Dear Author:

Attached you will find a pdf of the proofs of your article scheduled to appear in a forthcoming issue of *Neural Computation*.

Please print out the pdf of your proof on standard size paper (8 ½ x 11) single-sided, and check for accuracy and consistency, making especially sure to check the accuracy of material that only you can verify (such as numerical data or spelling of proper names). Please also check all figures on the proofs for orientation, legibility, and overall quality and initial your approval next to all figures. Throughout, please limit your changes to those necessary to correct errors or inconsistencies. Major changes to the manuscript or replacement of figures may not be made. Note that changes resulting in repagination of the issue will not be accepted. Please do not send a revision of your article and do not alter the pdf that was sent to you. If you are unable to get the corrections back to us within the requested time frame, we can not guarantee that your changes will be incorporated.

Please mark corrections in red pen directly on the proofs. Please also include a typed list detailing changes. If each figure is approved please indicate this on the list. If you are not approving the figures and new figures have been sent, please indicate why. Please send only one set of corrections on one master copy and one list of changes, even if there is more than one author making changes.

Within three days of receipt of the email notification, please use an overnight delivery service to send your corrected proofs and detailed list to the contact below.

Should you decide to fax or email your corrections, please be sure to provide the following:

- A list containing all changes, including figure approval.
- A copy of every page and figure where corrections are marked.
- If, because of error in placement or reproduction, adjustments to figures are required or new figures are sent, include on the list of changes an explanation of why this is necessary.

Contact Information:
Eric Witz
MIT Press Journals
55 Hayward St.
Cambridge, MA 02142-1315
Email: ewitz@mit.edu
Phone: (617) 258-0586
Fax: (617) 258-6779

Thank you for your cooperation.

Yours sincerely,

Eric Witz
Production Coordinator
MIT Press Journals
Email: ewitz@mit.edu

LETTER Communicated by Thomas Wennekers

A Neurocomputational Approach to Prepositional Phrase Attachment Ambiguity Resolution

Kailash Nadh
k.nadh@mdx.ac.uk
Christian Huyck
c.huyck@mdx.ac.uk
School of Engineering and Information Sciences, Middlesex University, London NW4 4BT, U.K.

A neurocomputational model based on emergent massively overlapping neural cell assemblies (CAs) for resolving prepositional phrase (PP) attachment ambiguity is described. PP attachment ambiguity is a well-studied task in natural language processing and is a case where semantics is used to determine the syntactic structure. A large network of biologically plausible fatiguing leaky integrate-and-fire neurons is trained with semantic hierarchies (obtained from WordNet) on sentences with PP attachment ambiguity extracted from the Penn Treebank corpus. During training, overlapping CAs representing semantic similarities between the component words of the ambiguous sentences emerge and then act as categorizers for novel input. The resulting average resolution accuracy of 84.56% is on par with known machine learning algorithms.

1 Introduction

Ambiguity resolution is an important challenge in natural language processing. This may involve syntactic or semantic ambiguities that make parsing of sentences into symbolic representations difficult. One type of ambiguity is prepositional phrase (PP) attachment ambiguity, which arises when a PP follows a verb phrase (VP) and a noun phrase (NP). Example 1 is a canonical example.

Example 1. *I saw the girl with the telescope*. Here, the phrases *(VP saw)*, *(NP the girl)*, and *(PP with the telescope)* can be combined in two ways. The PP can attach to the NP, yielding the semantic interpretation that *the girl* has *the telescope (VP saw (NP the girl (PP with the telescope)))*, or the PP can attach to the VP, yielding the semantic interpretation that *the telescope* is the instrument of the verb *saw (VP saw (NP the girl) (PP with the telescope))*. It is important to make the correct attachment to get the correct meaning of the sentence. An incorrect attachment decision can lead to a cascade of parsing errors (Lin, 1998).

Humans typically resolve these ambiguities without even noticing them, using their rich semantic knowledge. Example 1 is truly ambiguous because *telescope* is good for seeing and can be easily carried.

Computational systems frequently resolve this ambiguity based on the heads of the phrases—a quadruple (v, $n1$, p, $n2$) where v is the head verb, $n1$ is the head noun of the NP, p is the preposition, and $n2$ is the head noun of the PP (Moens, Calder, Klein, Reape, & Zeevat, 1989).[1] In example 1, the quadruple is *(saw, girl, with, telescope)*, and most PP attachment ambiguities can be resolved with just the quadruple.

Automatic PP attachment ambiguity resolution is usually carried out on quadruples and evaluated on standard data sets, as is the work described in this letter. Previously seen quadruples occurring in a test set can be resolved with relatively good accuracy, but most ambiguities have not been seen before. That is, cases of PP ambiguity occur frequently in text and are usually unique to the reader or even all previously produced English. Sparseness of data makes a one-to-one comparison and classification difficult as real-world data are inherently sparse (Atterer & Schütze, 2007). Despite this, the sparseness of available corpora, for example, one as diverse as the Penn Treebank (PTB; Marcus, Santorini, & Marcinkiewicz, 1993), remains a challenge.

As shown in our previous work (Nadh & Huyck, 2009) a method of measuring semantic similarity between ambiguous quadruples for the purpose of classification can overcome the sparseness problem relatively well to resolve PP attachment ambiguities. In that work, word sense hierarchies were obtained from WordNet (Miller, 1990) for each verb and noun in a quadruple, and the attachment decisions were stored in a lattice. This was a shallow representation of the semantics of the attachment decisions.

This letter describes a neurocomputational approach to PP ambiguity resolution, where the semantic similarity method from our previous work is applied to a neural associative memory model of Hebbian cell assemblies (CAs) (Hebb, 1949; see section 4.1). The neural associative memory encodes the semantics of the attachment decisions, which is then used to make attachment predictions based on how it reacts to novel input.

Using a simple shallow parsing algorithm, we extract examples of PP attachment ambiguity from the PTB corpus in the form of quadruples. Word sense hierarchies for each verb and noun in a quadruple and the words themselves are encoded in a neural network. The corpus of quadruples is split into a training and testing set. The quadruples of the training set are presented to the network along with the correct attachment. Over time,

[1] *Head of a phrase* is standard linguistic terminology that refers to the main item in the phrase or sentence.

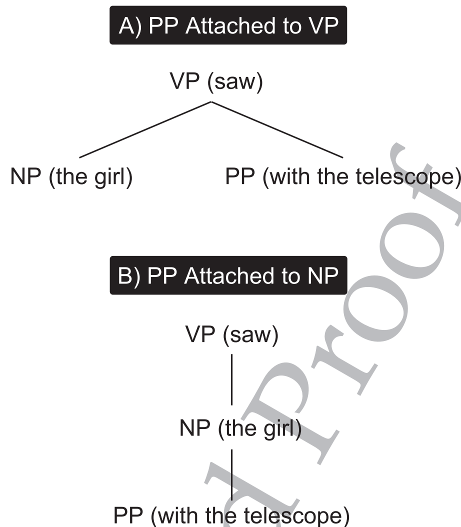

Figure 1: Example of PP attachment ambiguity.

overlapping CAs representing semantic relationships between sense hierarchies of the quadruples emerge via a neurobiologically plausible learning mechanism. The network is then presented with a test set. With the learned CAs acting as categorizers for this novel input, the network achieves an average resolution accuracy of 84.56%, comparable to traditional machine learning models that have been applied to similar data sets (Ratnaparkhi, Reynar, & Roukos, 1994; Nakov & Hearst, 2005). This suggests that biologically plausible CA models are able to perform tasks usually limited to machine learning techniques, highlighting their potential and the importance of further research into them.

The letter is organized as follows. Section 2 details the PP attachment ambiguity, and section 3 discusses prior work on the automatic resolution of this ambiguity. Section 4 discusses CAs, the neural model used in the simulations, and the learning algorithm; section 5 discusses the data used in the simulations; section 6 details the simulations; and section 7 discusses the results and future work.

2 PP Attachment Ambiguity

The canonical example is illustrated in Figure 1, where Figure 1A shows the PP attached to the VP, with the telescope used as the instrument for seeing. Figure 1B shows the case where the PP attaches to the noun. This illustration based on syntax trees is consistent with most current syntactic theories (Jackendoff, 2002).

PP attachments cannot be reliably decided solely based on the sentences in which they are contained, because context (i.e., adjacent sentences) influences decisions. However, subjects can be forced to make a decision on individual sentences, typically called the null context. Most work on automatic resolution, including this letter, is done on the null context, and most take only individual tuples into consideration (Zhao & Lin, 2004; Agirre, Baldwin, & Martinez, 2008; Ratnaparkhi et al., 1994; Stetina & Nagao, 1997; Toutanova, Ng, & Manning, 2004; Nadh & Huyck, 2009). While it can be said with certainty that no system can ever resolve all cases of the ambiguity with only the information from individual sentences, null context resolution systems seem to be a step in the right direction as they continue to produce incrementally better results.

In example 1, 90% of the subjects typically prefer attachment to the verb (Ford, Bresnan, & Kaplan, 1982) in the null context. It is, however, truly ambiguous, and in the null context, many subjects opt for the NP attachment. For many sentences, virtually all subjects select one interpretation when tested in the null context; however, the correct interpretation depends on the sentence. A case similar to example 1, but one that can be disambiguated relatively easily, is the sentence *I saw the girl with the cupcakes*. Since cupcakes are not conventionally instruments for seeing, it makes sense for the PP to attach to the NP meaning *the girl holding the cupcakes* was *seen*.

There is evidence that argumenthood influences the decision (Schutze & Gibson, 1999), suggesting that all things being equal, the preferred interpretation is the PP attaching to the verb as an argument. Early computational systems used heuristics for disambiguation. For instance, right association (Kimball, 1973) suggests that the PP should always attach to the NP, the minimal attachment heuristic (Frazier, 1983) suggests that the PP should always attach to the verb, and the preposition-centric heuristic (Huyck, 2000), which is effective in the case of *of*, always attaches to the NP. Another technique is disambiguation using three-tuples (verb, first noun, preposition; Hindle & Rooth, 1993), as opposed to the widely used four-tuple models such as this work.

3 Related Work

Many systems have been used to resolve PP ambiguity, but the lack of a standard test set makes comparison difficult (see section 7). Ratnaparkhi et al.'s (1994) maximum entropy model used lexical information within verb phrases obtained from the PTB Wall Street Journal (WSJ) corpus with no external semantic knowledge, in contrast to this work, which uses semantics from WordNet. They trained a maximum entropy model and a binary hierarchy of word classes derived by mutual information clustering from the corpus, assigning a probability to either of the possible attachments. Their method yielded an accuracy of 81.6%. The model and binary hierarchy acted as a simple form of semantics.

Stetina and Nagao's (1997) decision tree and semantic dictionary method for word sense disambiguation attained an accuracy of 88.1%. They extracted semantic hierarchies from WordNet for ambiguous quadruples and applied an algorithm to disambiguate the entire training set, which was then converted into a decision tree based on their semantic similarity and attachment decisions. Novel quadruples were applied to the tree, and a decision was based on where they positioned themselves in the tree based on semantic similarity.

Toutanova et al.'s (2004) random walk model, which used a Markov chain model on WordNet synsets, gained an accuracy of 87.5%. Here, the Markov chain, along with the synsets encodes the shallow semantics.

Nakov and Hearst (2005) model queried *n*-grams of ambiguous quadruples against search engines, using the Web as a very lage, implicit training data set. Based on the frequencies of occurrences of different *n*-grams on the Web, the model predicted attachment decisions, achieving an accuracy of 83.82%.

Agirre's et al.'s (2008) sense and parse tree model using semantic classes for parse disambiguation achieved an accuracy of 86.5%. They used two statistical parsers, but instead of supplying them with ambiguous sentences as is, substituted their constituent words with semantic classes extracted from WordNet. Their results showed that incorporating semantic information can enhance the baseline performance that does not incorporate any such information.

In all of these cases, the semantics are shallow and make use of known or derived semantic hierarchies. The semantics of the individual words are not known, but a subsumption hierarchy containing the words is known. Known attachment decisions, based on quadruples, are stored, and novel decisions are based on a combination of the hierarchy and the stored decisions.

Very recently, our own model, which uses a large lattice of hierarchical categories for disambiguation, achieved a resolution accuracy of 88.1%. Quadruples extracted from the PTB WSJ corpus and word sense hierarchies obtained from WordNet were used to construct the lattice. Attachment decisions were made by comparing the semantic similarity of sense hierarchies of unknown quadruples with previously learned examples in the lattice. Semantic similarity of quadruples was measured by counting the number of shared elements in the lattice. For example, the quadruple (*saw, girl, with, telescope*) and the quadruple (*see, boy, with, binocular*) intersected in the lattice at instances such as (*perceive, person, with, instrument*).

The simulations described in this letter use the same quadruples and semantic hierarchies and apply them on a CA-based associative memory model. The CAs take the place of the lattice for measuring semantic similarity.

4 The Neural Network

The simulation described in this letter is modeled with CAs (see section 4.1), constructs that are thought to be the neural basis of concepts. The simulations are based on a neural network with some neurobiologically plausible characteristics. This network has been used for modeling a wide range of tasks (Huyck, 2000, 2001, 2008; Huyck & Nadh, 2009; Byrne & Huyck, 2010; Huyck et al., 2011). The neurons (see Section 4.3) accurately model neural spiking behavior (Huyck, 2011) under a varying input regime. There is extensive evidence that the adjustment of synaptic weights in the brain adheres to a Hebbian learning rule (Abbott & Nelson, 2000), and the simulation described in this paper uses a Hebbian learning rule. The synaptic connections in the model are sparse like those connections in the brain (Churchland & Sejnowski, 1992). The specific gross topology (see section 6.1), however, is unlikely to have a close relation with any particular brain area.

4.1 Hebbian CAs. CA theory postulates that concepts are represented in the brain by the simultaneous activation of large numbers of neurons that have high mutual synaptic strengths (Hebb, 1949). There is extensive evidence that this is indeed the case (Harris, 2005; Maurer, Cowen, Burke, Barnes, & McNaughton, 2003; Pasupathy & Connor, 2002; Funahashi, 2001; Pulvermuller, 1999; Schoenbaum, 1998; Nicolelis, Baccala, Lin, & Chapin, 1995; Foster & Alexander, 1971).

CAs are reverberating circuits of spatially distributed groups of neurons that have high mutual synaptic strength (Wennekers & Palm, 2000). They are learned by a Hebbian learning rule, whereby modifications in the synaptic transmission efficacy are driven by correlations in the firing activity of presynaptic and postsynaptic neurons (Gerstner & Kistler, 2002). That is, if two neurons A and B cofire, where A is the presynaptic neuron transmitting a signal to the postsynaptic neuron B, the strength of the transmission synapses or the mutual synaptic strength increases. This is known as the Hebbian learning rule.

The rule affects A and B such that the likelihood of B responding to signals coming from A is higher in the future. The more two neurons cofire, the higher their mutual synaptic strength grows and the larger the likelihood becomes. Thus, the repeated cofiring of a group of neurons in response to certain stimuli can raise their mutual synaptic strength, making a CA that will respond to similar stimuli in the future, thus forming a neural representation of the stimuli. A CA may become active, or ignite, when a small subset of its neurons fires. The high intra neuronal synaptic strength may cause it to enter a state of reverberation as its neurons undergo cascades of firing, even after the stimulus that triggered it is removed.

If the group of firing neurons happens to belong to more than one CA, activity may spread to other CAs. Prolonged coactivation may cause different

CAs to merge, and learned lateral inhibition may cause certain parts of a CA to win over other parts, eventually disassembling into separate CAs as a result of competitive learning. In Hebbian theory, such processes of formation of CAs account for long-term memory and their reverberating behavior for short-term memory.

Because CAs may share neurons and neurons in different CAs may be connected, CAs may be associated. So they are also the basis of the the fundamental cognitive process of associative memory (Anderson & Bower, 1980).

CAs have been used in various computational models of associative memory (Wennekers & Palm, 2000; Huyck, 2001; Wennekers, 2007; Knoblauch, Kupper, Gewaltig, Körner, & Körner, 2007; Huyck & Nadh, 2009; Nadh & Huyck, 2010). Thus, computational models of Hebbian CAs provide a neurobiologically plausible associative memory mechanism. The system described in this letter uses CAs that emerge from a large fatiguing leaky integrate-and-fire (FLIF) neural network.

4.2 CAs and Associative Memory. CAs can be associated with each other by the same mechanism that drives their formation. Just as repeated cofiring of neurons affects their synaptic strength such that they form a CA, cofiring of neurons belonging to two different CAs may increase their mutual synaptic strength, thus associating the CAs themselves. Either of these two CAs may ignite the other in the future via their synaptic association. Synaptic association may underlie the association of repeatedly co-occurring concepts such as *pen* and *paper*, where the concepts may not share distinctively similar features.

A more complex form of association is possible where CAs representing concepts that have similar features may encode those features with a common subset of neurons that represent them (Osan, Chen, Feng, & Tsien, 2011). For example, features such as *tail* and *fur* shared by *dog* and *cat* may be encoded by a common group of neurons between the CAs representing them. In the brain, many complex cognitive processes are encoded across large, overlapping circuits of neurons (LaBar, Gitelman, Parrish, & Mesulam, 1999; Kable, Lease-Spellmeyer, & Chatterjee, 2002). Overlapped encoding supports dynamics such as generalisation and emergent novelty from the complex interaction of the different subsets of neurons encoding various concepts (Nadh & Huyck, 2010; Nadh, 2010). In addition, overlapped encoding of CAs provides increased capacities of encoding information over a given number of neurons.

4.3 The FLIF Neural Model. The FLIF neuron is an extension of the leaky integrate-and-fire (LIF) neuron (Maass & Bishop, 2001; Gerstner, 2002) that models many features of biological neurons. If a neuron does not fire, it leaks activation. In the FLIF model, the neuron fatigues after firing, resembling biological behavior. The activation A of a neuron i at time t is shown in equation 4.1:

$$A_i(t) = \frac{A_i(t-1)}{\delta} + \sum_{j \in V_i} w_{ji}. \tag{4.1}$$

Activation A is the sum of incoming activation and remnant activation from the previous time step $t-1$, reduced by the decay constant δ, where $\delta > 1$. The incoming activation is the sum of activation of neurons $j \in V_i$, V_i being all presynaptic neurons of i that fired at $t-1$, weighted by the connection from neuron j to i.

When $A > \theta$, where θ is the activation threshold, the neuron fires, losing its activation, $A = 0$. Firing is a binary event, and activation of w_{ji} is sent to all neurons i to which the firing neuron j has a connection.

Fatiguing changes θ; $\theta_{t+1} = \theta_t + F_t$. F_t is positive (F_+) if the neuron fires at t and negative (F_-) otherwise. Firing raises θ, thus reducing a neuron's ability to fire, as it has to integrate more activation. θ decreases with inactivity but is always greater than or equal to the original value.

4.4 Hebbian Learning. A correlatory Hebbian learning rule (Huyck, 2004) drives the learning in the FLIF neural network. Synaptic weights are modified based on the following equations:

$$\Delta^+ w_{ij} = (1 - w_{ij}) * \lambda, \tag{4.2}$$

$$\Delta^- w_{ij} = -(w_{ij} * \lambda). \tag{4.3}$$

w_{ij} is the synaptic weight from neuron i to j and λ is the learning rate. If neurons i and j fire simultaneously, w_{ij} increases based on the Hebbian rule (see equation 4.2). If only neuron i fires, w_{ij} decreases based on the anti-Hebbian rule (see equation 4.3). w_{ij} approximates the likelihood that neuron j fires when neuron i fires.

Neurons in the FLIF network are either excitatory or inhibitory, but never both, where they transmit positive or negative potential respectively. This is an abstraction of the excitatory and inhibitory nature of neurons determined by how their receptors respond to chemical neurotransmitters (Vicario-Abejn, Collin, McKay, & Segal, 1998). The neurons in the network do not have self-connections.

The network may be partitioned into smaller subnets to achieve modularity (see section 6.1). These subnets have a distance-biased topology similar to that of biological neurons (Churchland & Sejnowski, 1992), where excitatory neurons synapse with neurons in their immediate vicinity and neurons in another area in the subnet via a long-distance axon. Because distance is

```
(
  (S
    (NP (PRP I))
    (VP (VBD saw)
      (NP (DT the) (NN girl))
      (PP (IN with)
        (NP (DT the) (NN telescope))))
    (. .)
  )
)
```

Figure 2: Annotated version of the sentence *I saw the girl with the telescope*.

important, the subnet topology is toroidal. Inhibitory and intersubnet connections are not distance biased, but random within the appropriate subnet. Equation 4.4 shows the connectivity rule:

$$\begin{aligned} C_{ij} &= 1 \quad \text{if } r < (1/(d * v)), \\ C_{ij} &= 0 \quad \text{if } r > (1/(d * v)). \end{aligned} \tag{4.4}$$

There exists a connection between neuron i and j in the network only if $C_{ij} = 1$, where r is a random number between 0 and 1, d is the neuronal distance, and v is the connection probability. Distance d ranges from 1 to 4 across all subnets in the simulation, as it has been observed to work well.

5 Data Sets

Attachment decisions are learned from the PTB WSJ corpus. Word hierarchies are extracted from WordNet.

Attachment training and test data are extracted from the PTB using a simple algorithm. The PTB is an English corpus annotated with part of speech (Charniak, 1997) and syntactic structure by lexicographers (Marcus et al., 1993). Each sentence in the corpus is represented in a standard parenthesised tree structure, as shown in Figure 2.

A simple recursive decent parsing algorithm searches the corpus for ambiguous trees of the form *(VP (*) (NP *) (PP-* (*) (NP (*))))*, or *(VP (*) (NP * (PP-* (*) (NP (*)))))*, where * is a wild card representing one or more tagged items. The results are then converted into quadruples (v, $n1$, p, $n2$). For instance, the quadruple extracted from Figure 2 would be *saw girl with telescope*. Some statistics of the data set are presented in Table 1. The

Table 1: Data Set.

Total sentences processed	49,208
Sentences with PP ambiguity	7810
Sentences assigned to the training set	4810
Sentences assigned to the test set	3000
Training set	
Sentences with verb attachments	3847
Sentences with noun attachments	963
Total unique verbs (v), and their word senses	2666
Total unique nouns (n1), and their word senses	4746
Total unique nouns (n2), and their word senses	5706

```
telescope, scope
 => magnifier
  => scientific instrument
   => instrument
    => device
     => instrumentality, instrumentation
      => artifact, artefact
       => whole, unit
        => object, physical object
         => physical entity
          => entity
```

Figure 3: Word sense hierarchy of the noun *telescope*.

training set statistics include counts of the word sense hierarchies obtained from WordNet, and *v1, n1*, and *n2* do not include duplicates.

The word sense hierarchy of a word is a lexical tree made from sequences of hypernyms (superordinate terms) in different levels, where each level is followed by the synset (set of synonyms) of its superordinate term. In semantic net terminology, if x is a hypernym of y, $y - IS\_A - x$; here y might be instantiated by *dog* and x by *mammal*. Figure 3 illustrates a sense hierarchy of the noun *telescope*.

In WordNet, words such as *see* may have many senses and thus may belong to many synsets. In the simulation described in this letter, in cases where multiple senses are present, the first is used and the rest discarded as WordNet orders senses based on the frequency of occurrence in various corpora, with the most common sense being the first (Lee, Lee, & Yun, 2000).

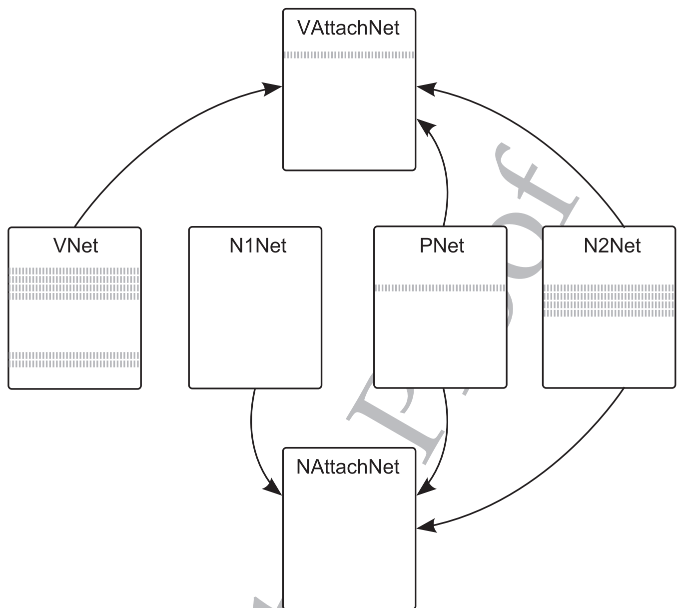

Figure 4: Subnets in the system.

While this may yield false positives in some cases, the probability of the first sense being correct remains high due to the frequency heuristic.

6 The Simulation

A FLIF neural network divided into subnets representing all possible training and test quadruples is used. Word sense hierarchies are encoded on the neural network as overlapping CAs via a Hebbian learning mechanism. Training quadruples are presented along with their attachment decisions. After learning, when presented with a novel quadruple, the attachment decision is derived based on how it activates the network depending on its similarity with the previously learned training quadruples.

6.1 Network Architecture. The system's neural network is partitioned into smaller subnets, as illustrated in Figure 4. *VNet*, *N1Net*, *PNet*, and *N2Net* are the four input subnets that represent each of the four components of a quadruple *v, n1, p, n2*, where *VNet* represents the head verb *v*; *N1Net* represents the head noun *n1* of NP; *PNet* represents the preposition *p*; and *N2Net* represents the head noun *n2* of PP. *VAttachNet* and *NAttachNet*

Table 2: Network Parameters.

	λ Learning	θ Threshold	$F_+ = F_-$ Fatigue	δ Decay	Excitatory Neurons	Number of Neurons
VNet	0.2	4.0	0.6	1.2	80%	53,320
N1Net	0.2	4.0	0.6	1.2	80%	94,920
PNet	0.2	4.0	0.6	1.2	80%	1080
N2Net	0.2	4.0	0.6	1.5	80%	101,520
VAttachNet	0.1	3.5	0.6	1.2	75%	38,470
NAttachNet	0.1	3.5	0.6	1.2	75%	38,470
Total						327,780

encode the verb attachment and noun attachment decisions of the known quadruples presented during training.

Every neuron in *VNet*, *PNet*, and *N2Net* connects to 90 random neurons in *VAttachNet* with an initial synaptic weight of 0.08. Similarly, every neuron in *N1Net, PNet*, and *N2Net* connects to 90 random neurons in *NAttachNet* with an initial synaptic weight of 0.08. The number of connections per neuron was the maximum permitted by computational limitations—primarily, the available memory of the Java heap. Other network parameters were determined by manual parameter exploration. These are presented in Table 2. The parameters were adjusted so as to facilitate slow and gradual learning. Since a large number of quadruples was to be learned, the fatigue rates were set high so that CAs would not be strong enough to continually persist over training cycles and affect the learning of subsequent quadruples.

6.2 Semantic Hierarchies as Overlapping CAs. During initialization, each word in $v, n1, n2$ and its sense words obtained from WordNet from all training quadruples are designated a pattern of 20 neurons in *VNet, N1Net*, and *N2Net*, respectively. Similarly, each unique instance of p is designated a pattern of 20 neurons in *PNet*. For instance, the first 20 neurons in *VNet* belong to the first v, the second 20 to the first word of its sense hierarchy, and so on for every word in its sense hierarchy and every v. This essentially makes the subnet a bag of words made of every v and its sense hierarchy in the training set. Only the first sense hierarchy of a word is considered, as explained in section 5. Also, each subnet has only one instance of a word, even if it is present in multiple quadruples. This serial arrangement does not encode association between words and their sense hierarchies. That is, *telescope* and *magnifier* has no physical overlap in the subnet. However, the associations are encoded in the synapses between the two sets of neurons by gradual learning.

Every training quadruple is presented to the network for 100 cycles each. Neurons belonging to each of the words in v, $n1$, $n2$, and their sense hierarchies, and the *p* are externally stimulated in their corresponding subnets. If a quadruple has a verb attachment, a set of 10 neurons representing it as a whole is externally stimulated in *VAttachNet* and 10 neurons in *NAttachNet* in the case of a noun attachment. This is illustrated in Figure 4, where a hypothetical quadruple with a verb attachment being learned is shown. Each shaded row in the subnets represents exactly one word (20 neurons) being stimulated. *V1Net* and *N2Net* have multiple rows active, representing one word and words in their sense hierarchies. Cofiring of this set of neurons with different word neurons in the three input subnets and the preposition subnet causes their intersubnet synaptic weights to increase, gradually forming CAs representing the quadruple in the appropriate attachment subnet. Similarly, CAs representing words and their sense hierarchies are formed in the three input subnets, and CAs representing different prepositions are formed in the preposition subnet. Word CAs representing word sense hierarchies may have a high degree of overlap with other words based on their semantic similarity. For instance, *boy* and *girl* CAs in *N1Net* share a large percentage of neurons as their sense hierarchies are very similar, for example, (*girl, adult, person, someone . . .*—*boy, adult, person, someone . . .*). These CAs may change as new words with semantic similarities are encountered as learning progresses.

If a quadruple *B* with a verb attachment is similar to a previously learned quadruple *A*, *B*'s inputs to the four input subnets may ignite the *A* CA in *VAttachNet* while *B* is being learned. This may result in the *B* CA overlapping with the already active *A*, forming a CA that encompasses the properties of both quadruples. Thus, similar quadruples have CAs that overlap based on the degree of similarity. This creates groups of overlapping CAs in the subnets that ignite dynamically based on inputs. If a novel quadruple is presented, CAs of quadruples most similar to them will ignite, and the system assumes that they have similar PP attachments. This is the essence of the simulation, where attachment decisions for novel inputs are made based on how they activate existing CAs. If a test quadruple excites more neurons in *NAttachNet* than *VAttachNet* over a fixed number of cycles, it is assumed that the quadruple has a noun attachment.

If the system had been trained on example 1, testing on that example would typically cause many more neurons to fire in the VAttachNet than in the NAttachNet. First, having cofired, the synapses from the neurons in the quadruple and their sense hierarchies, and the specifically stimulated neurons in the VAttachNet would have high weights. When restimulated during testing, the quadruple neurons would cause those initial neurons to fire. Second, as other neurons in the VAttachNet may have fired during the training presentation of the quadruple, they would also have high incoming synaptic weights from the quadruple and might fire. Third, as the specifically stimulated neurons in the VAttachNet may have cofired with other VAttachNet neurons in later training episodes, those other VAttachNet neurons might fire. In the second and third cases, a VAttachNet CA

Table 3: Test Results.

Attachment	Total Number in the Test Set	Correct Predictions	Accuracy
Verb	2416	2416	100%
Noun	584	234	40.06%
Total	2650	3000	88.33%

has spread beyond the initial 10 neurons stimulated during training from example 1.

Of course, due to the sparse nature of PP ambiguity, it is unlikely that any training quadruple will occur in the test; indeed none does. However, the effect still persists due to the hierarchical nature of quadruple encoding. When a test quadruple is presented, some of the neurons from its sense hierarchies stimulate neurons in the VAttachNet and in the NAttachNet.

6.3 Testing and Results. After training, learning in the subnets was switched off, and the 3000 novel quadruples from the test set were presented to the four input subnets for 100 cycles each. The number of neurons the inputs caused to fire in *VAttachNet* and *NAttachNet* was recorded over these cycles, and the decision was attributed to the most active subnet. If *VAttachNet* was most active, the decision was verb attachment, and if *NAttachNet* was most active, the decision was noun attachment. Over six trials, the system correctly disambiguated the test set with an average accuracy of 84.56% (standard deviation of 2.85). One of the trials yielded an accuracy of 88.33%. However, this cannot be considered a definitive measure of the model's performance, and further tests are required to verify its significance. It is important to note that unlike other machine learning models, quadruples with the preposition *of* were not included in the training and test sets due to computational size limitations. It is known that the preposition *of* always attaches to the NP. Including these omitted quadruples would yield better accuracy.

The results from a typical trial run, here from the trial that yielded the highest accuracy, are presented in Table 3. All incorrect predictions had noun attachments and had a larger number of neurons active in *VAttachNet* than *NAttachNet*. This seems to be due to the large number of training quadruples with verb attachments (79.9%) and the sparseness of the quadruples. That is, since words in the four input subnets do not have duplicates, frequently reoccurring verbs and nouns have stronger connections to *VAttachNet* due to the sheer volume of quadruples with verb attachments.

CAs may have excitatory connections with each other based on the connection rule (see equation 4.4), but their intersynaptic weights are generally low because neurons in different CAs seldom cofire. Table 4 shows the Pearson's correlation coefficients of three different network states of *VerbAttach*

Table 4: Pearson's Correlation Coefficient of Different States of *VerbAttach*.

put stock on list	*put touch on compromise*	0.52
put touch on compromise	*enter venture in april*	0.11
enter venture in april	*put stock on list*	0.1

Table 5: Comparison with Previous Work.

Method	Result
Maximum entropy model (Ratnaparkhi et al., 1994)	81.6%
Decision trees and WordNet (Stetina & Nagao, 1997)	88.1%
Nearest-neighbor method (Zhao & Lin, 2004)	86.5%
Learning random walk models for inducing word dependency distributions (Toutanova et al., 2004)	87.5%
Corpus-based PP attachment ambiguity resolution with a semantic dictionary (Nakov & Hearst, 2005)	83.8%
Semantic hierarchies for lattice construction (Nadh & Huyck, 2009)	88.1%
Semantic hierarchies as overlapping CAs	
Average	**84.56%**
Highest in a single trial	88.33%

during training. The coefficients are a measure of linear dependence and show the similarity of different states of the network, where the states are represented by firing neurons. A high coefficient suggests that a large number of the same neurons fired during both instances. The network states shown in the table were measured at cycles $t = 2250$, $t = 2750$, and $t = 5150$, while the network was training on the quadruples *put stock on list*, *put touch on compromise*, and *enter venture in april* respectively. The quadruples have verb attachments and were chosen so as to illustrate the differences in CAs representing them.

7 Discussion and Conclusion

Many systems have been used to resolve PP ambiguity, but the lack of a standard test set makes comparison difficult. Nonetheless, results from some previous PP attachment ambiguity resolution systems are in presented in Table 5.

The intersubnet connections (see section 6.1) were limited to 90 synapses per neuron due to computational resource constraints (available memory), totaling over 18 million connections from the four input subnets to the attachment subnets. This was due to the large number of neurons present in the subnets. In contrast, neurons in the brain have thousands of synapses (DeFelipe, Marco, Busturia, & Merchn-Prez, 1999). Each trial run of the simulation took about 8 hours to complete, with the presentation of a single quadruple to the subnets taking an average of 3 seconds, prolonging the development time significantly. For this reason, the significance of the single trial result of 88.33% could not be verified. Increasing the number of connections per neuron may have produced better results as the connectivity of word CAs spread across the large subnets would be better. Also, the size of both attachment subnets was kept in proportion to the maximum number of attachment decisions of either type of attachment in the training set in order to reduce bias. Thus, a large number of neurons in *NAttachNet* never participated in any activity as the number of noun attachments in the training set is considerably lower (20.1%) than the verb attachments. If *NAttachNet* had fewer neurons than *VAttachNet*, increased connection density from the four input subnets would activate a larger number of neurons in *NAttachNet* frequently, biasing the results.

When humans make attachment decisions, they are governed by the semantics of the quadruples, the full phrases, and the context. Hebbian theory states that the semantics of the words are stored in CAs, and the CAs in turn are part of a large associative memory. If humans made attachment decision, even based on just quadruples, their rich semantics would almost certainly enable better performance. In this letter, the semantics of the base word CAs are very simple having been derived directly from the WordNet hierarchy. These are a proxy for the rich semantics humans store for words. Similarly, the CAs that have been learned in the simulations for attachment decisions are a proxy for the much richer semantics stored in human associative memory.

It is also fair to note that while the simulation has a degree of biological plausibility, it is a far from perfect model of brain function. The FLIF neural model is a point model, and existing compartmental models (Hodgkin & Huxley, 1952) are more accurate. The learning rule, while Hebbian, cannot account for biological behavior of less than 10 ms. differences. Similarly, the topology is far from biologically accurate in both number of synapses per neuron and overall number of neurons. Nonetheless, while far from biological realism, the system is a step in that direction for a machine learning system.

Although CAs of individual words and attachment decisions form as a result of the external stimulation of the 20 designated neurons for each word, large, overlapping CAs emerge internally from coactivation of these individual CAs. While it is difficult to assess the nature of these emergent CAs as they change dynamically as learning progresses, it is clear from the results that they embody characteristics of their constituent word CAs as an associative memory. While word CAs inherently have no meaning and semantic relationships between them are encoded in the network simply by coactivation of their neurons, the emergent CAs seem to encompass their hierarchical relationships in some form.

To verify that CA dynamics are indeed what enabled the classification of novel inputs, the trials were repeated five times with all intranet neuronal connections in *VAttachNet* and *NAttachNet* dettached. This ensured that no CAs formed in either subnets and that there was no recurrent activity between their constituent neurons. The trials yielded a poor resolution accuracy of 19.09% (standard deviation of 0.62), confirming that it is indeed CAs that form in the two attachment subnets that enable the task.

There is evidence that CAs in the brain are involved in higher-order cognitive phenomena (Pulvermuller, 1999; Funahashi, 2001; Pasupathy & Connor, 2002). Although the physical characteristics of biological CAs are known to an extent, their precise dynamics, which give rise to higher-order phenomena, including the fundamental process of associative memory, are yet to be understood. Nonetheless, computational models of CAs have been used to model a wide range of tasks (Huyck, 2000; Wennekers, 2007; Knoblauch et al., 2007). While the model is expensive in terms of computation, this letter is another such piece of work aimed at understanding the nature of CAs, which used the PP attachment disambiguation task as the means of doing so.

Acknowledgments

This work was partially supported by EPSRC grant GR/R13975/01.

References

Abbott, L., & Nelson, S. (2000). Synaptic plastiticy: Taming the beast. *Nature Neuroscience*, *3* 1178–1183.

Agirre, E., Baldwin, T., & Martinez, D. (2008). Improving parsing and PP attachment performance with sense information. In *Proceedings of ACL-08: HLT* (pp. 317–325). Columbus, OH: Association for Computational Linguistics.

Anderson, J. R., & Bower, G. H. (1980). *Human associative memory: A brief edition*. Erlbaum.

Atterer, M., & Schütze, H. (2007). Prepositional phrase attachment without oracles. *Computational Linguistics*, *33*(4), 469–476.

Byrne, E., & Huyck, C. (2010). Processing with cell assemblies. *Neurocomputing*, *74*(1–3), 76–83.

Charniak, E. (1997). Statistical techniques for natural language parsing. *AI Magazine*, *18*, 33–44.

Churchland, P., & Sejnowski, T. (1992). *The computational brain*. Cambridge, MA: MIT Press.

DeFelipe, J., Marco, P., Busturia, I., & Merchn-Prez, A. (1999). Estimation of the number of synapses in the cerebral cortex: Methodological considerations. *Cerebral Cortex*, *9*, 722–732.

Ford, M., Bresnan, J., & Kaplan, R. (1982). A competence-based theory of syntactic closure. In J. Bresnan (Ed.), *The mental representation of grammatical relations*. Cambridge, MA: MIT Press.

Foster, J., & Alexander, G. (1971). Neuron activity related to short-term memory. *Science*, *173*, 652–654.

Frazier, L. (1983). Processing sentence structure. In K. Raynoer (Ed.), *Eye movements in reading*. New York: Academic Press.

Funahashi, S. (2001). Neuronal mechanisms of executive control by the prefrontal cortex. *Neuroscience Research*, *39*, 147–165.

Gerstner, W. (2002). Integrate-and-fire neurons and networks. In M. Arbib (Ed.), *The handbook of brain theory and neural networks* (Vol. 2, pp. 577–581). Cambridge, MA: MIT Press.

Gerstner, W., & Kistler, W. K. (2002). Mathematical formulations of Hebbian learning. *Biological Cybernetics*, *87*, 404–415.

Harris, K. (2005). Neural signatures of cell assembly organization. *Nature Reviews Neuroscience*, 6, 399–407.

Hebb, D. O. (1949). *The organization of behavior*. New York: Wiley.

Hindle, D., & Rooth, M. (1993). Structural ambiguity and lexical relations. *Computational Linguistics*, *19*, 103–120.

Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology*, *117*, 500–544.

Huyck, C. (2000). A practical system for human-like parsing. In W. Horn (Ed.), *Proceedings of the 14th European Conference on Artificial Intelligence*. Fairfax, VA: IOS Press.

Huyck, C. R. (2001). Cell assemblies as an intermediate level of cognition. In S. Wermster & J. Austin (Eds.), *Emergent neural computational architectures based on neuroscience* (pp. 383–397). New York: Springer.

Huyck, C. R. (2004). Overlapping cell assemblies from correlators. *Neurocomputing*, *56*, 435–439.

Huyck, C. R. (2008). CABot1: A videogame agent implemented in fLIF neurons. In *Proceedings of Cybernetic Intelligent Systems, 7th IEEE International Conference* (pp. 1–6). Piscataway, NJ: IEEE.

Huyck, C. (2011). Parameter values for FLIF neurons. In *Complexity, Informatics and Cybernetics: IMCIC 2011*. Winter Garden, FL: International Institute of Informatics and Systemics.

Huyck, C. R., Belavkin, R., Jamshed, F., Nadh, K., Passmore, P., Byrne, E., et al. (2011). CABot3: A simulated neural games agent. In *Proceedings of the International Joint Conferences on Artificial Intelligence*. AAAI Press/International Joint Conferences on Artificial Intelligence.

Huyck, C. R., & Nadh, K. (2009). Multi-associative memory in fLIF cell assemblies. In A. Howes, D. Peebles, & R.C.E. Cooper (Eds.), *9th International Conference on Cognitive Modelling* (pp. 81–87). Manchester, UK.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. New York: Oxford University Press.

Kable, J. W., Lease-Spellmeyer, J., & Chatterjee, A. (2002). Neural substrates of action event knowledge. *J. Cognitive Neuroscience*, *14*, 795–805.

Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition*, *2*, 15–47.

Knoblauch, A., Kupper, R., Gewaltig, M.-O., Körner, U., & Körner, E. (2007). A cell assembly based model for the cortical microcircuitry. *Neurocomputing*, *27*, 649–677.

LaBar, K. S., Gitelman, D. R., Parrish, T. B., & Mesulam, M. (1999). Neuroanatomic overlap of working memory and spatial attention networks: A functional MRI comparison within subjects. *Neuroimage*, *10*, 695–704.

Lee, C., Lee, G., & Yun, S. J. (2000). Automatic WordNet mapping using word sense disambiguation. In *Proceedings of the 2000 Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora: Held in Conjunction with the 38th Annual Meeting of the Association for Computational Linguistics*, *13*, (pp. 142–147). Stroudsburg, PA: Association for Computational Linguistics.

Lin, D. (1998). Dependency-based evaluation of MINIPAR. In *Proc. Workshop on the Evaluation of Parsing Systems*. Sussex: University of Sussex.

Maass, W., & Bishop, C. M. (2001). *Pulsed neural networks*. Cambridge, MA: MIT Press.

Marcus, M. P., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of english: The Penn Treebank. *Computational Linguistics*, *19*, 313–330.

Maurer, A. P., Cowen, S. L., Burke, S. N., Barnes, C. A., & McNaughton, B. L. (2003). Organization of hippocampal cell assemblies based on theta phase precession. *Hippocampus*, *16*, 785–794.

Miller, G. (1990). Wordnet: An on-line lexical database. *International Journal of Lexicography*, *3*, 235–244.

Moens, M., Calder, J., Klein, E., Reape, M., & Zeevat, H. (1989). Expressing generalizations in unification-based grammar formalisms. In *Proceedings of the Fourth Conference on European Chapter of the Association for Computational Linguistics, EACL* ’89, (pp. 174–181). Stroudsburg, PA: Association for Computational Linguistics.

Nadh, K. (2010). Learnt novelty with overlapping cell assemblies. In *Proceedings of the York Doctoral Symposium* (pp. 47–54). York: University of York.

Nadh, K., & Huyck, C. R. (2009). Prepositional phrase attachment ambiguity resolution using semantic hierarchies. In *The Ninth IASTED International Conference on Artificial Intelligence and Applications*. Calgary, AB: ACTA Press.

Nadh, K., & Huyck, C. R. (2010). A pong playing agent modelled with massively overlapping cell assemblies. *Neurocomputing*, *73*, 2928–2934.

Nakov, P., & Hearst, M. (2005). Using the Web as an implicit training set: Application to structural ambiguity resolution. In *Proceedings of HLT-NAACL*. Stroudsburg, PA: Association for Computational Linguistics.

Nicolelis, M., Baccala, L., Lin, R., & Chapin, J. (1995). Sensorimotor encoding by synchronous neural ensemble activity at multiple levels of the somatosensory system. *Science*, *268*, 1353–1358.

Osan, R., Chen, G., Feng, R., & Tsien, J. Z. (2011). Differential consolidation and pattern reverberations within episodic cell assemblies in the mouse hippocampus. *PLoS ONE*, *6*(2), e16507.

Pasupathy, A., & Connor, C. (2002). Population coding of shape in area V4. *Nature Neuroscience*, *5*, 1332–1338.

Pulvermuller, F. (1999). Words in the brain’s language. *Behavioral and Brain Sciences*, *22*, 253–336.

Ratnaparkhi, A., Reynar, J., & Roukos, S. (1994). A maximum entropy model for prepositional phrase attachment. In *Proceedings of the ARPA Workshop on Human Language Technology* (pp. 250–255). Stroudsburg, PA: Association for Computational Linguistics.

Schoenbaum, G. (1998). Cell assemblies and the ghost in the machine. In H. Eichenbaum & J. Davis (Eds.), *Neuronal ensembles: Strategies for recording and decoding* (pp. 81–116). New York: Wiley.

Schutze, C., & Gibson, E. (1999). Argumenthood and English prepositional phrase attachment. *Journal of Memory and Language*, *40*, 409–431.

Stetina, J., & Nagao, M. (1997). Corpus based PP attachment ambiguity resolution with a semantic dictionary. In *Proceedings of Workshop on Very Large Corpora* (pp. 66–80). New York: Springer.

Toutanova, K., Ng, A. Y., & Manning, C. (2004). Learning random walk models for inducing word dependency distributions. In *Proceedings of International Conference on Machine Learning*. New York: ACM.

Vicario-Abejn, C., Collin, C., McKay, R.D.G., & Segal, M. (1998). Neurotrophins induce formation of functional excitatory and inhibitory synapses between cultured hippocampal neurons. *Journal of Neuroscience*, *18*, 7256–7271.

Wennekers, T. (2007). A cell assembly model for complex behaviour. *Neurocomputing*, *70*, 1988–1992.

Wennekers, T., & Palm, G. (2000). Cell assemblies, associative memory and temporal structure in brain signals. In R. Miller (Ed.), *Time and the brain: Conceptual advances in brain research* (Vol. 2, pp. 251–274). Newark, NJ: Harwood Academic Publishers.

Zhao, S., & Lin, D. (2004). A nearest-neighbor method for resolving PP-attachment ambiguity. In *Proceedings of the International Joint Conference on Natural Language Processing* (pp. 428–434). New York: Springer.

Received August 10, 2011; accepted December 9, 2011.