal* (arg2)
- *impacting process* for arg2 - *no impacting process*
- *change of state* of arg2 - *no change of state*
- *psychological state* - *no psychological state*
- argl has *objective* - *no objective*
- *effective action* - *no effective action*
- *high intensity* of the process - *low intensity*
- argl has *interest on process* - *no interest on process*

Again, for each of these subsets, one feature is active, and the other is inactive. For instance, in the sense of sentence (1) above, for *break* the following features are active: *control of action, direct process triggering, direction to goal, impacting process, change of state, no psychological state, objective, effective action, high intensity,* and *interest on process.* In the sense of sentence (2), the following features are active: *no control of action, indirect process triggering, direction to goal, impacting process, change of state, no psychological state, no objective, effective action, high intensity,* and *no interest on process.* As one can see, two different readings for the same verb *break.*

But when the user enters the verb *break* into HTRP, the system does not know which *break* is intended. And, the network input is the "average" of the two readings of *break.* Again, some of .the microfeatures will be undetermined. And again, the system will arrive at the missing values for the intended reading of *break.*

4.2 Initial Symbolic Rules

The HTRP "thematic rules" inspired by Haegeman [1991] and McRae *et al.* [1997] for 13 types of verbs (8 different verbs and 5 alternative readings) were also implemented. The rules are *if-then* rules (logical implications), and they are implemented as an *and* gate, i. e., if an input is absent, the unit should not be activated. Unlike classical logic, each element in the antecedent part of the rules is weighted in a fuzzy way, by the connection weight of the respective element in the network. Then, for a unit to be active, all its inputs together should be such that their sum is enough to activate the unit (see figure 2).

For each thematic role there are two 'hidden' rules whose antecedents map the units belonging to the input layer and whose consequents map hidden units - one for the verb, and the other for the noun (see figure 1). For instance, for the thematic role AGENT in BIW, there is no initial rule for the noun (N), because any noun can in principle be an AGENT. The system, after learning, will decide which nouns could be AGENTS. But for the verb (V), the rule is:

If for verb (0.2 *control of action)* + (0.2 *direct process triggering)* + (0.2 *impacting process)* + (0.2 *objective)* + (0.2 *interest on process)*
Then V

//(0.5 V) + (0.5 N) *then* thematic role = AGENT.

4.3 The Learning Step

The training sentences are generated by a sentence generator, alternating verbs and nouns. Both semantically sound *and* ill-formed sentences are generated. For BIW, learning begins after the introduction of initial symbolic rules as connection weights of the network. The algorithm used is the supervised *backpropagation* [Rumelhart *et al.,* 1986]. After 3,000 training cycles, 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.

One interesting consequence of learning is that the system is able to categorize on the basis of the complementarity of the verb microfeatures for most subsets. Consider the system without initial knowledge (RIW); in this case, the initial connection weights for each subset of microfeatures are random. Since, during training the sentences exhibited mutually exclusive values within each subset of microfeatures, the final connection weights are found to be complementary in the sense that their respective values are of opposite signals. That is, the network incorporates the complementarity of microfeatures in virtue of its architecture and experience.

4.4 Final Rules

Rule extraction consists in reversing the process of initial rule insertion, in BIW. That is, the net weights are assessed and a weighted antecedent is obtained, corresponding to the connection weight. This rule is fuzzy because it allows for weighted antecedents in the production rule. 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 fuzzy rule extraction from the network, after training, for both versions of HTRP is based on Fu [1993], Setiono and Liu [1996], and Towell and Shavlik [1993].

For RIW, the final rules for the thematic role AGENT are the following:

'Hidden* rules:

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)*

ThenV

If for noun (1.7 *human)* + (0.2 soft) + (3.1 me-Jiuro + *I.8 large)* + (0.2 3-D) + (0.2 *rounded)* + (1.4 *unbreakable)* + (3.7 *animate)*

ThenN

'Output' rule:

If(-7;3 V) + (6.9 N) *then* thematic role = AGENT.

Notice that almost all antecedents of the 'hidden' rule are negative for the verb. But the antecedent of the 'output' rule (-7.3 V) is also negative for the verb, which means that the negative signals cancel each other out.

Notice also that, for the verb, many microfeatures were highly biased by learning: *control of action, direct process triggering, impacting process, change of state, objective, effective,* and *interest on process.* In relation to the noun rule, the AGENT learned by the net is mainly *medium* and *animate,* and less prominently, *human, large,* and *unbreakable.*

For BIW, that is, with the introduction of initial symbolic rules, for the thematic role AGENT there is the following final rule for the verb:

'Hidden' rule:

If for verb (0.9 *control of action)* + (1.2 *direct process triggering)* + (0.8 *direction to goal)* + (0.5 *impacting process)* + (0.4 *change of state)* + (0.1 *no psychological state)* + (1.2 *objective)* + (-0.1 *effective)* + (0.2 *high intensity)* + (1.2 *interest on process)*

ThenV

As one can see, considering the initial rule antecedents, all features were highly strengthened by learning, with the exception of *impacting process,* which rose only from 0.2 to 0.5. That is, the system can be said to reinforce the initial features.

There is a final rule for the noun too:

'Hidden' rule:

If for noun (-1.6 *human)* + (-0.3 *soft)* + (-2.3 *medium* + -0.8 *large)* + (-0.7 *3-D)* + (-0.6 *rounded)* + (-0.6 *unbreakable)* + (-2.6 *animate)*

ThenN

'Output' rule:

If (7.1 V) + (-7.1 N) *then* thematic role = AGENT.

Since the 'output' rule shows a negative antecedent for the noun (-7.1 N), all the negative weights of the 'hidden' rule antecedents become positive. So, the AGENT learned by the net is mainly *human, medium* and *animate,* and, with less prominence, *soft, targe, 3-D, rounded,* and *unbreakable.*

Notice that there are small differences between the final hidden rules for nouns in RIW and BIW, although one might expect them to be the same because both for BIW and RIW no initial rules for nouns are provided. Such difference stems from (i) the connectionist architecture employed, which takes into account both verb and noun inputs to activate the thematic role output (see figure 1); and (ii) from the backpropagation algorithm, which causes vert) weights to influence noun weights during the error backpropagation step.

To illustrate and compare the differences between RIW and BIW, a summary of the weights for the verb, concerning the thematic role AGENT, is presented in table 3. Recall that these values are used to weigh the microfeatures in the antecedents of the symbolic rules.

Mf	*ca*	*dt*	dg	*im*	*cs*	*np*	*ob*	ef	*hi*	ip
I	0.2	0.2		0.2	-		0.2		-	0.2
FR	0.6	1.0	0.1	0.9	1.1	0.1	2.2	0.6	-0.2	0.8
FB	0.9	1.2	0.8	0.5	0.4	0.1	1.2	-0.1	0.2	1.2

Note: *Mf* = semantic microfeature; *ca = control of action; dt = direct triggering; dg = direction to goal; im = impacting process; cs = change of state; np = no psychological state; ob = objective; = effective; hi = high intensity; ip = interest on process; I* = initial weights; *FR* = final weights for RIW; *FB* = final weights for BIW.

Table 3. A comparison between initial and final weights.

Notice that, when initial knowledge is input to the system (BIW), there is a tendency of strengthening the initial weights. When no initial knowledge is provided (RIW), the final weights are quite close to those in BIW. This can only be taken as evidence that the final weights reflect the available symbolic knowledge (about a thematic role) from the examples and from the architecture, since in this case the initial weights are arbitrary.

5 Conclusion

In the realms of connectionist Natural Language Processing, several systems use the notion of thematic role modeling (e.g., McClelland and Kawamoto [1986], McClelland *et al.* [1989], St. John and McClelland [1990], Jain [1991], and Miikkulainen [1996]). Also, at least one recent paper [Chan and Franklin, 1998] implementing a hybrid system makes use of the notion of case roles, which is close to the concept of thematic relations. The present system departs from all these in that it relies on the role of semantic entailments in thematic relations, i.e., in the way it makes use of theoretical knowledge from linguistics.

HTRP implements a symbolic-connectionist hybrid approach to thematic role processing. In this approach, the advantages of symbolic systems (ease of knowledge representation, understanding through logical inference, etc.) are combined with the advantages of connectionism (learning, generalization, fault tolerance, etc.) to yield a more discriminating thematic role processing, that is sensitive to the subtleties involved in such linguistic phenomenon.

The representation of semantic features adopted in this system would also easily allow for new words to be entered in order to increase its lexicon, once their semantic microfeature arrays are supplied. In HTRP a single network accounts for each verb-noun pair; 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. Another interesting result that should be emphasized regards RIW. Even in a system without initial knowledge, the final rules extracted from the network fully correspond to the symbolic theory that explains them. That is, it seems that the HTRP architecture together with training is enough for the system to arrive at the correct semantic grid of a sentence.

Acknowledgements

We thank Marcio L. de Andrade Netto, Ester M. Scarpa, and Plinio A. Barbosa, of Unicamp, for their extensive comments and inspiration on drafts of this paper. We also thank IJCAI99 reviewers, for their very useful suggestions. Of course, remaining mistakes are our own.

References

[Chan and Franklin, 1998] S. W. K. Chan and J. Franklin. Symbolic Connectionism in Natural Language Disambiguation. *IEEE Transactions on Neural Networks,* Vol. 9, No. 5, 739-755, 1998.

[Dowty, 1989] 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.

[Fodor and Pylyshyn, 1988] J. A. Fodor and Z. W. Pylyshyn. Connectionism and Cognitive Architecture: A Critical Analysis. *Cognition,* Vol. 28, 3-71, 1988.

[Franchi and Cancado, 1998] C. Franchi and M. Cancado. Thematic Hierarchy (in Portuguese). *Technical Report.* Unicamp/USP and UFMG, Brasil, 1998.

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

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

[Jain, 1991] A. N. Jain. Parsing Complex Sentences with Structured Connectionist Networks. *Neural Computation,* 3:110-120,1991.

[Lawrence *et al,* 1999] S. Lawrence, C."L. Giles, and S. Fong. Natural Language Grammatical Inference with Recurrent Neural Networks. *IEEE Trans, on Knowledge and Data Engineering* (accepted for publication), 1999.

[McClelland *et aL,* 1989]. J. L. McClelland, M. F. St. John, and R. Taraban. Sentence Processing: A Parallel Distributed Processing Approach. *Language and Cognitive Processes,* 4 (3/4), 287-335., 1989.

[McClelland and Kawamoto, 1986] 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. L. McClelland and D. E. Rumelhart. A Bradford Book, MIT Press, 1986.

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

[Miikkulainen, 1996]. R. Miikkulainen. Subsymbolic Case-role Analysis of Sentences with Embedded Clauses. *Cognitive Science* 20, 47-73, 1996.

[Omlin and Giles, 1996] C. W. Omlin and C. L. Giles. Rule Revision with Recurrent Neural Networks. *IEEE Trans, on Knowledge and Data Eng.* ,8(1): 183-188, 1996.

[Raposo, 1992] E. P. Raposo. *Teoria da Gramdtica - A Faculdade da Linguagem.* Caminho - colecgao universitaria, Lisboa, Portugal, 1992.

[Rosa, 1997] J. L. G. Rosa. A Thematic Connectionist Approach to Portuguese Language Processing. In *Proceedings of the IASTED International Conference on Artificial Intelligence and Soft Computing.* Banff, Canada, 240-243, July 27-31, 1997.

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

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

[Shavlik, 1994] J. W. Shavlik. Combining Symbolic and Neural Learning. *Machine Learning,* 14(3): 321-331, 1994.

[St. John and McClelland, 1990]. M. F. St. John and J. L. McClelland. Learning and Applying Contextual Constraints in Sentence Comprehension. *Artificial Intelligence, 46:211-25S,* 1990.

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

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