Title
Strong Semantic Systematicity from Unsupervised Connectionist Learning

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 17(0)

Authors
Hadley, Robert F.
Hayward, Michael B.

Publication Date
1995

Peer reviewed

Strong Semantic Systematicity from Unsupervised Connectionist Learning

Robert F. Hadley
School of Computing Science
Simon Fraser University
Burnaby, B.C., V5A 1S6
hadley@cs.sfu.ca

Michael B. Hayward
MPR Teltech Ltd.
8999 Nelson Way
Burnaby, B.C., V5A 4B5
hayward@mprgate.mpr.ca

Abstract

A network exhibits strong semantic systematicity when, as a result of training, it can assign *appropriate* meaning representations to novel sentences (both simple and embedded) which contain words in syntactic positions they did not occupy during training. Herein we describe a network which displays strong semantic systematicity in response to *unsupervised* training. During training, two-thirds of all nouns are presented only in a single syntactic position (either as grammatical subject or object). Yet, during testing, the network correctly interprets thousands of sentences containing those nouns in novel positions. In addition, the network generalizes to novel levels of embedding. Successful training requires a corpus of about 1000 sentences, and network training is quite rapid.

1. Introduction

Fodor's and Pylyshyn's arguments (1988) to the effect that human thought and language exhibit both compositionality and systematicity are by now widely known. Although connectionists have questioned whether humans display these attributes in the form that F&P describe, most now agree that *in some important sense* humans do exhibit some form of linguistic systematicity.

In 1989–90, a number of connectionists reported results which established that connectionist networks (hereafter, c-nets) could exhibit forms of linguistic generalization, which, *prima facie*, qualify as systematicity. These results were obtained without recourse to mere implementation of "classical" symbolic methods, and so, it appeared that one of F&P's major conclusions was falsified. However, in Hadley, 1992, 1994a, a learning based conception of systematicity was introduced, and various degrees of systematicity were distinguished, ranging from weak syntactic to strong semantic systematicity. Hadley (1994a) examined six different connectionist systems and argued that, in all probability, none of these systems displayed the strong forms of systematicity that humans display. As a consequence, it appeared that a variant of F&P's original challenge stood unscathed.

Recently, however, some researchers claim to have satisfied Hadley's definition of strong systematicity, though not his formulation of semantic systematicity. In one instance (Phillips, 1994), this claim clearly requires qualification, since (as Phillips has acknowledged, personal communication) the system involved cannot process embedded sentences as required by Hadley's definition. In another instance (Christiansen & Chater, 1994), a claim to strong generalization is restricted to a single syntactic context (conjunctive noun phrases). Discussion of this claim, together with those of Niklasson & van Gelder (1994) is given in Hadley, 1994b, where reservations are explored. In any event, none of the work just cited addresses semantic aspects of systematicity and compositionality, although F&P's (1988) presentation of these concepts did seem to involve semantic issues (such as the capacity to understand the *meaning* of novel sentences and the need to banish semantic equivocation in logical inference).

Before proceeding, it will be helpful to characterize strong semantic systematicity, as defined in (Hadley, 1994b). Briefly stated, we may say that a network exhibits strong semantic systematicity when, as a result of training, it can assign *appropriate* meaning representations to novel sentences (both simple and embedded) which contain words in syntactic positions they did not occupy during training. The training set (or corpus) involved should not only refrain from presenting all words in all syntactic positions, but should so refrain for a significant fraction of the training vocabulary. Furthermore, a sentence counts as novel only if it contains a word in a syntactic position (e.g., subject) that it did not occupy *at any level of embedding* during the training phase.

Now, given F&P's emphasis on *understanding* novel sentences, and our contention (Hadley, 1994a) that humans display at least strong *semantic* systematicity, we have sought a connectionist system which clearly displays these properties in the context of a simple recursive language. In the following pages, we describe a system of c-nets which satisfies this requirement. However, we should stress that we do not see our model as a refutation of F&P's basic thesis. Rather, we have sought to forge a genuine synthesis between connectionist methodology and a powerful classical insight, viz., that activating complex semantic representations entails activating their semantic constituents. Significantly, representations within our model do not involve static strings, but emerge from connectionist methods not considered in F&P's 1988 paper (e.g., binding nodes [Smolensky, 1990] and activation decay). Also, although we certainly

S → NP V NP
NP → N | N RC
N → Mary | Jane | Sally | Susan | Vicky | Fran
| Abe | Bill | Carl | Dave | Earl | Fred
V → likes | knows | treats | calls | draws | helps
| races | sees
REL-PRO → who
RC → REL-PRO V NP

Figure 1: The Grammar of L.

would not claim 'cognitive fidelity' for our model, we believe that the present research takes meaningful steps in the direction of cognitive plausibility. Support for this belief will emerge in later sections, but we may summarize several relevant accomplishments here:

(a) The model exhibits strong semantic systematicity. Following training on a recursive grammar, the system successfully processes, with complete accuracy, substantially deeper levels of sentence embedding than occur during training (thus attaining level 4 in Niklasson's and van Gelder's (1994) generalization hierarchy).

(b) During training, two-thirds of all nouns are *not* presented in all legal positions. However, during testing, those nouns are each successfully presented in positions novel to those words.

(c) When embedding is restricted to a maximum depth of one, as occurs during training, over one million sentences are candidates for inclusion in the training corpus. Yet, the network is successfully trained on a subset of about one thousand sentences.

(d) All network learning is unsupervised; forms of Hebbian training are used throughout. It is widely believed that Hebbian learning is probably closer to biological reality than the commonly used method of backpropagation of error.

(e) Once training is complete, the network not only displays strong semantic systematicity (hereafter, semantic-S), but a straightforward explanation of this fact exists. The network's behavior is transparent.

2. Task and Basic Strategy

A system of c-nets (hereafter, simply called the 'c-net' or the 'network') is given the task of attaining semantic-S in the context of learning the semantics of a simple recursive language. Sentence meanings (construed here as propositions) are represented within a special layer that loosely corresponds to a traditional semantic network (cf. Schubert, 1976). Upon completion of training, the entire c-net should produce appropriate meaning representations in response to any sentence taken from the language L, provided embedded sentences do not occur at depths greater than level three. The grammar of L is given in Figure 1.

Corresponding to each sentence of L (where the maximum level of embedding may be as great as three) is a proposition or 'meaning' representable in the semantic network. However, the training corpus consists of a tiny fraction of all sentences whose maximum level of embedding is *one* (i.e., during training, no relative clause contains another relative clause). As each of the training sentences is presented as input, its corresponding proposition is actively represented within the semantic network. Sentences are presented to an input layer one word at a time, and as training progresses, associations are learned between word (or lexical) nodes in the input layer and nodes in the semantic (network) layer. Activation propagated from the input layer is spread within the semantic layer, where the sequencing of thematic role activation must be learned.

3. Architecture and Representation Methods

Before delving into representational details, we wish to stress that nodes and links within our network should not *in any way* be construed as the counterparts of biological objects (e.g., neurons and axons). Rather, following Smolensky (1988), we intend that our nodes, links, and processes should be taken as fairly high-level abstractions. Nodes and links, for example, might in fact correspond to *patterns of activity* whose biological substrates are left entirely open.

3.1 Overall Structure

The network is comprised of an input layer and a semantic (output) layer. The input layer is a linear array of 21 nodes, each node corresponding to a single lexical item (a word). When a word is presented as input, the corresponding lexical node is activated. By contrast, the semantic layer has considerable internal structure, and contains four distinct types of nodes. These are: concept nodes, proposition nodes (pnodes), binding nodes, and thematic site nodes. Pnodes also have internal structure and serve to integrate concept nodes into unified propositions. Lexical nodes in the input layer are fully connected, by means of tunable links, to each concept node and to the *core* of each pnode in the semantic layer. After training, concept nodes and pnodes provide semantic content for the lexical items, and for this reason are called 'semantic nodes'. We recognize, of course, that semantic nodes do not possess *intrinsic* semantic content. Rather, their semantic role presumably arises through their participation in semantic grounding processes (cf. Hadley, 1989).

3.2 The Semantic Layer, Internal Structure

Structure within the semantic layer is provided primarily by links between concept nodes and pnodes. Pnodes are of two types: master pnodes and modifier pnodes (mod-pnodes). Both types have the cognitive role of integrating concepts of objects and actions into a coherent whole. There is a single 'master' pnode, which unites constituents into the main proposition expressed by a complete sentence (see Figure 2).

By contrast, mod-pnodes have a subservient role; they represent propositions which modify particular concept nodes. Mod-pnodes can be 'bound' to concept nodes for specific periods by means of modifier sites (see π in Figure 2). In the current implementation, there are three mod-pnodes.

Pnodes of each type are small networks in their own right, consisting of a core node (connected to lexical nodes in the input layer) and attached site nodes. Each pnode core is connected to its satellite sites by directional links, and vice-versa. Site nodes represent various the-

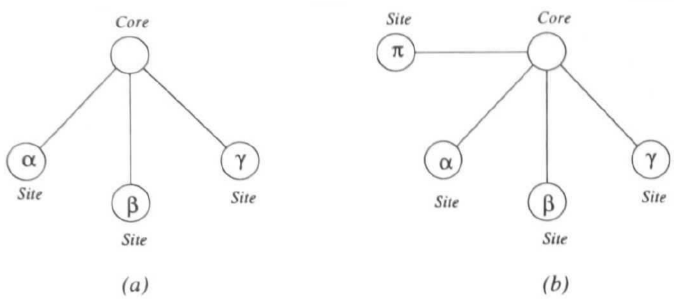

Figure 2: A master pnode (a) and a modifier pnode (b). Site nodes are part of each pnode constellation. The *thematic role* represented by each site is indicated as follows: π indicates a *modifier* site, α represents an *agent* site, β an *action* site, and γ is a *patient* site.

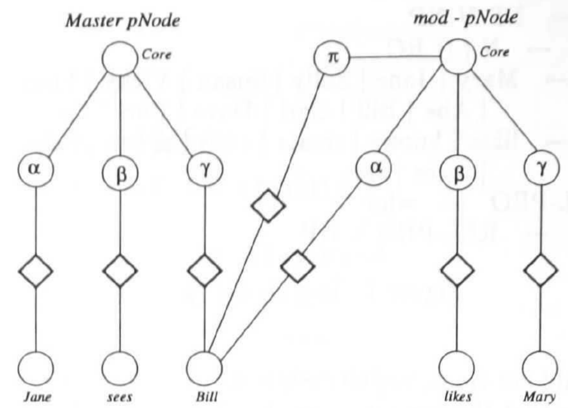

Figure 3: Semantic representation for 'Jane sees Bill who likes Mary'. Diamond shaped nodes depict binding nodes. Only active binding nodes are shown here.

matic roles, including agent, patient, and action roles. In addition, mod-pnodes each possess a 'modifier site'. The thematic role associated with each site is fixed, and site nodes can enter into 'bindings' with a concept node by means of *binding nodes* (see below).

The sites involved in a given pnode form a competitive, winner-take-all cluster. As is customary in c-net simulations, inhibitory links in all winner-take-all (WTA) clusters in this network remain at the virtual level. Links between pnode cores and sites are actual, however, and when a hitherto inactive core receives activation (from below) which exceeds its threshold, the core 'fires' and sends activation to its associated site nodes (for details, see sections 5 and 7). Weights on links from each core to its sites are *tuned* during network training when the core is prompted to fire by an activated site. As will emerge, this tuning enables each pnode constellation to learn the *sequence* in which its sites are stimulated during training. We regard each pnode (which includes both core and sites) as a *module* having the specific cognitive function of learning sequences.

Now, just as pnodes are of two types, so are concept nodes; they represent either actions or objects. Each 'action concept node' is connected to *every* thematic site that represents an action role. Similarly, each 'object concept' is connected to each agent site and each patient site. Every connection between a concept node and a role site is mediated by a 'binding node' (cf. Smolensky, 1990), which intercepts activity from the concept node to the site node (Figure 3).

Links to various binding nodes emanate from a given site. Once a site enters into a binding, both that site and the effective binding node remain active, and the site will not send activation to any binding node except the one to which it is actively bound.[1] For this reason, a site can only bind with a single concept node at a time (via the mediating binding node).

Concept nodes can enter into bindings with particular sites when their mutual *binding node* receives adequate input from both the site and the concept node (see section 7). Binding nodes reside on the connection between each pnode site and each semantically appropriate concept node.

In order to ensure that *appropriate* bindings are formed, binding nodes compete in a WTA fashion (see section 7 for details). This competition is separate from that involving site nodes. However, both *kinds of* competition are essential to the formation of appropriate bindings during sentence interpretation. In this regard, a notable aspect of site competition is that, once a site enters into an active binding, it no longer competes with other sites. Given our assumption that pnodes are abstract objects, possessing a specific cognitive function, we think it plausible that thematic role sites should behave this way. Furthermore, this assumption is consistent with standard connectionist techniques. For, we need only suppose that virtual modifier links (cf. Feldman & Ballard, 1982) emanate from each *site node* to all inhibitory links coming into or leaving that site. When a binding occurs, a binding node fires and sends activation to the involved site (see figure 3). As a result, the site node attains a high activation level (+3) and the site's modifier links then block any activation flow through the site's inhibitory links.

Apart from competitive clusters involving, respectively, binding nodes and site nodes, WTA competition also occurs between semantic nodes. 'Winning' semantic nodes are selected both during network training and testing.

4. Training Data

Training and test data are generated using the grammar of L. One hundred separate training corpora have been generated, and the model has tested successfully when trained on each. A given training corpus consists of 1370 sentences, with any given sentence duplicated at most once. Each training corpus contains 75% simple (N V N) sentences and 25% complex sentences, containing relative clauses. The latter have a maximum embedding depth of one. About half of the complex sentences contain a relative clause in both the subject and object NPs. All 12 nouns appear in each training corpus, but only four of these nouns are permitted to appear in both subject and object positions.

[1] We assume that connections at sites could have evolved to have this specific property.

5. Network Training

Successful network training requires less than one complete pass through a single training corpus. As previously indicated, each sentence is presented to the input layer one word at a time. Presentation of a word amounts to activating one of the 21 lexical nodes in the input layer. The appropriate lexical node is set to +1 just until the next word is presented. Throughout the processing of a complete sentence, the corresponding propositional representation is active within the semantic layer. All concept nodes and pnode cores involved in that proposition are set to their maximum activation levels (+1). Site nodes involved in the active pnodes are also set initially to +1, though their maximum level is +3. In every active propositional representation, the master pnode will be active. One of the three mod-pnodes will be *randomly* selected for inclusion in the representation whenever some proposition modifies a concept (this occurs only when relative clauses are present in the input).

The network's trainable links are those occurring between the input layer and semantic layer, and those emanating from pnode cores to pnode (thematic) sites. Other links within the semantic layer serve to establish bindings and to spread activation. Every lexical input node is connected to each semantic node (i.e., to the concepts and pnode cores). On each 'word iteration', activation flows from an active input node to semantic nodes. Subsequently, activation is relayed from semantic nodes to binding nodes, and ultimately to pnode sites and cores.

5.1 Training Links Between Layers

Each time an input word is presented, a semantic node corresponding to the word's 'meaning' will be active within the semantic layer. This holds true even for the relative pronoun ('who'), which not only denotes some individual each time it is used, but signals that some proposition modifies that individual. Thus, mod-pnodes are strongly correlated with occurrences of 'who'. However, many *spurious semantic nodes* will also be active when a given word appears as input.

A simple Hebbian learning algorithm, which merely strengthened weights between simultaneously active nodes in the input and semantic layers, would suffice to discover the strongest correlations. However, this approach would still assign substantial weights to moderate, though spurious, correlations. A better method, adopted here, is to integrate Hebbian learning with a simple form of competition.

To understand this Hebbian variant, recall that many links flow into each semantic node from the lexical layer. Moreover (as is common), there is a fixed maximum (+1) for the sum of all weights on links coming into a semantic node. Initially, the weight on each of these links is .001.[2] Each time an active input node sends activation to the semantic layer, weights are incremented on every link which connects that node to a semantic node that is active (+1) within the current propositional representation. Let S be any of these active semantic nodes. Then the link (L) coming into S is incremented as follows:

$$increment = R * .0005$$

where R is the ratio of the *current weight on L* to the *current average weight of links into S.* Note that the ratio R can cause links with 'above average' weight to be rewarded significantly. As learning proceeds, the learning on *winning links* accelerates, and little weight is assigned to links which reflect spurious co-occurrences. Moreover, if many input nodes have significant correlation with a single node, as happens with the master pnode, no clear winning link emerges, since the ratio R remains near unity for most links into that semantic node.

Training halts when all semantic nodes have reached their weight maximums, i.e., when no semantic node has any weight left to distribute among its incoming links. Typically, this happens after 950–1050 sentences have been processed. When training is complete, links reflecting *correct* word-concept associations are frequently 100 times greater than those reflecting spurious co-occurrence.

Presently, we shall examine training which occurs within the semantic layer. Before doing so, a final issue is germane to the testing phase, described in section 7. During training, firing thresholds for semantic nodes are irrelevant, since those semantic nodes which are active in a given semantic representation have already been set to their maximum level of activation. However, we assume that, as semantic nodes receive varying levels of input during training, they acquire thresholds. Upon completion of training, the firing threshold of each semantic node is taken to be 80% of the largest input stimulus ever presented to that node. This means that, typically, firing thresholds for semantic nodes will be above .7.

5.2 Training within the Semantic Layer

Within the semantic layer, training occurs at pnodes only. In particular, weights on links flowing from pnode cores to site nodes are tuned as a side effect of spreading activation, which is initiated at concept nodes. In loose analogy with training between layers, we have assumed a fixed limit of *one* for the sum of all core-to-site weights at each pnode. Learning ceases at any pnode which has reached this weight limit.

As previously mentioned, a complete semantic representation remains active throughout the processing of a given sentence. Within this active representation, active concept nodes are *bound* to the particular thematic sites by active binding nodes (see Figure 3). All nodes participating in this representation are set to +1, which is the maximum for concept nodes and pnode cores.

As each word is presented to the input layer, activation is propagated upwards to semantic nodes. Each semantic node receives activation equal to the weight on the link coming into that node from the active input node. Upon receiving this input surge (or 'boost'), *semantic* nodes enter into a WTA competition. That node whose received boost is largest will win the competition. We have assumed that semantic nodes will spend their excess 'boost' when competing, but not their initial level of activation.

Now, training within a pnode constellation occurs as

[2] Random weights close to .001 would serve equally well. For simplicity, we have chosen a uniform initial weight.

a side effect of a concept node's winning the WTA competition. For, when a concept node wins (e.g., 'Jane' in figure 3), it spreads activation towards binding nodes. When this spreading activation reaches an *inactive* binding node, nothing happens because a binding node will fire only when it senses activation from both a site node and a concept node. However, when activation from a concept node reaches *any active* binding node, that node fires and sends activation to the involved site node. Referring to figure 3, if 'Jane' were selected as winner of the semantic competition, activation would reach the agent site on the master pnode. However, if instead 'Bill' were to win, three separate sites would receive activation, including the modifier and agent sites on the mod-pnode. For simplicity, we shall assume that the first word of the sentence, 'Jane', causes the Jane node to fire.

Once a given site (α, in this case) receives activation from a binding node, it will jump from its current activation (+1 in this case) to its maximum level, +3. At this point, the α site is at a higher activation level than the remaining two sites. However, since all sites at the current pnode are involved in active bindings, no WTA competition occurs between the sites (their mutual inhibitory links are currently blocked, as described earlier). Now, given that site α has just been boosted to its highest activation level, α fires towards the pnode core, but retains its +3 activation.[3] Activation received at the core causes the core to fire in turn. As a result, the core sends activation to each of its sites in proportion to the existing weight on the link leading from the core to the given site. The 'boost' now received by each site is added to the site's activation level, unless that site is already at its maximum (+3) level.

Thus far, activation levels for sites α, β, and γ would be, respectively, +3, $(1 + j)$, and $(1 + k)$, where j and k are substantially less than one. Consequently, the agent site, α, is at the highest level by far. Also, now that activation has passed along *trainable* links, from an active core to active sites, weight modification will occur. In particular, weights will be modified by adding $(activation(s) * .0005)$ to the existing weight on the link from core to site s. Clearly, in the present example, the link to site α will receive a significantly larger increment than links to β and γ.

Once weight modification between core and sites is complete, a new cycle begins, for the affected sites have not received sufficient activation to cause them to fire. If we assume that the next input word causes the 'sees' node to win, spreading activation will pass through a binding node, and cause site β to jump to a level of +3. β, in turn, will now excite the core, which will again spread activation to the sites. Weight modification will again occur, this time with α and β both at +3, while the patient site, γ, has substantially less activation. As a consequence, the weight into α receives *another* (comparatively) sizable increment, and though the weight into β will now be substantially augmented, γ's weight receives a comparatively small increase. By the time the third cycle is completed (after processing 'Bill'), weights into agent, action and patient sites will be ordered, by descending magnitude, in the same sequence as those sites were boosted to their maximum (+3) values (other things being equal).

[3]It is somewhat unusual for nodes to both fire and retain significant levels of activation. However, this assumption is consistent with connectionist principles as set forth in Feldman & Ballard (1982).

Due to the nature of conceptual binding, the α and π sites on any mod-pnode will always be bound to the same concept (during training). In such situations, if the concept node is chosen as winner of a WTA competition, both the α and π sites will receive simultaneous activation. Thus, those two sites will always undergo identical training. For this reason, weights into the α and π sites at a given mod-pnode will be identical, though the weights at *separate* pnodes will vary. Because α and π sites are equally weighted with respect to a pnode's core, both sites will tie in any WTA competition between sites which occurs during the testing (i.e., comprehension) phase, described in section 7.

6. Test Data

Once the network has been trained on a given corpus, it is tested on a separate set of over 6000 sentences. Of these 4000 are randomly generated; no restrictions are placed upon any word's syntactic position, aside from grammaticality. Three quarters of these sentences contain embedded clauses, frequently in both subject and object NPs. The remaining 2000+ sentences all present some noun in a position it did not occupy during training (i.e., the word did not occupy that position at any level of embedding during training). Of these (roughly) 2000 sentences, only 500 are simple sentences. In addition, six handcrafted sentences are included in the set. These handcrafted sentences test the network's ability to generalize to deep levels of embedding.

7. Testing Performance

Once the c-net has been fully trained, its 'comprehension' is tested on each of the 6006 sentences described in the preceding section. Testing of a given sentence involves the seriatum presentation of words to the network's input layer. In response to each word, a semantic node is activated within the semantic layer and, via spreading activation, some 'binding' will occur (i.e., a binding node will be activated). Once the sentence has been processed, a pattern of active nodes and bindings exists within the semantic layer. (A node is considered active if its level of activation exceeds its firing threshold.) This pattern of activation is compared to the sentence's *correct* propositional representation, and 'perfect matches' are noted. The process by which thematic role sites are activated and bindings are set is somewhat complex. Given space limitations, we can present only an outline of this process here.

7.1 Sentence Comprehension – Outline.

For each sentence in the training corpus:

(1) All nodes throughout the network are initially set to zero. Cognitively, this might be caused by a preceding silence or by successful comprehension of the previous sentence.

(2) The master pnode is activated, initially via its core. Activation spreads to the three site nodes and a WTA

competition ensues among the sites. A single site (presumably the agent site) wins and attains an activation level of +3. In theory, activation of the master pnode occurs just as some initial stimulus is being recognized *as a word*, perhaps by a 'preprocessor' which activates a particular lexical node once the word is recognized.

(3) A lexical node is activated and activation flows from the lexical layer to the semantic layer. A WTA competition ensues among the semantic nodes. The winning node is selected as follows: If one or more previously inactive nodes surpasses it firing threshold, that node wins which surpassed its threshold by the largest amount. (Ties are extremely unlikely, though a unique winner would be selected in that case.) Otherwise, that presently active node which received the largest input boost wins. In either case, the winner attains an activation of +1.[4]

(4) If the 'winner' from step 3 was a pnode, call it P. P's core is now active (+1). The following occurs:

- A WTA competition ensues among P's site nodes, which are all presently unbound. Multiple (tying) winners are possible.
- A winning site(s) at P is (are) chosen and assumes activation of +3. All other active sites and concept nodes *undergo some decay* (a .01 decrement to their activation level).

(5) Binding now occurs between the most active site (or sites, if there is a tie) and the most active concept node that can bind with that site(s). Activation *decay* ensures that the most recently activated site(s) and concept are the most active nodes.

(6) Binding action from the previous step causes a binding node to fire. This in turn spreads activation to the site involved in that binding, and the site relays activation to the core of the involved pnode. If there are still unbound sites at that pnode, a new WTA competition is triggered among just those sites, and the winning site attains +3 activation. All other active sites and concept nodes undergo some decay. This decay ensures that correct bindings occur during and after recursive embeddings. (See Stevenson, 1994, for a similar use of decay in a massively parallel parser.)

(7) Return to step 3.

8. Test Results

As mentioned in section 4, the network has been separately trained and tested on 100 distinct training corpora. Each of the 100 training sessions produces a uniquely weighted c-net, which, in turn, is tested on the 6006 sentence test corpus described in section 6. As each sentence in each trial is tested, its output is compared to the correct 'target' representation for that sentence. *In all cases*, a perfect match occurred between the network's real and target output. This held true even when test sentences involved maximum levels of embedding (a depth of three consecutive relative clauses).

[4]This 'disjunctive' winner selection rule is consistent with a standard WTA competition among semantic nodes. We merely need to assume that a node which has just surpassed its firing threshold discharges an activation greater than one, i.e., greater than the 'input boost' received by presently active nodes.

9. Discussion

The fundamental goal of research presented here has been to demonstrate that strong semantic systematicity can be achieved through unsupervised connectionist learning. In terms of the definition presented in section 1, we believe this goal has clearly been achieved. Not only is the *trained* c-net able to process sentences containing words in novel positions, but overall network behavior is transparent. Given the forms of Hebbian learning involved, it is clear how words become associated with their conceptual counterparts and how the proper sequence for thematic site activation is learned by p-nodes. These aspects, together with the combinatorial power of *binding nodes*, explain the model's ability to process words occurring in novel syntactic positions. Note that training of p-node links is crucial to the resultant systematicity.

In addition, although our model undoubtedly ignores some concerns for cognitive plausibility, we have sought to attain plausibility wherever possible. For example, (a) we have used unsupervised learning methods throughout, (b) relatively small training corpora have been employed, (c) *most* nouns were not presented in all positions during training, (d) the network generalizes to deep levels of embedding.

References.

Christiansen, M.H. & Chater, N. (1994). Generalization and connectionist language learning. *Mind and Language*, 9, 273-287.

Feldman, J.A. & Ballard, D.H. (1982). Connectionist models and their properties, *Cognitive Science*, 6, 205-254.

Fodor, J.A. & Pylyshyn, Z.W. (1988). Connectionism and cognitive architecture: a critical analysis. *Cognition*, 28, 3-71.

Hadley, R.F. (1989). A default-based theory of procedural semantics. *Cognitive Science*, 13, 107–137.

Hadley, R.F. (1992). Compositionality and systematicity in connectionist language learning, *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*, Bloomington, Indiana, 659-664.

Hadley, R.F. (1994a). Systematicity in connectionist language learning. *Mind and Language*, 9, 247-272.

Hadley, R.F. (1994b). Systematicity revisited: reply to Christiansen and Chater and Niklasson and van Gelder, *Mind and Language*, 9, 431-444.

Niklasson, L.F. & van Gelder, T. (1994). On being systematically connectionist. *Mind and Language*, 9, 288-302.

Phillips, S. (1994). Strong systematicity within connectionism: the tensor-recurrent network, *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*, Atlanta, GA, 723-727.

Schubert, L. (1976). Extending the expressive power of semantic networks, *Artificial Intelligence*, 7, 163-198.

Smolensky, P. (1988). On the proper treatment of connectionsim, *Behavioral and Brain Sciences*, 11, 1-23.

Smolensky, P. (1990). Tensor product variable binding and the representation of symbolic structures in connectionist systems. *Artificial Intelligence*, 46, 159-216.

Stevenson, S. (1994). A unified model of preference and recovery mechansims in human parsing, *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*, Atlanta, GA, 824-829.