Journal of Universal Computer Science, vol. 16, no. 21 (2010), 3245-3277
submitted: 13/1/09, accepted: 25/11/10, appeared: 1/12/10 © J.UCS

Biologically Plausible Connectionist Prediction of Natural Language Thematic Relations

João Luís Garcia Rosa
(Department of Computer Science
NILC - Interinstitutional Center for Research and Development in Computational Linguistics
University of São Paulo at São Carlos, Brazil
joaoluis@icmc.usp.br)

Juan Manuel Adán-Coello
(Computer Engineering Faculty
Pontifical Catholic University of Campinas, Brazil
juan@puc-campinas.edu.br)

Abstract: In Natural Language Processing (NLP) symbolic systems, several linguistic phenomena, for instance, the thematic role relationships between sentence constituents, such as AGENT, PATIENT, and LOCATION, can be accounted for by the employment of a rule-based grammar. Another approach to NLP concerns the use of the connectionist model, which has the benefits of learning, generalization and fault tolerance, among others. A third option merges the two previous approaches into a hybrid one: a symbolic thematic theory is used to supply the connectionist network with initial knowledge. Inspired on neuroscience, it is proposed a symbolic-connectionist hybrid system called BioθPred (Bio*logically plausible thematic (θ) symbolic-connectionist* Pred*ictor*), designed to reveal the thematic grid assigned to a sentence. Its connectionist architecture comprises, as input, a featural representation of the words (based on the verb/noun WordNet classification and on the classical semantic microfeature representation), and, as output, the thematic grid assigned to the sentence. BioθPred is designed to "predict" thematic (semantic) roles assigned to words in a sentence context, employing biologically inspired training algorithm and architecture, and adopting a psycholinguistic view of thematic theory.
Key Words: thematic (semantic) role labeling, natural language processing, biologically plausible connectionist models
Category: I.2.7, I.2.6, I.5.4, I.2.4

1 Introduction

In sentences such as (1), one can intuitively find an AGENT (*the man*) and a PATIENT (*the girl*). The Government and Binding linguistic theory [Cho81] refers to the roles words usually have in relation to a verb as *thematic roles* (θ-roles), so that one can say that the verb *hit* assigns a thematic grid with the following roles [AGENT, PATIENT] to this sentence. But linguistic theory also assumes that this structure can change, depending on the sentence. So, to sentence (2), the same verb *hit* assigns a different thematic grid, since *car* is CAUSE, instead of

AGENT. The difference between (1) and (2) arises from the semantic properties concerning the subject of verb *hit* - the conscious animate noun *man* in (1) in relation to the inanimate noun *car* in (2); thus, the thematic grid assigned to (2) - [CAUSE, PATIENT] - is different from the grid assigned to (1).

$$\textit{The man hit the girl.} \tag{1}$$

$$\textit{The car hit the tree.} \tag{2}$$

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. The predicate assigns thematic roles to the arguments so each sentence can be associated with a thematic grid.

More than two decades ago, [MK86] proposed a system to deal with relationship patterns. Their system handled those patterns - the words of a sentence - in order to assign the correct case role to its constituents. Since then, many researchers proposed similar approaches.

Here, another Natural Language Processing (NLP) system, called BIOθPRED (which stands for BIO*logically plausible thematic (θ) symbolic-connectionist* PRED*ictor*), is proposed to identify the thematic grid assigned to a semantically sound input sentence. This system, based on linguistic thematic theory, incorporates many features including dealing with lexical and *thematic* ambiguities. BIOθPRED departs from a symbolic knowledge base concerning thematic theory, and after training, shows the thematic grid assigned to a sentence, one thematic role at a time[1]. In addition, a revised symbolic theory can be extracted from the connectionist architecture.

Unlike McClelland and Kawamoto system, in BIOθPRED 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.

BIOθPRED receives the sentence as input, presented in verb-noun pairs. Only meaningfully well-formed sentences belong to its training set. The motivation for this 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. Furthermore, children when learn language, are unlikely to have semantic anomalous sentences as examples. Consequently, BIOθPRED is designed to contain only semantically sound sentences in its training set.

[1] BIOθPRED employs a limited set of thematic roles. The reasons for this choice are presented in the next section.

For the so-called *semantic role labeling* (SRL), according to [PGX10], "for a given verb and given each constituent in a parse, the SRL task is to select from a pre-defined set the constituent's semantic role label with respect to the verb." So, BioθPred does a sort of SRL, although it depends on semantic features of the words, instead of almost exclusively syntactic-based, employed by corpora resources, such as PropBank [PGK05], FrameNet [FJP03], and VerbNet [KKRP08].

The next section presents thematic roles under a psycholinguistic view. Section 3 describes the connectionist representation of the adopted input data, which are distributed and based on a semantic feature set. Section 4 presents a discussion of two opposing standpoints regarding language acquisition, and how this influences the way connectionist systems learn. In section 5, there is an introduction to symbolic-connectionist hybrid systems, their advantages and drawbacks. And finally, in section 6, the proposed system BioθPred is presented, with concluding remarks in section 7.

2 Thematic Roles

The Government and Binding linguistic theory [Cho81, Cho86] states that thematic roles - the semantic relations between words in a sentence - are in the lexicon, so a specific verb assigns a single thematic grid, the structure containing the thematic roles assigned to a sentence. This is a "slot and filler" lexicalist view. For instance, the verb *judge* would assign an experiencer (i) and a theme (j), no matter in which sentence it occurs, like in $[I]_i$ *cannot judge* [*some works of modern art*]$_j$ [Ros07]. There are verbs, however, which assign different thematic grids to different sentences, for instance the verb *thaw* in sentences (3) and (4).

$$\textit{The chef thawed the soup}, \tag{3}$$

$$\textit{The wind thawed the ice cream.} \tag{4}$$

To the sentences (3) and (4), although the same verb is employed, are assigned different thematic grids. In one possible reading of sentence (3), the thematic grid assigned is [agent, patient] and in sentence (4), [cause, patient]. The reason is that *the chef*, in the intended reading of sentence (3), is supposed to have the *control of action*, that is, the intention of thawing. The same does not occur in sentence (4). *The wind* is not willing of thawing anything. A verb that assigns two different thematic grids to different sentences is called here *thematically ambiguous*. How to solve the thematic ambiguity of the verb *thaw* in a lexicalist view?

In BioθPred, the output units constitute the thematic grid assigned to a sentence, which is composed of up to seven thematic roles: AGENT, EXPERIENCER, CAUSE, PATIENT, THEME, LOCATION, and VALUE[2]. For BioθPred, some intuitive thematic role definitions are adopted, as follows. AGENT is the argument that controls the action expressed by the predicate. EXPERIENCER is a participant who does not have the control of an action and usually expresses 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 states. THEME is the participant affected indirectly by the action of the predicate, without changing states. LOCATION represents the place where the event expressed by the verb occurs or is direct to. VALUE is the argument that stresses the importance of something or somebody[3].

A set of verbs from WordNet was chosen for BioθPred. These verbs represent all kinds of semantic relationships BioθPred intends to treat.

In a non-lexicalist view (componential), one could have a representation for thematically ambiguous verbs, like *thaw* in sentences (3) and (4), that would allow them to function as predicates in several sentence types. From the verbs proposed for BioθPred, many are thematically ambiguous verbs, so they can assign more than one thematic grid.

The thematic role notion employed here is what some researchers call *abstract* thematic roles [GJ02].

English lexical resources, like PropBank [PGK05], FrameNet [FJP03], and VerbNet [KKRP08] provide thematic role labels. These labels can be used in annotation tasks. Although the lexical bases differ in the granularity of semantic labels, they are very compatible [Pal09]. Among them, FrameNet is the most fine-grained regarding thematic relationships [Ros10]. Systems that employ these resources use mainly syntactic information on the chosen features for training [PGX10]. Since BioθPred relies mainly on *semantic* representation for the

[2] In the development of the system, several thematic roles, including INSTRUMENT were tried. But the role INSTRUMENT, as well as TIME and others, would be essential in a context where they are necessary to express the use of tools for doing something (INSTRUMENT) or in temporal events (TIME). In the case of BioθPred, these roles are not necessary, because it describes relationships in which these roles do not occur. In other words, BioθPred lexicon does not contemplate verbs which assign such roles. That's the reason they have not been considered anymore, so the final system uses the only seven roles that the employed verbs assign.

[3] How to name a semantic role for *money*? FrameNet employs the role MONEY itself [FB10]. According to VerbNet [Pal10], it should be used ASSET for nouns that specify values (price paid): "Asset: used for the Sum of Money Alternation, present in classes such as Build-26.1, Get-13.5.1, and Obtain-13.5.2 with 'currency' as a selectional restriction." [Pal10]. However, Dowty adopts the usual THEME [Dow90], and Andrews [And85], EXTEND. On the other side, Sowa uses RESOURCE [Sow00]. As one can see, there is no agreement about how to call this specific thematic (semantic) role. So, it is decided to call it VALUE: this way, one could abstract the meaning of the intended role.

words and since it applies to subject-verb-object sentences instead of complex sentences, it is considered irrelevant direct comparisons among such systems.

According to [AS10], "semantic (thematic) roles are a critical aspect of linguistic knowledge because they indicate the relations of the participants in an event to the main predicate." [AS10] propose a computational approach that shows that children and adults use associations between general thematic roles such as AGENT and THEME. In addition, they suggest that semantic roles are verb-based and evolve over time. Also, semantic roles can be learned from the data children are exposed to, through a process of generalisation and categorisation. Regarding cognitive neuroscience, [DIH09] studies how cortical plasticity is employed to drive thematic role assignment in language acquisition. These works show the importance of dealing (computationally also) with semantic roles. This way, innovative approaches to Semantic Role Labeling must take into account semantic aspects of sentence constituents.

3 Microfeatural Distributed Representation

[WP85] and [MK86] employ semantic feature generalization in their pioneering proposals, and much of subsequent works adopt this kind of representation: [Jai91], [JW90], [MJT89], [Mii93], [MD91], [Ros07], [Ros09], [RF99], [RS04], [JM89], [JM90]. These papers present basically the discussion of language processing through a machine learning technique called artificial neural networks. According to them, artificial neural networks are a good approximation of the low-level mental architecture, but not for high-level symbol manipulation processes [FP88]. They also state that subsymbolic systems (i.e., parallel distributed processing approach, or distributed connectionism, or distributed neural network) seem to capture several properties of human-like information processing like learning from examples, context sensitivity, generalization, behavior robustness, and intuitive reasoning. They argue that these properties have been very difficult to model with the traditional symbolic techniques. These papers describe sentence processing models that aim to explain how syntactic, semantic, and thematic constraints are combined in sentence comprehension, and how this knowledge can be coded in the network. The models are based partly on semantic feature encoding of the words. These works are pointed as highly structured connectionist architectures to parse complex syntactic structure sentences. Sometimes, the words are entered one at a time, and the network output consists on local representations for possible assignments of words into phrases, phrases into clauses, and phrases into roles in each clause, and for the possible relationships of the clauses. The consistent activation of the output units represents the sentence interpretation.

3.1 Systems that make use of distributed representations

According to [vG92], a representation is *distributed* when it can be described by means of a function. In a sense, the units in a connectionist network have function roles.

[Mii93] discusses several kinds of representations adopted by many authors. A popular approach for forming distributed representations is the semantic feature encoding, used by [WP85] and [MK86]. In their connectionist systems, inputs are not raw sentences but semantic microfeature representations of the constituent structures of sentences. Words are represented as semantic microfeatures arrays. According to them, every word is described by an array of bits in which each subset holds an associated meaning, like *human-non human*, *soft-hard*, *male-female*, and so on. For nouns and verbs, the features are grouped in several dimensions. Each dimension consists of a mutually exclusive value set and, in general, each word is represented by an array in which one, and only one, value in each dimension is on and all the other values are off for that word.

This kind of representation is meaningful by itself. It is possible to extract information just by examining the representation, without having to train a network to interpret it. In addition, these representations are semantically well constructed; that is, they can be related to a semantic theory.

On the other hand, such patterns must be preencoded, so they remain fixed. Adapting the representations to the actual task and data cannot optimize the performance. Because every one of the concepts must be classified along the same dimensions, the number of dimensions becomes very large, and many of them are irrelevant to the particular concept. It is not an easy task to decide what dimensions are necessary and useful [vG89].

There is also the epistemological question of whether the process of deciding what dimensions to use is justifiable or not. Hand-coded representations are always more or less ad hoc and biased. In some cases, it is possible to make the task trivial by a clever encoding of the input representations [Mii93].

Developing internal representations in hidden layers of a backpropagation network avoids these problems. Hinton's family tree network is a good example [Hin86, Hin90]. That network consists of input, output, and three hidden layers. The input and output layers are localists: exactly one unity is dedicated to each item. The hidden layers next to the input and output layers contain considerably fewer units, which force these layers to form compressed distributed activity patterns for the input and output items. The development of these patterns occurs as an essential part of learning the processing task, and they end up reflecting the regularities of the task.

[Elm89, Elm90] proposes another variant of the same approach. A simple recurrent network is trained to predict the next word in input word sequence. The network hidden layer develops structured representations for the words based on how the words occur in sequences.

In FGREP approach [Mii93], the representations for symbols are developed automatically while the network is learning the processing task. Since the representations are adapted according to the backpropagation error signal, they end up coding the properties of the input elements that are most crucial to the task.

3.2 Distributed representations in BioθPred

In BioθPred, a distributed semantic microfeatural representation is employed, inspired by [WP85] and [MK86] representations. The chosen features are related to a psycholinguistic thematic theory. For verbs, it is based on *WordNet*[4] classification for verbs [Fel90] and on a thematic framework[5]. For nouns, it is based mainly on WordNet classification for nouns [Mil90]. The chosen semantic features for verbs in BioθPred are strongly based on a non-lexicalist representation; that is, the thematic role assignment componentially depends on the whole sentence. For instance, taking the verb *walk*, (7) and (8) are the thematic grids assigned to (5) and (6), respectively.

$$\textit{The man walked to the restaurant,} \tag{5}$$

$$\textit{The man walked the dog.} \tag{6}$$

$$[\textsc{agent}, \textsc{location}], \tag{7}$$

$$[\textsc{agent}, \textsc{theme}]. \tag{8}$$

To explain the difference, one can resort to the notion that thematic roles are elements with semantic content [Dow89]. In this case, it seems that sometimes (e.g. in sentence (5) direct process triggering is required by the verb *walk* in relation to the first argument (*the man*), while no such direct triggering is required in sentence (6)). Thus, one could say that *direct process triggering* is a feature to be associated with the verb.

The same is true for the verb *thaw*, regarding a different feature: *control of action*. In sentence (3) presented previously, *control of action* is part of the game, while in (4) *weather* assumes a central role.

[4] WordNet version 3.0: `http://wordnet.princeton.edu/obtain`

[5] Not all of the features employed are based on WordNet. 10, out of 25, are based on a "thematic framework", that is, they are relevant in a context where semantic role labeling plays a leading role. Is is used an adaptation of McClelland and Kawamoto's representation [MK86] for these features.

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. The componential 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 assigns a single thematic grid, because this will depend on the whole sentence in which the verb occurs. In summary, a non-lexicalist approach is preferable [RF00].

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

Table 1: The twenty five semantic microfeature dimensions for verbs, based on WordNet verb classification and on a thematic framework.

10	*01*
body	*no body*
change	*no change*
cognition	*no cognition*
communication	*no communication*
competition	*no competition*
consumption	*no consumption*
contact	*no contact*
creation	*no creation*
emotion	*no emotion*
motion	*no motion*
perception	*no perception*
possession	*no possession*
social	*no social*
stative	*no stative*
weather	*no weather*
control of action	*no control of action*
direct process triggering	*indirect process triggering*
direction of action to source	*direction of action to goal*
impacting process	*no impacting process*
change of state	*no change of state*
psychological state	*no psychological state*
objective action	*no objective action*
effective action	*no effective action*
high intensity of action	*low intensity of action*
interest on process	*no interest on process*

The schema on table 1 displays the semantic features for verbs. These features are based on WordNet classification for verbs [Fel90] and on a thematic framework [MK86]. For each of these dimensions, one feature is active, and the other is inactive. For instance, table 2 shows the features for the verb *hit* in the senses of sentences (1) and (2).

Table 2: The semantic microfeature dimensions for verb *hit* in the sense of sentence (1) (hit_1) and in the sense of sentence (2) (hit_2).

feature	hit_1	hit_2
body	no	no
change	no	no
cognition	no	no
communication	no	no
competition	yes	no
consumption	no	no
contact	yes	yes
creation	no	no
emotion	no	no
motion	yes	yes
perception	no	no
possession	no	no
social	no	no
stative	no	no
weather	no	no
control of action	yes	no
process triggering	direct	indirect
direction of action	goal	goal
impacting process	yes	yes
change of state	yes	yes
psychological state	no	no
objective action	yes	no
effective action	yes	yes
intensity of action	high	high
interest on process	yes	no

As one can see, two different readings for the same verb. But when the user enters the verb *hit* into BioθPred, the system does not know which *hit* is intended. And, the input pattern makes use of a third value to represent uncertainty, in dimensions where the different readings of the word disagree. It means that in cases in which the two readings agree with the values of an input dimension, this dimension holds the agreed value in the input representation. In cases in which the two readings disagree, the feature displays the value 0.5 in the input representation (represented by the "?" sign). The goal is to verify whether the system can come up with the correct values for such unspecified slots or positions in the input array. Therefore, some of the microfeatures will be undetermined and the system should arrive at the missing values for the intended reading of *hit*.

In addition to the thematic ambiguity of verbs, the system can also handle the problem of lexical ambiguity of nouns. For ambiguous nouns (*bat*, for instance), the input employs a representation similar to that of verbs (values in a three-valued logic).

Table 3 shows the microfeatures for nouns, based on WordNet classification for nouns [Mil90]. Here, there are two-bit groups representing the semantic features, except for *change*, *cognition*, *communication*, *competition*, *emotion*, *motion*, *perception*, *social*, *stative*, and *weather*, which are represented by only one bit.

4 The Starting Large Approach

[RP99] and [Elm93] discuss the importance of starting "small" or starting "large" regarding language acquisition. Elman argues that 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: starting with 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 BIOθPRED. 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 [Elm90], a recurrent artificial 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. A system that accounts for thematic role relationships employing a recurrent network is presented in [Ros02]. In BIOθPRED, a bi-directional architecture is employed, which has memory too, but in a biologically more plausible fashion.

[LGF00] invoke the necessity of having negative examples in the training set for the network to learn the correct grammatical relations between the words in sentences. But this consideration is not plausible from a language acquisition standpoint, since it is not reasonable to suppose that children receive nega-

Table 3: The thirty semantic microfeature dimension groups for nouns, based mainly on WordNet noun classification.

feature	value 1 (*01*)	value 2 (*10/1*)	value 3 (*11*)	value 4 (*00/0*)
action	act	process	state	*n/a*
life	animal	person	plant	*n/a*
element	artifact	quantity	substance	*n/a*
property	attribute	location	possession	*n/a*
corporeal	body	cognition	feeling	*n/a*
society	commun.	event	relation	*n/a*
nature	time	nat. obj.	nat. phen.	*n/a*
miscellaneous	group	motive	shape	*n/a*
size	small	medium	large	*n/a*
consistency	soft	medium	hard	*n/a*
form	rounded	angular	irregular	*n/a*
fragility	breakable	unbreakable	-	*n/a*
instrument	tool	utensil	food	*n/a*
adulthood	prof. adult	adult	child	*n/a*
gender	male	female	-	*n/a*
body	object	subject	-	*n/a*
change	-	subject	-	*n/a*
cognition	-	subject	-	*n/a*
communication	-	subject	-	*n/a*
competition	-	subject	-	*n/a*
consumption	object	subject	-	*n/a*
contact	object	subject	-	*n/a*
creation	object	subject	-	*n/a*
emotion	-	subject	-	*n/a*
motion	-	subject	-	*n/a*
perception	-	subject	-	*n/a*
possession	object	subject	-	*n/a*
social	-	subject	-	*n/a*
stative	-	subject	-	*n/a*
weather	-	subject	-	*n/a*

Legend: *n/a* = not applicable, commun. = communication, nat. obj. = natural object, nat. phen. = natural phenomenon, prof. adult = professional adult.

tive examples, or semantically unsound sentences, when they are listening to their parents. [Elm93] and [RP99] share this position. So, for the sake of cognitive coherence, BioθPred is implemented with no error output, and of course, without semantically unsound sentences in its training set, present in other systems [RF99, RF00].

5 Symbolic-Connectionist Hybrid Systems

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 are not adequate for manipulation of high level symbols [FP88]. 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 exhibit the ability to learn from experience.

But, their critics emphasize that they lack transparency, that is, one does not know how they work, how they develop internal representations. This is a huge drawback. 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. In addition, 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 Symbolic-Connectionist Hybrid Systems, which bring the opposing AI paradigms into closer contact, allowing for symbolic knowledge to be introduced in as well as extracted from neural networks [GBG01]. In these systems one can combine symbolic approach benefits, like expressive power of general logical implications, ease of knowledge representation, and understanding through logical inference, with connectionism advantages already mentioned.

The extraction of symbolic knowledge from trained artificial neural networks permits the exchange of information between connectionist and symbolic knowledge representations and has been of great interest to understand what the artificial neural network actually does [Sha94]. Additionally, a significant decrease in training time can be obtained by training networks with initial knowledge [OG96].

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, had been revised by the connectionist learning. The symbolic knowledge generated by the net can be extracted in a way comparable to initial symbolic knowledge insertion.

Since symbolic-connectionist hybrid systems include initial knowledge, training is supposed to take less time. In the hybrid approach adopted here, the symbolic knowledge is represented through connection weights between artificial neural network processing units. For instance, a connectionist schema, as shown in figure 1, can represent a logical rule (Eq. (9)), with weighed antecedents A and B, and consequent C. The antecedents are weighed, 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.

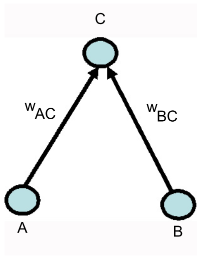

Figure 1: A schema for the rule $(w_{AC} * A)$ and $(w_{BC} * B) \rightarrow C$.

$$(w_{AC} * A) \text{ and } (w_{BC} * B) \rightarrow C. \tag{9}$$

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 computational implementations [HU95].

6 The Biologically Plausible BioθPred System

The search for models (architecture and algorithm) that are more biological plausible attempts to "rescue" the initial neural motivation of connectionist systems. And, of course, computational performance cannot be ignored. It is shown that, in the case of BioθPred, the alleged more biologically motivated system is also more computationally efficient. That is, experiments show that backpropagation was surpassed by BioθPred for a cognitive task, such as semantic role labeling.

In this section, the BioθPred system is presented: its connectionist architecture, the biologically plausible supervised learning algorithm, and simulation experiments: the way training sentences are generated, the inclusion of initial symbolic data, training, and the extraction of symbolic data, which confirm the expected data and extend initial knowledge. Finally, a real example is demonstrated.

6.1 BioθPred connectionist architecture

The connectionist network used in BIOθPRED is structured in three layers: the *input layer* α with `A` units, to which the input sentence is made available, word by word; the *hidden layer* β with `B` units, which allows the network to develop internal representations; and the *output layer* γ with `C` units, from which the assigned thematic grid representations are generated by the system.

The implemented architecture is bi-directional, with a hundred input units ($\mathtt{A} = 100$), fourteen hidden units ($\mathtt{B} = 14$), and seven output units ($\mathtt{C} = 7$), one for each of the seven thematic roles: AGENT (A), PATIENT (P), EXPERIENCER (E), THEME (T), LOCATION (L), CAUSE (C), and VALUE (V). In this case, the architecture classification schema, according to [Sun03], can be *single-module*, employing distributed representation. Each sentence is presented one word at a time to the hundred-unit input layer α (see figure 2). Notice that there are different slots for verbs and nouns. Notice also the bi-directional links between hidden (β) and output (γ) layers, while there are unidirectional links from input (α) to hidden (β) layer.

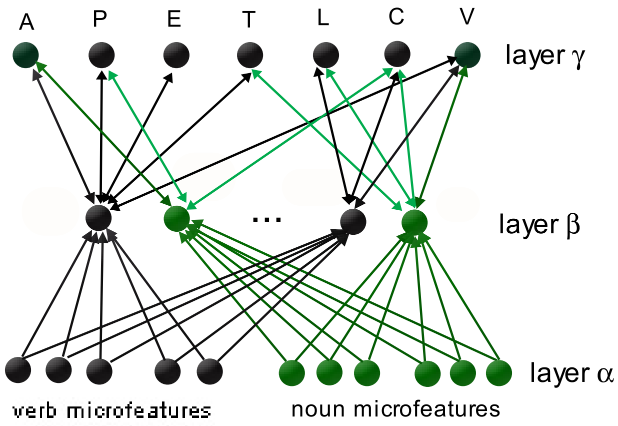

Figure 2: The connectionist architecture of BIOθPRED. Thematic roles in the output layer γ: A = AGENT, P = PATIENT, E = EXPERIENCER, T = THEME, L = LOCATION, C = CAUSE, and V = VALUE.

Since input sentences are pre-parsed [Ros07], sentences other than declarative could be accepted[6]. This paper reports only the connectionist learning with

[6] The reference for pre-parsing is for Portuguese language [Ros07], although this procedure may be considered language independent. Pre-parsing is referred as a prelim-

declarative sentences, but a key-structure with all the nouns compounding the sentence and the main verb is generated. This way, sentences like (10), (11) and (12) would generate the same key-structure (cat-open-door) which predicts the same thematic grid [AGENT, THEME] assigned by verb *open*.

$$The\ cat\ opened\ the\ yellow\ door. \tag{10}$$

$$The\ yellow\ door\ was\ opened\ by\ the\ cat. \tag{11}$$

$$Did\ the\ cat\ open\ the\ yellow\ door? \tag{12}$$

6.2 Biologically plausible supervised learning

For each sentence presented, an output is computed, based on the input pattern and on the current values of net weights. The actual output can be quite different from the "expected" output, i.e. the values that it should have in the correct reading of the sentence, that is, the correct thematic grid assigned to the input sentence. During training, each output is compared to the correct reading, supplied as a "master input." This master input should represent what a real language learner would construct from the context in which the sentence occurs. Learning may be described as the process of changing the connection weights to make the system output correspond, as close as possible, to the master input.

Instead of the computationally successful, but considered to be biologically implausible [Cri89] supervised Back-propagation [RHW86], the learning algorithm BIO*Rec* employed in BIOθPRED is inspired by the Recirculation [HM88] and GeneRec [O'R96] algorithms, and consists of two phases.

In the *expectation* phase[7] (figure 3), when input x, representing the first word of a sentence through semantic microfeatures, is presented to input layer α, there is propagation of these stimuli to the hidden layer β (*bottom-up propagation*) (step 1 in figure 3). There is also a propagation of the previous actual output o^p, which is initially empty, from output layer γ back to the hidden layer β (*top-down propagation*) (steps 2 and 3).[8] Then, a hidden expectation activation (h^e) is generated (Eq. (13)) for each and every one of the B hidden units, based on inputs and previous output stimuli o^p (sum of the bottom-up and top-down propagations - through the sigmoid logistic activation function σ). Then, these hidden signals propagate to the output layer γ (step 4), and an

inary symbolic processing that eliminates the so-called *stop words*, such as determiners, prepositions, and also adjectives. This key format of the (pre-parsed) sentence is used as input to the connectionist network.

[7] [O'R96] employs the terms "minus" and "plus" phases to designate *expectation* and *outcome* phases respectively in his GeneRec algorithm.

[8] The superscript p is used to indicate that this signal refers to the *previous* cycle.

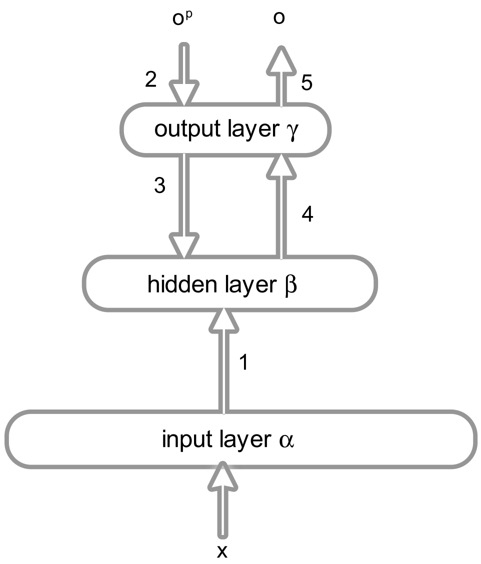

Figure 3: The expectation phase of the BIO*Rec* algorithm.

actual output o is obtained (step 5) for each and every one of the C output units, through the propagation of the hidden expectation activation to the output layer (Eq. (14)) [Ros02]. w^h_{ij} are the connection (synaptic) weights between input (i) and hidden (j) units, and w^o_{jk} are the connection (synaptic) weights between hidden (j) and output (k) units[9].

$$h^e_j = \sigma(\Sigma^{\mathtt{A}}_{i=0} w^h_{ij}.x_i + \Sigma^{\mathtt{C}}_{k=1} w^o_{jk}.o^p_k \qquad 1 \leq j \leq \mathtt{B} \tag{13}$$

$$o_k = \sigma(\Sigma^{\mathtt{B}}_{j=0} w^o_{jk}.h^e_j) \qquad 1 \leq k \leq \mathtt{C} \tag{14}$$

In the *outcome* phase (figure 4), input x is presented to input layer α again; there is propagation to hidden layer β (bottom-up) (step 1 in figure 4). After this, expected output y (step 2) is presented to the output layer and propagated back to the hidden layer β (top-down) (step 3), and a hidden outcome activation

[9] i, j, and k are the indexes for the input (A), hidden (B), and output (C) units respectively. Input (α) and hidden (β) layers have an extra unit (index 0) used for simulating the presence of a bias [Hay99]. This extra unit is absent from the output (γ) layer. That's the reason i and j range from 0 to the number of units in the layer, and k from 1. x_0, h^e_0, and h^o_0 are set to +1. w^h_{0j} is the bias of the hidden neuron j and w^o_{0k} is the bias of the output neuron k.

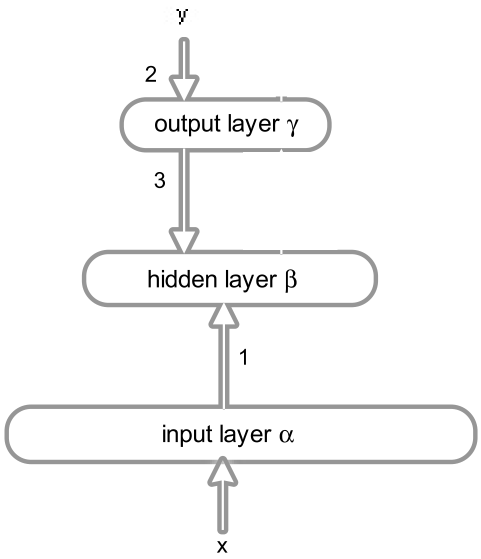

Figure 4: The outcome phase of the BIO*Rec* algorithm.

(h^o) is generated, based on inputs and on expected outputs (Eq. (15)). For the other words, presented one at a time, the same procedure (*expectation* phase first, then *outcome* phase) is repeated [Ros02]. Recall that the architecture is bi-directional, so it is possible for the stimuli to propagate either forwardly or backwardly.

$$h^o_j = \sigma(\Sigma_{i=0}^{\mathtt{A}} w^h_{ij}.x_i + \Sigma_{k=1}^{\mathtt{C}} w^o_{jk}.y_k) \qquad 1 \leq j \leq \mathtt{B} \tag{15}$$

In order to make learning possible the synaptic weights are updated through the delta rule[10] (Eqs. (16) and (17)), considering only the local information made available by the synapse. The learning rate η used in the algorithm is considered an important variable during the experiments [Hay99].

$$\Delta w^o_{jk} = \eta.(y_k - o_k).h^e_j \qquad 0 \leq j \leq \mathtt{B}, \quad 1 \leq k \leq \mathtt{C} \tag{16}$$

$$\Delta w^h_{ij} = \eta.(h^o_j - h^e_j).x_i \qquad 0 \leq i \leq \mathtt{A}, \quad 1 \leq j \leq \mathtt{B} \tag{17}$$

[10] The learning equations are essentially the delta rule (Widrow-Hoff rule), which is basically error correction: “The adjustment made to a synaptic weight of a neuron is proportional to the product of the error signal and the input signal of the synapse in question.” ([Hay99], p. 53).

6.3 Simulation experiments

In this subsection, simulation experiments for BIOθPRED are presented. Training sentences represented by their semantic microfeatures are entered, word by word, to the input layer α (training step). Each sentence is accompanied by its thematic grid since the adopted training algorithm BIO*Rec* is supervised. After training, test sentences are entered in order to check the system learning of the correct thematic grids (recognition step).

During training, the system employs a sentence generator which generates only semantically well-formed sentences. That is, in training, there are no undetermined input values. Every (lexically or thematically) ambiguous word is related to a specific meaning regarding the sentence in which it occurs. After training, when the user enters a thematically ambiguous verb (or a lexically ambiguous noun), the word is simply entered as it is written, that is, without any additional semantic information. So, since at input layer the word comes apart from the sentence, it is unknown which meaning is intended. In this case, some semantic microfeatures have their values "undetermined." For instance, see the word *hit* on table 2, for the semantic microfeature *control of action*, one reading has the value "yes" while the other presents "no." When the user enters *hit*, the value of this feature is unknown, so it is set on 0.5 (neither "yes" nor "no"). The system will arrive to the correct value because it learned sentence patterns with the two readings, so, based on context, it is able to "discover" which is the correct reading (a kind of pattern recognition).

The sentence generator supplies different training sentences, according to semantic and syntactic constraints, absolutely necessary for a system to learn without negative examples [RP99]. Instead of entering the sentences by hand, they are generated automatically by a frame set for each one of the verbs, including alternative readings of the thematically ambiguous verbs. Each frame specifies a verb, a noun set and a list of possible fillers for each noun. So, the sentence frame *the human eats the food* is a generator for sentences in which the subject *human* is replaced by one of the words in the human list, like *man*, and *food* is replaced by one of the words in the food list, like *soup*, since *eat* assigns the following thematic roles: an AGENT (the one that eats) and a PATIENT (the food that is eaten). Then the sentence *the man ate the soup* could be generated. And the output for this sentence would be the assigned thematic grid [AGENT, PATIENT]. See the frames and the assigned thematic grids (outputs) for some verbs on table 4.

The generator replaces the categories present in frames by the words for each category (see examples on table 5), in order to furnish BIOθPRED with the input sentences. Table 6 shows the thematic grids assigned by some verbs in training sentences.

The data used in the experiments are realistic in the way that they reflect

Table 4: Some frames of the sentence generator for training and the thematic grids assigned to them in BioθPred. See table 5.

training frame for *dress*	thematic grid
the human dresses the human	[AGENT,THEME]
training frames for *eat*	thematic grids
the human eats the food	[AGENT,PATIENT]
the animal eats the food	[AGENT,PATIENT]
training frames for *fight*	thematic grids
the human fights the fought	[AGENT,PATIENT]
the human fights the human	[AGENT,PATIENT]
training frames for $have_1$	thematic grids
the human has the value	[EXPERIENCER,VALUE]
the human has the thing	[EXPERIENCER,VALUE]
training frame for $have_2$	thematic grid
the human has the human	[EXPERIENCER,THEME]
training frame for hit_1	thematic grid
the hitter hits the hit	[CAUSE,PATIENT]
training frames for hit_2	thematic grids
the human hits the human	[AGENT,PATIENT]
the human hits the animal	[AGENT,PATIENT]
training frames for *judge*	thematic grids
the human judges the judged	[EXPERIENCER,THEME]
the human judges the human	[EXPERIENCER,THEME]
training frame for *paint*	thematic grid
the human paints the painted	[AGENT,PATIENT]
training frames for *see*	thematic grids
the human sees the seen	[EXPERIENCER,THEME]
the human sees the human	[EXPERIENCER,THEME]
the human sees the animal	[EXPERIENCER,THEME]
training frames for $thaw_1$	thematic grids
the cook thaws the food	[AGENT,PATIENT]
the human thaws the food	[AGENT,PATIENT]
training frame for $thaw_2$	thematic grid
the natural phenomenon thaws the food	[CAUSE,PATIENT]
training frame for $walk_1$	thematic grid
the human walks to the place	[AGENT,LOCATION]
training frame for $walk_2$	thematic grid
the human walks the walked	[AGENT,THEME]

situations found "in the wild." The method used for generating sentences for the training and test sets (i.e. by filling out the slots of sentence frames) creates a representative set of training or test instances, because the chosen frames are typical for the kinds of sentences BIOθPRED intends to deal with.

Table 5: The categories for the frames in the sentence generator and an example of a noun that may fill out the slots for each category (table 4).

category	*noun*
animal	*bat*
cook	*chef*
food	*soup*
fought	*war*
hit	*tree*
hitter	*car*
human	*man*
judged	*competition*
natural phenomenon	*heat*
painted	*door*
place	*house*
seen	*movie*
thing	*ball*
value	*money*
walked	*dog*

If all possible inputs and outputs are shown to a connectionist network employing a supervised training procedure, the net will find a weight set that approximately maps the inputs to the outputs. For many artificial intelligence problems, however, it is impossible to provide all possible inputs. To solve this problem, the training algorithm uses the generalization mechanism, i.e. the network will interpolate when inputs, which have never been received before, are supplied. In the case of this system, since words are described by microfeatures arrays, there are words with related meanings (like, for instance, *woman* and *girl*). These words are expected to contain many microfeatures in common, so the distance between their microfeatures arrays is small, favoring generalization.

When BIOθPRED runs, it is shown a menu through which the user can train the system to learn the correct thematic grids assigned to input sentences. The training set was chosen in order to contain representative verbs and nouns of each thematic category present in BIOθPRED. After training, the system saves the weight matrixes (w^h and w^o) corresponding to the learned knowledge. The

Table 6: Assigned "thematic grids" of BioθPred complete sentences for some verbs during training step.

verb/θ-ROLE	AGE	PAT	EXP	THE	LOC	CAU	VAL
dress	θ			θ			
eat	θ	θ					
fight	θ	θ					
$have_1$			θ				θ
$have_2$			θ	θ			
hit_1		θ				θ	
hit_2	θ	θ					
judge			θ	θ			
paint	θ	θ					
see			θ	θ			
$thaw_1$	θ	θ					
$thaw_2$		θ				θ	
$walk_1$	θ				θ		
$walk_2$	θ			θ			

Legend (thematic role (θ-role)): AGE = AGENT, PAT = PATIENT, EXP = EXPERIENCER, THE = THEME, LOC = LOCATION, CAU = CAUSE, VAL = VALUE.

user can also check the thematic grid assigned to an input sentence. In this case, they are asked to input a sentence manually or to let the system generate it.

The sentences generated automatically by the test generator are different from the sentences generated by the training sentence generator, although their thematic frames are basically the same (the difference relies on the choice of the words involved). In this case, only the default readings for thematically ambiguous verbs are generated, simulating a user entering sentences to be analyzed. The user does not need to know which thematic reading is expected for the verb; BioθPred will decide, based on sentence context, which will be the correct reading and, consequently, arrive at the expected thematic grid assigned to that sentence.

Recall that BioθPred is a symbolic-connectionist hybrid system, so the next three subsections present (1) the cycle of symbolic data insertion and extraction, (2) initial symbolic knowledge insertion, in order to improve training performance, decreasing training time, and (3) the final symbolic data extraction, to revise the symbolic thematic theory.

6.3.1 The cycle of symbolic data in the symbolic-connectionist hybrid BioθPred

The cycle of symbolic data insertion and extraction into/from BioθPred is shown in figure 5, adapted from [TG99]. Instead of beginning with random connection weights (no initial knowledge), BioθPred starts with some biased symbolic knowledge inserted into the connectionist architecture, and, through a learning procedure that makes use of a sentence generator, produces final symbolic data. The sentence generator employs a training set, consisting of syntactic and semantically sound sentences provided by a symbolic theory. Symbolic data, extracted from the connectionist architecture after learning, revise the initial symbolic theory and provide up-to-date information for the sentence generator. It had been proved that the set of rules and the network, from which it is extracted, are equivalent [HU95].

6.3.2 Initial symbolic knowledge insertion and training

After the introduction of initial symbolic knowledge concerning thematic roles as connection weights, the network begins to learn through presentations of semantically sound sentence-thematic grid pairs. The sentence generator produces the input sentences.

Initial knowledge is inserted based on the architecture displayed in figure 6. In this architecture, symbolic concepts about verbs and nouns are mapped on to network elements, according to the hybrid approach as shown in figure 1. As mentioned before, the symbolic knowledge is represented through connection weights between network units. So, logical rules with weighed antecedents can be obtained.

Negative and positive values for nouns and verbs are assigned. In the hidden layer β there are the conjunction of the verb inputs in *HV* and the conjunction of the noun inputs in *HN*. These two units are connected to one unit in the output layer γ, corresponding to a specific thematic role.

Symbolic knowledge regarding a thematic framework is considered in the symbolic-connectionist hybrid system BioθPred. Initial "symbolic thematic rules" for verbs are implemented for each one of the thematic roles (table 7) [Hae91, MFA97]. For nouns, there are initial rules only for the thematic roles agent and cause (table 8), because both may be assigned to subjects of thematically ambiguous verbs (see tables 6 and 4). The other thematic roles are considered noun-independent. 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 weighed 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.

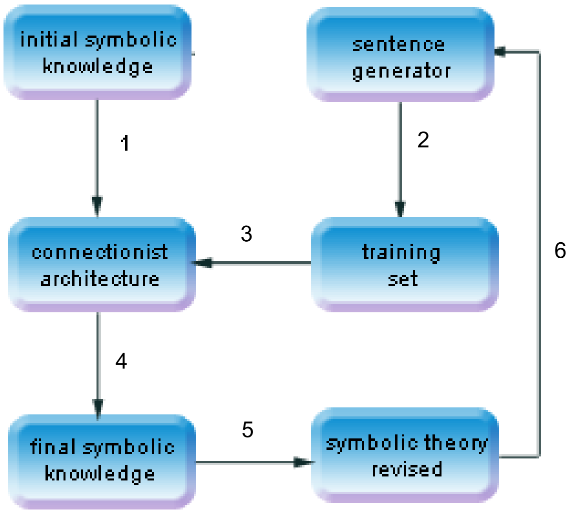

Figure 5: The cycle of symbolic data insertion and extraction into/from a connectionist architecture. Steps: (1) Initial symbolic knowledge is inserted as synaptic weights into the network; (2) The sentence generator supplies semantically sound sentences for training; (3) Training sentences feed the connectionist architecture; (4) After learning is finished, final symbolic knowledge is extracted from the connectionist architecture; (5) Final knowledge is used to revise the initial symbolic theory; (6) The revised symbolic theory provides up-to-date information for the sentence generator.

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 (*HV*), and the other for the noun (*HN*) (see figure 6). For instance, for the thematic role AGENT, the rules are:

If verb comprises (*control of action*) and (*direct process triggering*) and (*impacting process*) and (*objective action*) and (*interest on process*) then HV_{AGENT}

If noun is (*person*) and (*body-subject*) and (*no cognition*) and (*competition*) and (*creation-subject*) then HN_{AGENT}

If (0.5 HV_{AGENT}) and (0.5 HN_{AGENT}) then thematic role = AGENT.

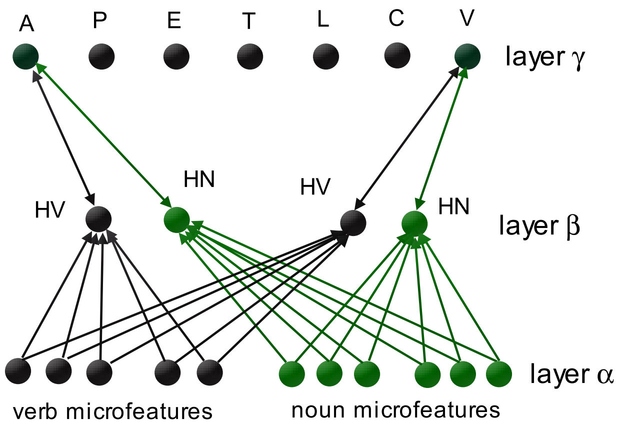

Figure 6: BIOθPRED architecture for initial rule insertion and final rule extraction.

Table 7: Initial symbolic knowledge regarding some verb microfeatures inserted into the network as connection weights.

θ-R/*Mf.*	*ca*	*pt*	*ds*	*ip*	*cn*	*ps*	*ob*	*ea*	*ia*	*in*
AGENT	*yes*	*yes*	-	*yes*	-	-	*yes*	-	-	*yes*
PATIENT	-	-	*no*	*yes*	*yes*	-	-	*yes*	*yes*	-
EXPERIENCER	-	-	*yes*	-	*no*	-	*no*	*no*	*no*	*no*
THEME	-	-	*no*	*no*	*no*	-	-	*yes*	*no*	-
LOCATION	*yes*	*yes*	*no*	-	*no*	-	-	*yes*	-	-
CAUSE	*no*	*no*	-	-	-	-	*no*	-	-	*no*
VALUE	*yes*	*yes*	-	-	*no*	-	-	*yes*	-	*yes*

Legend: Semantic microfeatures (Mf): *ca = control of action, pt = direct process triggering, ds = direction of action to source, ip = impacting process, cn = change of states, ps = psychological state, ob = objective action, ea = effective action, ia = high intensity of action, in = interest on process.*

6.3.3 Final symbolic knowledge extraction

After about 45,000 training cycles, which corresponds to an average output error[11] of 10^{-3}, the system will be able to predict the thematic roles assigned to an input sentence.

[11] The average output error is the difference between "actual" output and "expected" output, and it is obtained from the average squared error energy formula [Hay99] for each set of different sentences presented to the network.

Table 8: Initial symbolic knowledge regarding some noun microfeatures inserted into the network as connection weights.

θ-R/*Mf.*	*li*	*el*	*na*	*bo*	*cg*	*cp*	*ct*	*cr*	*we*
AGENT	*person*	-	-	*subject*	*no*	*yes*	-	*yes*	-
CAUSE	*animal*	*artifact*	*yes*	*object*	-	*no*	*subject*	*no*	*yes*

Legend: Semantic microfeatures (Mf.): *li = life, el = element, na = natural phenomenon, bo = body, cg = cognition, cp = competition, ct = contact, cr = creation, we = weather.*

In relation to accuracy, the system presents precision rate of 94%[12], since only seven words revealed inadequate thematic roles in 120 words belonging to test sentences, in a limited, but sufficient, set of test sentences.

As soon as training finishes, the symbolic rules can be obtained from the architecture by running an extraction procedure [Fu93, SL96, TS93]. Rule extraction consists in reversing the process of initial rule insertion. That is, the net weights are assessed and a weighed 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.

Initial and extracted rules for thematic roles AGENT and CAUSE are presented below. The semantic microfeatures are in *italics* and are weighed by a normalized real number. Only microfeatures representing relevant semantic features for the thematic role (weight above 50%) are displayed. The closer to 1.0, the more precise is the value. In order to understand the conjunction of extracted rules see figure 6:

AGENT

initial hidden rule for verb

If verb comprises (*control of action*) and (*direct process triggering*) and (*impacting process*) and (*objective action*) and (*interest on process*) then HV_{AGENT}

initial hidden rule for noun

If noun is (*person*) and (*body-subject*) and (*no cognition*) and (*competition*) and (*creation-subject*) then HN_{AGENT}

extracted hidden rule for verb:

If verb comprises (1.0 *no change*) and (0.9 *no contact*) and (0.5 *no perception*) (0.6 *no weather*) and (0.9 *control of action*) and (0.9 *direction of action to*

[12] According to [JM00], *precision* is the number of correct answers given by the system divided by the number of answers given by the system.

source) and (0.9 *objective action*) and (0.5 *no effective action*) and (1.0 *interest on process*) then HV_{AGENT}

extracted final rule:
If 1.0 HV_{AGENT} and -0.1 HN_{AGENT} then thematic role = AGENT

CAUSE

initial hidden rule for verb
If verb comprises (*no control of action*) and (*indirect process triggering*) and (*no objective action*) and (*no interest on process*) then HV_{CAUSE}
initial hidden rule for noun
If noun is (*animal*) and (*artifact*) and (*natural phenomenon*) and (*no body-subject*) and (*no competition*) and (*contact-subject*) and (*no creation-subject*) and (*weather*) then HN_{CAUSE}

extracted hidden rule for verb
If verb comprises (0.7 *change*) and (0.6 *contact*) and (0.4 *weather*) and (0.7 *no control of action*) and (0.6 *direction of action to goal*) and (0.7 *no objective action*) and (0.7 *no interest on process*) then HV_{CAUSE}

extracted final rule:
If 1.0 HV_{CAUSE} and 0.5 HN_{CAUSE} then thematic role = CAUSE

The schema above shows initial and extracted rules for two thematic roles (AGENT and CAUSE). For instance, taking the rules for AGENT, one can notice that all the microfeatures considered relevant as initial knowledge were highlighted by connectionist learning, mainly *control of action*, *objective action*, and *interest on process*. The system discovered new semantic features for the verb that assigns the thematic role AGENT: *no change*, *no contact*, *no perception*, *weather*, *direction of action to source*, and *no effective action*.

The noun to which is assigned the thematic role AGENT is not taken into consideration, since the "output" rule shows no significant value for HN_{AGENT}.

Notice that, when initial knowledge is input to the system (for verbs), there is a tendency of strengthening the initial weights. 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, in cases when initial weights are arbitrary.

One interesting consequence of learning is that the system is able to categorize on the basis of the complementarity of the verb microfeatures for many subsets. Since during training the sentences exhibited mutually exclusive values within some subsets of microfeatures, the final connection weights are found to be complementary in the sense that their respective values are of opposite signs. That is, the network incorporates the complementarity of microfeatures by virtue of its architecture and experience.

6.4 A real example

Consider the connectionist architecture of Figure 2. Recall that initial symbolic knowledge concerning verbs and nouns is inserted into the network as connection weights. Take sentence 18 as an example. This sentence is fed to the input layer α, word by word, through semantic features. First, the word *woman* is entered into the noun slot, with the following semantic microfeatures: *no action*, *life: person*, *no element*, *no property*, *no corporeal*, *no society*, *no nature*, *no miscellaneous*, *size: medium*, *consistency: soft*, *form: angular*, *no fragility*, *no instrument*, *adulthood: adult*, *gender: female*, *body: subject*, *no change*, *cognition: subject*, *communication: subject*, *competition: subject*, *consumption: subject*, *contact: subject/object*, *creation: subject*, *emotion: subject*, *motion: subject*, *perception: subject*, *possession: subject*, *social: subject*, *stative: subject*, and *no weather*.

$$\textit{The woman dressed the girl.} \tag{18}$$

The word *woman* represented by its semantic microfeatures propagates to the hidden layer β - *bottom-up* propagation. Also, an empty output o^p is fed to the output layer γ, and then to the hidden layer β - *top-down* propagation. The sum of these signals ((Eq. (13))) is propagated to the output layer γ, generating an output signal o (Eq. (14)) ("expectation" phase - figure 3).

After this phase, the word *woman* is fed again to the input layer α, propagating to the hidden layer β - *bottom-up* propagation. A target output y[13] is propagated from the output layer γ to the hidden layer β - *top-down* propagation, generating a signal (Eq. (15)) ("outcome" phase - figure 4). This way, the connection weights could be updated (Eqs. (16) and (17)).

Only when the verb *dress* is fed to the input layer α, the thematic role assigned to the subject *woman* could be predicted (in this case, AGENT). The verb repeats the cycle of the noun, except for its features: *body*, *no change*, *no cognition*, *no communication*, *no competition*, *no consumption*, *no contact*, *no creation*, *no emotion*, *no motion*, *no perception*, *no possession*, *no social*, *no stative*, *no weather*, *control of action*, *direct process tiggering*, *direction to source*,

[13] In this case, the first word of a sentence, no target output will be provided ($y = 0$) because the system does not know the verb yet. Only after the verb input, there will be a correct thematic role presented at the output as signal y (target or desired output).

impacting process, *no change of states*, *no psychological state*, *objective*, *effective action*, *low intensity of action*, and *interest on process*.

Of course, in the beginning of training, the obtained output o is far from the target output y, therefore the error is large. It means that y does not represent yet the thematic role it should represent for the noun *woman*. After the verb, it is time for the object noun *girl*, whose features are: *no action*, *life: person*, *no element*, *no property*, *no corporeal*, *no society*, *no nature*, *no miscellaneous*, *size: medium*, *consistency: soft*, *form: angular*, *no fragility*, *no instrument*, *adulthood: child*, *gender: female*, *body: subject*, *no change*, *cognition: subject*, *communication: subject*, *competition: subject*, *consumption: subject*, *contact: subject/object*, *creation: subject*, *emotion: subject*, *motion: subject*, *perception: subject*, *possession: subject*, *no social*, *stative: subject*, and *no weather*. Note that the features for *girl* are very similar to the features for *woman*, except for *adulthood* and *social*. This means that the system would be able to generalize over similar inputs. After learning, the system will give the thematic role for *girl*, i.e., THEME.

When many sentences are shown to the network, it will be able to reduce the error to a minimum acceptable, so it can be said that the system learned the thematic roles of training sentences. In the recognition step, if the user enters sentence 18, for instance, BioθPred will be able to reveal the thematic role for each word entered: for *woman*, no thematic role appears because the verb is still unknown, for *dress*, the system reveals the thematic role for *woman*, regarding the verb *dress*: AGENT. When finally the user enters the word *girl*, the system will give THEME at the output. In addition, final symbolic rules can be extracted from the network, confirming and updating the initial symbolic knowledge related to the thematic grid of the sentence.

7 Conclusion

BioθPred 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.

In connectionist NLP 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. 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, provided that their semantic microfeature arrays are supplied.

BioθPred adopts pre-specified semantic microfeatures, although its microfeatures are partially based on WordNet, which is considered an ontology based on semantics [O'H04]. In addition, there is a psycholinguistic concern about which features should be considered important in a thematic framework. A next version of this system could contemplate a kind of semantic features learning, like FGREP architecture [Mii93, MD91], which develops distributed representations for its input/output words.

BioθPred presents as a novelty a more biologically plausible architecture and training procedure based on [OM00]. The bi-directional connectionist architecture is designed to account for chemical and electrical synapses that occur in the cerebral cortex, and the training procedure makes use of this architecture resulting in a model based on neuroscience.

In this system, the architecture employed is feed-forward, although bi-directional. A recurrent architecture, in the sense of [Elm90] was also considered, but it proved to be not as efficient as the bi-directional feed-forward version, although it is well known that recurrent architectures are adequate to temporal processing tasks, like NLP.

BioθPred is able to classify and categorize the intended mutually exclusive microfeatures within a sub-array, 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, in addition to initial biasing, induces the connection weights related to pairs of semantic features to be taken as complementary (see initial knowledge of BioθPred on tables 7 and 8). That is, some sort of internal representation of implications has been developed for thematic roles, which are not introduced as inputs to the network.

In order to better determine the effectiveness of BioθPred for thematic role labeling, it would be very helpful to have some comparison to existing algorithms. Comparisons between the employed algorithm and Back-Propagation for similar NLP applications can be found in [Ros04, RS04].

Acknowledgments

João Luís G. Rosa thanks Fapesp - Fundação de Amparo à Pesquisa do Estado de São Paulo, Brazil, for the research support under project number 2008/08245-4, from which this paper originated. Also, the authors would like to thank the anonymous reviewers for their constructive criticism and useful suggestions.

References

[And85] Andrews, A.: The major functions of the noun phrase; In Shopen, T., editor, *Language Typology and Syntactic Description - Clause Structure*, volume 1. Cambridge University Press, 1985.

[AS10] Alishahi, A. and Stevenson, S.: A computational model of learning semantic roles from child-directed language; *Language and Cognitive Processes*, 25(1):50–93, 2010.

[Cho81] Chomsky, N.: *Lectures on Government and Binding: the Pisa Lectures* Holland: Foris Publications, 1981.

[Cho86] Chomsky, N.: *Knowledge of Language: its Nature, Origin, and Use* New York: Praeger Publishers, 1986.

[Cri89] Crick, F. H. C.: The recent excitement about neural networks; *Nature*, 337:129–132, 1989.

[DIH09] Dominey, P. F., Inui, T., and Hoen, M.: Neural network processing of natural language: Ii. towards a unified model of corticostriatal function in learning sentence comprehension and non-linguistic sequencing; *Brain and Language*, 109(2-3):80–92, 2009.

[Dow89] Dowty, D.: On the semantic content of the notion of 'thematic role'; In G. Chierchia, R. T., B. H. Partee, editor, *Properties, types and meaning.* Dordrecht, Kluwer, 1989.

[Dow90] Dowty, D.: Thematic proto-roles, argument selection, and lexical semantic defaults; *Language: Journal of the Linguistic Society of America*, 3(67):547–619, 1990.

[Elm89] Elman, J. L.: Structured representations and connectionist models; In *Proceedings of the 11th Annual Conference of the Cognitive Science Society*, pages 17–23. Hillsdale, NJ: Erlbaum, 1989.

[Elm90] Elman, J. L.: Finding structure in time; *Cognitive Science*, 14:179–211, 1990.

[Elm93] Elman, J. L.: Learning and development in neural networks: the importance of starting small; *Cognition*, 48:71–99, 1993.

[FB10] Fillmore, C. J. and Baker, C. F.: Framenet - frame report: commerce_buy; http://framenet.icsi.berkeley.edu/index.php?option=com_frontpage&Itemid=1, 2010.

[Fel90] Fellbaum, C.: English verbs as a semantic net; *International Journal of Lexicography*, 3(4):278–301, 1990.

[FJP03] Fillmore, C. J., Johnson, C. R., and Petruck, M. R. L.: Background to framenet; *International Journal of Lexicography*, 16:235–250, 2003.

[FP88] Fodor, J. A. and Pylyshyn, Z. W.: Connectionism and cognitive architecture: a critical analysis; *Cognition*, 28:3–71, 1988.

[Fu93] Fu, L. M.: Knowledge-based connectionism for revising domain theories; *IEEE Transactions on Systems, Man, and Cybernetics*, 23(1):173–182, 1993.

[GBG01] Garcez, A. S. D., Broda, K., and Gabbay, D. M.: Symbolic knowledge extraction from trained neural networks: a sound approach; *Artificial Intelligence*, 125:155–207, 2001.

[GJ02] Gildea, D. and Jurafsky, D.: Automatic labeling of semantic roles; *Computational Linguistics*, 28(3):245–288, 2002.

[Hae91] Haegeman, L.: *Introduction to Government and Binding Theory* Blackwell, 1991.

[Hay99] Haykin, S.: *Neural Networks - a Comprehensive Foundation* Prentice Hall, 2 edition, 1999.

[Hin86] Hinton, G. E.: Learning distributed representations of concepts; In *Proceedings of the Eighth Annual Conference of the Cognitive Science Society*, pages 1–12. Hillsdale, NJ: Erlbaum, 1986.

[Hin90] Hinton, G. E.: Mapping part-whole hierarchies into connectionist networks; *Artificial Intelligence*, 46:47–75, 1990.

[HM88] Hinton, G. E. and McClelland, J. L.: Learning representations by recirculation; In Anderson, D. Z., editor, *Neural Information Processing Systems*, pages 358–66. American Institute of Physics, New York, 1988.

[HU95] Honavar, V. and Uhr, L.: Integrating symbol processing systems and connectionist networks; In Goonatilake, S. and Khebbal, S., editors, *Intelligent Hybrid Systems*, pages 177–208. John Wiley and Sons Ltd., 1995.

[Jai91] Jain, A. N.: Parsing complex sentences with structured connectionist networks; *Neural Computation*, 3:110–120, 1991.

[JM89] John, M. F. S. and McClelland, J. L.: Applying contextual constraints in sentence comprehension; In D. S. Touretzky, G. E. H. and Sejnowski, T. J., editors, *Proceedings of the 1988 Connectionist Models Summer School*, pages 338–46, San Mateo, CA, 1989. Morgan Kaufmann.

[JM90] John, M. F. S. and McClelland, J. L.: Learning and applying contextual constraints in sentence comprehension; *Artificial Intelligence*, 46:217–258, 1990.

[JM00] Jurafsky, D. and Martin, J. H.: *Speech and Language Processing - an Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* Prentice Hall, 2000.

[JW90] Jain, A. N. and Waibel, A. H.: Incremental parsing by modular recurrent connectionist networks; In Touretzky, D. S., editor, *Advances in Neural Information Processing Systems*, volume 2, pages 364–71. Morgan Kaufmann, San Mateo, CA, 1990.

[KKRP08] Kipper, K., Kohonen, A., Ryant, N., and Palmer, M.: A large-scale classification of english verbs; *Language Resources and Evaluation Journal*, 42(1):21–40, 2008.

[LGF00] Lawrence, S., Giles, C. L., and Fong, S.: Natural language grammatical inference with recurrent neural networks; *IEEE Transactions on Knowledge and Data Engineering*, 12(1):126–140, 2000.

[MD91] Miikkulainen, R. and Dyer, M. G.: Natural language processing with modular pdp networks and distributed lexicon; *Cognitive Science*, 15:343–399, 1991.

[MFA97] McRae, K., Ferretti, T. R., and Amyote, L.: Thematic roles as verb-specific concepts; *Language and Cognitive Processes*, 12(2/3):137–176, 1997.

[Mii93] Miikkulainen, R.: *Subsymbolic Natural Language Processing - an Integrated Model of Scripts, Lexicon, and Memory* A Bradford Book, The MIT Press, 1993.

[Mil90] Miller, G. A.: Nouns in wordnet: a lexical inheritance system; *International Journal of Lexicography*, 3(4):245–264, 1990.

[MJT89] McClelland, J. L., John, M. F. S., and Taraban, R.: Sentence processing: a parallel distributed processing approach; *Language and Cognitive Processes*, 4(3/4):287–335, 1989.

[MK86] McClelland, J. L. and Kawamoto, A. H.: Mechanisms of sentence processing: Assigning roles to constituents of sentences; In McClelland, J. L. and Rumelhart, D. E., editors, *Parallel Distributed Processing, Volume 2 - Psychological and Biological Models.* A Bradford Book, MIT Press, 1986.

[OG96] Omlin, C. W. and Giles, C. L.: Rule revision with recurrent neural networks; *IEEE Transactions on Knowledge and Data Engineering*, 8(1):183–188, 1996.

[O'H04] O'Hara, T. P.: *Empirical Acquisition of Conceptual Distinctions via Dictionary Definitions* PhD thesis, NMSU CS, 2004.

[OM00] O'Reilly, R. C. and Munakata, Y.: *Computational Explorations in Cognitive Neuroscience - Understanding the Mind by Simulating the Brain* A Bradford Book, The MIT Press, Cambridge, Massachusetts, USA, 2000.

[O'R96] O'Reilly, R. C.: Biologically plausible error-driven learning using local activation differences: the generalized recirculation algorithm; *Neural Computation*, 8(5):895–938, 1996.

[Pal09] Palmer, M.: Semlink: Combining english lexical resources; In *Proceedings of GL2009 - 5th. Intl. Conf. on Generative Approaches to the Lexicon*, pages 9–15, Pisa, Italy, Sept. 17-19, 2009 2009.

[Pal10] Palmer, M.: Verbnet - a class-based verb lexicon; http://verbs.colorado.edu/ mpalmer/projects/verbnet.html, 2010.

[PGK05] Palmer, M., Gildea, D., and Kingsbury, P.: The proposition bank: An annotated corpus of semantic roles; *Computational Linguistics*, 31(1):71–106, 2005.

[PGX10] Palmer, M., Gildea, D., and Xue, N.: *Semantic Role Labeling* Morgan and Claypool Publishers, 2010.

[RF99] Rosa, J. L. G. and Françozo, E.: Hybrid thematic role processor: Symbolic linguistic relations revised by connectionist learning; In *Proceedings of 16th. International Joint Conference on Artificial Intelligence - IJCAI99*, pages 852–7, Stockholm, Sweden, 31 July-6 August 1999. Morgan Kauffman.

[RF00] Rosa, J. L. G. and Françozo, E.: Linguistic relations encoding in a symbolic-connectionist hybrid natural language processor; In Monard, M. C. and Sichman, J. S., editors, *Lecture Notes in Computer Science - Advances in Artificial Intelligence*, volume 1952, pages 259–68. Springer-Verlag Heidelberg, 2000.

[RHW86] Rumelhart, D. E., Hinton, G. E., and Williams, R. J.: Learning internal representations by error propagation; In *Parallel Distributed Processing, Volume 1 - Foundations*. A Bradford Book, MIT Press, 1986.

[Ros02] Rosa, J. L. G.: A biologically inspired connectionist system for natural language processing; In *Proceedings of the 2002 VII Brazilian Symposium on Neural Networks - SBRN 2002*, pages 243–8, Recife, Brazil, November, 11-14 2002. IEEE Computer Society Press.

[Ros04] Rosa, J. L. G.: A biologically motivated and computationally efficient natural language processor; In *Lecture Notes in Computer Science*, volume 2972, pages 390–9. Springer-Verlag Heidelberg, 2004.

[Ros07] Rosa, J. L. G.: A connectionist thematic grid predictor for pre-parsed natural language sentences; In D. Liu, Z. H. H. Z., S. Fei and Sun, C., editors, *Advances in Neural Networks - ISNN2007 - Lecture Notes in Computer Science*, volume 4492, pages 825–834. Springer-Verlag Berlin Heidelberg, 2007.

[Ros09] Rosa, J. L. G.: Symbolic-connectionist processor of natural language semantic relations; In *Proceedings of the IEEE Symposium Series on Computational Intelligence (IEEE SSCI 2009)*, pages 64–71, Music City Hotel, Nashville, TN, USA, March 30 - April 2 2009. IEEE Conference Proceedings.

[Ros10] Rosa, J. L. G.: Biologically plausible connectionist prediction of natural language thematic relations; In *Proceedings of the WCCI 2010 - 2010 IEEE World Congress on Computational Intelligence, IJCNN 2010 - International Joint Conference on Neural Networks*, pages 1127–1134, Centre de Convencions Internacional de Barcelona, Barcelona, Spain, July 18-23 2010. IEEE Conference Proceedings.

[RP99] Rohde, D. L. T. and Plaut, D. C.: Language acquisition in the absence of explicit negative evidence: How important is starting small?; *Cognition*, 72:67–109, 1999.

[RS04] Rosa, J. L. G. and Silva, A. B.: Thematic role assignment through a biologically plausible symbolic-connectionist hybrid system; In *Proceedings of the International Joint Conference on Neural Networks - IJCNN 2004*, pages 1457–62, Budapest, Hungary, July, 25-29 2004.

[Sha94] Shavlik, J. W.: Combining symbolic and neural learning; *Machine Learning*, 14(3):321–331, 1994.

[SL96] Setiono, R. and Liu, H.: Symbolic representation of neural networks; *IEEE Computer*, 29(3):71–77, 1996.

[Sow00] Sowa, J. F.: *Knowledge Representation: Logical, Philosophical, and Computational Foundations* Brooks/Cole Publishing Co., Pacific Grove, CA, 2000.

[Sun03] Sun, R.: Hybrid connectionist/symbolic systems; In *The Handbook of Brain Theory and Neural Networks, Second Edition*, pages 543–7. A Bradford Book, MIT Press, 2003.

[TG99] Taha, I. A. and Ghosh, J.: Symbolic interpretation of artificial neural networks; *IEEE Transactions on Knowledge and Data Engineering*, 11(3):448–463, 1999.

[TS93] Towell, G. and Shavlik, J. W.: Extracting refined rules from knowledge-based neural networks; *Machine Learning*, 13:71–101, 1993.

[vG89] van Gelder, T.: *Distributed Representation* PhD thesis, Department of Philosophy - University of Pittsburgh, Pittsburgh, PA, 1989.

[vG92] van Gelder, T.: Defining 'distributed representation'; *Connection Science*, 4(3/4):175–191, 1992.

[WP85] Waltz, D. L. and Pollack, J. B.: Massively parallel parsing: a strongly interactive model of natural language interpretations; *Cognitive Science*, 9:51–74, 1985.