A Single Layer Network Model of Center Embedding and Hierarchical Phrase Structure in Sentence Processing

Simon Dennis (Simon.Dennis@gmail.com)
Department of Psychology, 1835 Neil Avenue,
Columbus, OH 43210 USA

Dennis N. Mehay (mehay@ling.ohio-state.edu)
Department of Linguistics, 1712 Neil Avenue
Columbus, OH 43210 USA

Abstract

Recurrent connectionist models, such as the simple recurrent network (SRN, Elman, 1991), have been shown to be able to account for people's ability to process sentences with center embedded structures of limited depth without recourse to a competence grammar that allows unbounded recursion (Christiansen & Chater, 1999). However, the corpus analysis by Karlsson (2007) suggests that in spoken language people are restricted to a single level of embedding (see also, Reich, 1969). Models like the SRN may be too powerful as they can be trained to represent an arbitrary level of embedding, albeit with diminishing performance. We show that a single layer softmax network in which order is represented by slowly decaying activations on the input units can capture one level of embedding and only one level. In addition, the representations produced by this network capture the hierarchical phrase structure of sentences without employing a recursive mechanism. Furthermore, we trained the network on a corpus of 1,000,000 sentences of naturally occuring newswire text and were able to capture the relative processing difficulties of subject versus object extraction in relative clauses and questions, singly versus doubly embedded relative clauses and singly, doubly and triply embedded clausal modifiers each of which have been identified as important test cases for models of sentence processing (Gibson, 1998).

Keywords: sentence processing, connectionist networks, center embedding

Connectionist Sentence Processing

Chomsky (1957) argued that the presence of recursive structures such as center embedded clauses in principle rules out associative explanations of the language processing mechanism. This argument has been challenged in many ways both by disputing the empirical claim that humans are capable of processing recursive structures (Reich, 1969) and the computational claim that associative mechanisms, particularly associative mechanisms that employ hidden unit representations in the connectionist tradition, are unable to process recursive structures, at least of the depth observed in human performance (Christiansen & Chater, 1999).

Early attempts to investigate the capabilities of connectionist networks to capture linguistic structure fell into two approaches (Christiansen & Chater, 1999). In the first approach, networks were provided with tagged datasets that provided information about the extent and identity of constituents (Chalmers, 1990; Pollack, 1988) and required the network to generalize these mappings. While these models demonstrated the representational abilities of networks, the fact that they required labelled training data of a kind that is unlikely to be available to human learners meant that their relevance to the question of how linguistic structure is acquired was limited. A second approach involved learning simplified tasks such as idenitfying the $a^n b^n$ language from raw input strings using small networks (Wiles & Elman, 1995). This work demonstrated that recursive generalization was possible to a significant degree, but it remained unclear whether these results would apply to other kinds of recursion and with expanded vocabularies.

Christiansen and Chater (1999) expanded previous work significantly by demonstrating that the simple recurrent network (Elman, 1991) was capable of capturing the three main kinds of recursive structures that were proposed by Chomsky (1957) as problematic for finite state systems. These were *counting* recursion (e.g. ab, aabb, aaabbb) of the kind studied by Wiles and Elman (1995), *mirror* recursion (e.g. abba, aabbaa, abbbba) also known as center embedding and *identity* recursion (e.g. abbabb, aabbaabb), which captures structures found in Swiss German and in Dutch.

There are, however, three critical objections to proposing the simple recurrent network as a model of human sentence processing. The first is that the computational properties of the SRN may be too powerful to capture human performance. Although performance on doubly center embedded sentences in the networks trained by Christiansen and Chater (1999) was significantly lower than that of singly embedded sentences, this was at least in part a consequence of the fact that these structures appeared very infrequently in the training corpus. In principle, however, the SRN is capable of learning doubly embedded sentences.

Karlsson (2007) conducted a systematic corpus analysis of large scale corpora from English, Finnish, French, German, Latin, Swedish, and Danish and concluded that while examples of double center embedding could be found in written language they were practically absent from spoken language. Given that written language is subject to editing processes that might increase complexity, the spoken results would seem to be germane to determining the capabilities of the online sentence processor. In that case, there is little evidence for embedding beyond a single level - support for a hypothesis first proposed by Reich (1969).

In addition, Gureckis and Love (2007) found that in the serial reaction time task subjects could quickly acquire linearly separable patterns such as the serial AND or OR tasks, but failed to acquire a serial XOR task. They demosntrated that the learning pattern that subejcts displayed was consistant with a single layer connectionist network, but could not be captured by the simple recurrent network. Note that a single level of center embedding is equivalent in structure to the to the AND or OR tasks, but a double level of embedding requires a more complex mapping equivalent to the XOR task.

Secondly, the SRN does not provide a notion of hierarchical constituency (see Figure 1 for an example). That sentences should be decomposed in this way is supported by an array of linguistic evidence (see (Radford, 1988) for an overview). To take one example, consider the appropriate responses to teh question "Where are you going?". You might say "to the cinema" or "the cinema", but you would not say "cinema", or "to the". "to the cinema" is a prepositional phrase and "the cinema" is a noun phrase and these constituents are organized hierarchically.

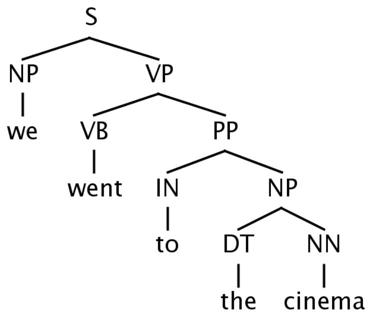

Figure 1: The phrase structure tree for "we went to the cinema". S = sentence, VP = verb phrase, NP = noun phrase, PP = prepositional phrase, NN = noun, VB = verb, IN = preposition, DT = determiner.

Finally, the SRN does not scale well. Ideally one would take the network, train it on a corpus of naturally occurring text and then test on example sentences taken from psycholinguistic experiments to assess its performance against that of humans. As vocabulary size, the number of hidden units and the corpus size increase, however, training times become prohibitive and so it is impractical to compare human and network performance directly. As a consequence training is restricted to small corpora selected to exhibit some specific property. The extent to which these constructed corpora are truely representative of the environmental statistics to which the human sentence processor is exposed is questionable, however. The failure to scale well has meant that the impact of the SRN in applied domains of computational linguistics has been limited.

In the next section, we outline a single layer network architecture that we believe addresses these concerns. Then we consider each of the issues in turn: performing one and only one layer of center embedding, capturing hierarchical phrase structure and the ability to scale to large naturally occuring corpora and to be compared directly to materials from psycholinguistic investigations showing how the model demonstrates these attributes.

The Architecture

The model that we will outline is a version of the syntagmatic paradigmatic model (Dennis, 2005). The SP model was designed as a model of verbal cognition. It is based on the distinction between **syntagmatic** associations that occur between words that appear together in utterances (e.g. run fast) and **paradigmatic** associations that occur between words that appear in similar contexts, but not necessarily in the same utterances (e.g. deep and shallow, c.f. Ervin-Tripp, 1970). The model has been used to account for a number of phenomena including long term grammatical dependencies and systematicity (Dennis, 2005), the extraction of statistical lexical information (syntactic, semantic and associative) from corpora (Dennis, 2003a), sentence priming (Harrington & Dennis, 2003), verbal categorization and property judgment tasks (Dennis, 2005), serial recall (Dennis, 2003b), and relational extraction and inference (Dennis, 2005, 2004).

Previous instantiations of the model have used string edit theory (SET) as a way of capturing the notion of paradigmatic association. While SET has been a fruitful mechanism with which to investigate the properties of the general theoretical framework, it has proven inadequate in some respects. In particular, it shares the property with the SRN that it does not provide any inherent constraint of the level of center embedding that people should be able to process. As a consequence, we propose to replace SET with a single layer network.

To model word distributions, we have implemented a single-layer softmax neural network (Bridle, 1990), where the activation inputs for each word in a sentence fire according to decayed weights for all surrounding words in each training and test sentence. The network has two weights between each word w and every other word w': one for w' when it occurs to the left of w and one for w' when it occurs to the right of w. In this way the network can learn the interactions between leftward and rightward word co-occurrences, as well as the influence of certain words when they occur *between* these co-occurrences.

The probability of a word w_k in the context of a sentence of m words $s = w_1 ... w_m$ has the following exponential form:

$$p(w_k \mid \vec{W} \cdot s) = \frac{\exp\left(\vec{W}_k \cdot \vec{\mathrm{D}}_s\right)}{\sum_{i=1}^{V} \exp\left(\vec{W}_i \cdot \vec{\mathrm{D}}_s\right)}$$

where V is the vocabulary size $\vec{W} = [W_1, ..., W_V]$ is a vector of vectors each of size $2V$ (i.e., the weight matrix) corresponding to the strengths learned between each word in the vocabulary and every word as it either appears to the right or left. The denominator is a normalizing sum so that the network is probability distribution. $\vec{\mathrm{D}}_s$ is a non-sparse vector of decayed counts corresponding to the number of times each word token in the context s currently appears to the left and right of the current word, and each such count is in a position corresponding to the $2V$-length weight vector. A contextual word token contributes to this count according to the exponential d-parameterized term $\exp\big(-d \cdot \mathrm{dist}(w_j, w_k)\big)$, so that the influence of words drops and then tapers off exponentially as the distance increases. Figure schematically illustrates the network with input strengths decaying as they go from white to black.

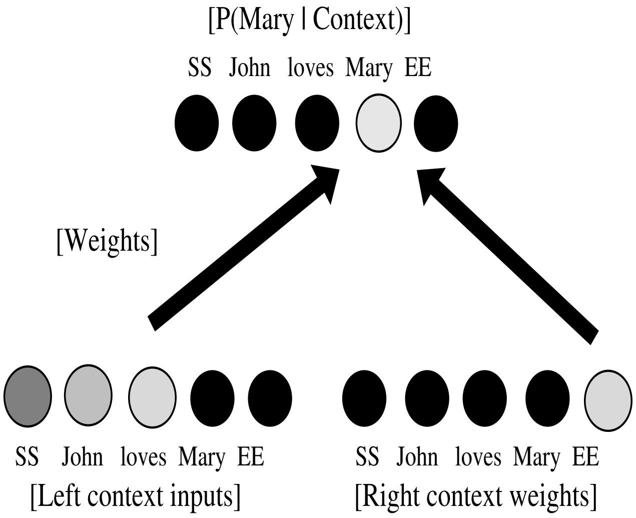

Figure 2: The sigle layered network architecture with decaying input units.

We train using simple gradient descent, as more sophisticated methods require multiple passes over the data set, which would be prohibitively time-consuming given the size (≈ 27 million words) of some of our data sets.

Center Embedding

Karlsson (2007) conducted the largest systematic corpus analysis of center embedding phenomena to date. He employed the Brown and the LOB corpora, The International Corpus of English. The British Component, an analysis of sentences by three 19th-century scholars (Jeremy Bentham, John Stuart Mill, and C. S. Peirce), more than one hundred descriptive, stylistic, and diachronic papers, grammars, and style manuals in multiple languages including latin and german, corpus examples provided by other authors in the literature and from the LINGUIST mailing list and naturalistic observation of newspapers and books over the years. While he was able to find 132 examples of double center embedding in written langauge, he only found three in spoken language. One of these was an anacadote, one was cited from other authors and just one example was actually observed. Karlsson (2007) concludes that double center embedding in spoken language is practically nonexistant supporting the conjecture by Reich (1969) that people are capable of the online processing of one level of center embedding and one only.

To demonstrate that an architecture can process a given level of embedding it is necessary to pick a binary distinction (such as plurality) and show that the netowrk is able to make the appropriate prediction in all cases. For instnace, if we take noun verb plurality agreement as an example, to demonstrate a single level of embedding one would have to show that the network could predict the correct form of the verb in the following cases:

```
NNs NNs VBs VBs
NNs NNp VBp VBs
NNp NNs VBs VBp
NNp NNp VBp VBp
```

where NNs stands for a singular noun, NNp stands for a plural noun, VBs stands for a singular verb and VBp stands for a plural verb. Following the final noun in these structures the next verb should be of matching plurality. This can be achieved by learning positive weights of approximately equal magnitude between the NNs input and the VBs output and the NNp input and the VBp output. The more difficult case comes when trying to predict VBs when the initial noun was a NNs but the intervening noun is an NNp (or symmetrically when the initial noun is a NNp and the intervening noun is an NNs). In this case, the NNp will be more strongly active and hence will continue to predict the NNp. However, the desired result can be achieved by introducing a negative weight between the VBp input node and the VBp output. Then when the VBp occurs it effectively turns itself off allowing the influence of the initial noun to control the subsequent prediction. When the model is trained it finds this solution. The use of negative weights to turn off the influence of previous words when a suitable dependent has appeared is an interesting mechanism could be used to explain how depnedency formalisms (e.g. link grmamar, Sleator & Temperly, 1993) could resolve constraints without recourse to search. In the large scale model that we train in the final section of the paper in the vacinity of 80% of the weights are negative indicating that to a large extend what the network learns is what cannot occur next.

Note, however, that such a mechanism is note sufficient to predict the verbs when two levels of embedding must be interpreted unless activations decay faster than $2^{-}d$ where d is the distance between the current item and the input item because under these conditions the activations of two previous nouns of one type will be greater than the activation of the immediately preceding noun. So, for instance, if you have the pattern NNs NNs NNp then the activations of the first and second singular nouns will add and become larger than that of the plural noun generating a prediction of a singular verb instead of the plural verb that should appear next. Simulations confirm that the network is unable to learn two levels fo embedding.

Hierarchical Phrase Structure

Given that there is no process by which linguistic units are combined to create larger linguistic units in the model outlined above, it may seem that the architecture would not be capable of capturing hierarchical grammatical structure. However, indiviudal words can be bound together indirectly by being paradigmatically bound to similar patterns. To see how this is possible consider the phrase structure tree for the simple sentence "we went to the cinema". As we outlined above "the cinema" is a noun phrase embedded within the prepostional phrase "to the cinema" which is in turn embedded in the verb phrase "went to the cinema" (see Figure 1).

Now consider the probability distributions that assigned to each of the words in this sentence (see figure 3 if the model is trained on the following corpus:

```
SS we saw EE
SS we went there EE
SS we went to Paris EE
SS we went to the cinema EE
```

where SS and EE are end of sentence markers.

```
we:      we 92 went 4
went:    went 87 to 4 we 3 saw 2 the 2
to:      to 77 the 7 there 6 went 4 saw 2 Paris 2
the:     the 62 Paris 14 to 8 cinema 5 there 4 saw 2
cinema:  cinema 62 Paris 14 the 5 there 5 saw 2 to 2
```

Figure 3: Probability distributions associated with each word in the sentence "we went to the cinema" when the model is trained on the small corpus provided in the text. Note numbers have been multiplied by 100 and any words with probabilities less than 0.01 have been removed. End of sentence markers were also removed for clarity. Learning rate was 0.02, decay was 0.01 and 200 iterations of learning were performed.

As a consequence of training the network learns that the word "saw" can appear between the word "we" and the end of sentence marker "EE". If the network is then presented with the sentence "SS we went to the cinema EE" the word "saw" is paradigmatically associated with each of the words "went", "to", "the" and "cinema" as they also appear between "we" and "EE". The word is in a sense taking on the role of a verb phrase symbol. Similarly, the word "there" is taking on the role of a prepositional phrase symbol and the word "Paris" is taking on the role of a noun phrase symbol. The network is sensitive to the distributional properties of word strings which are key tests of constituency within transformation approaches (Radford, 1988) and which underpin successful unsupervised grammar induction methods (Klein & Manning, 2001).

Of course, exposure to a more complete corpus would result in distributed patterns of activity that would correspond to the different phrase types, and it would then not be possible to identify individual words that correspond to individual phrase types. To test whether this meachanism would be sufficient to capture a more complicated example the model was trained on 10,000 example sentences generated with the following grammar:

```
NNP -> {mark, simon, paul, luke, john, bob, sue,
     viki, laura, kathryn, trish, alison, vladimir,
     wil, michael, jim, roger, bill, sofie, maddy,
     liv, mari, sydney, angela}
NN -> {dog, cat, bird, snowman, lawyer,
       fireman, cricketer, doctor}
VB -> {loves, detests, believes, sees, knows}
DT -> {a, the}
WP -> who
NP -> NNP |
      DT NN |
      DT NN WP VP
VP -> VB |
      VB NP
S -> NP VP
```

A learning rate of 0.02 and a decay rate of 0.01 were used and the model and a single iteration through the training examples was executed. Note that this grammar is recursive and permits arbitrarily deep trees. For instance, the sentence "the snowman detests the lawyer who knows jim" could be represented by the tree in Figure 4.

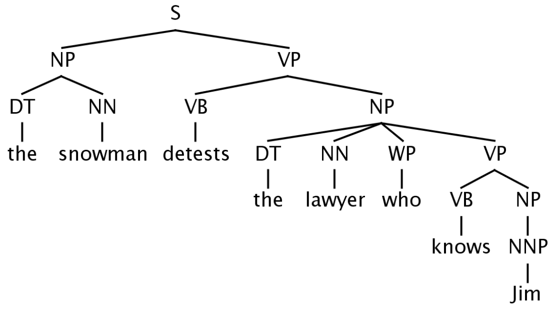

Figure 4: The phrase structure tree for "the snowman detests the lawyer who knows jim". NNP = noun proper, WP = Wh-pronoun.

For any word in this network, one can follow the links upward to identify the set of constituents to which it belongs. For instance, the word "jim" is part of an S, two VPs, two NPs and is tagged with an NNP. Next 10,000 new sentences were generated along with there labelled constituents and a second linear network was trained to predict the set of constituents (including the part of speech tags) that the word belongs to from the probability distribution over words that is associated with it. You might think of this network as playing the role of a linguistic looking at the internal representations for words and deciding on the constituents which they must appear in. To test whether there exists sufficient information in these probability distributions to distinguish the constituents to which the word belongs a third labelled exemplar set was constructed and the linear network was used to predict the constituents to which each word belonged.

Each word was assigned a single part of speech tag and so the prediction is more straightforward for these classes - although note the word itself does not contribute directly to the probability distribution to which it is assigned, only the context words do. The mean prediction for each of the part of speech tags is given in Table 1. Performance is good except perhaps for the proper nouns, probably because there are many multiword sequence that have the same context as proper nouns in the grammar.

S	E	DT	VB	NN	NNP	WP
0.99	1.00	0.99	0.95	0.96	0.44	0.99

Table 1: Mean linear prediction for each part of speech. E = end of sentence marker

Figure 5 shows the mean predictions for the VP and NP constituents. Note the challenge is not just to determine whether a word belongs to a noun phrase or a verb phrase, but also to determine how many noun phrases and verb phrases it is a part of. The graph shows the predicted number against the actual. While performance declines as the embedding level increases it is clear that the network is capable of distinguishing embedding at different levels.

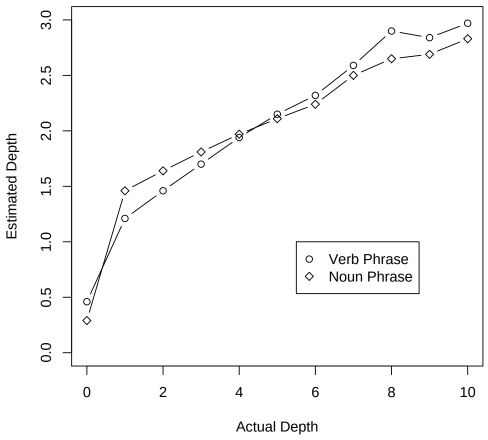

Figure 5: The mean predicted versus actual level of noun phrase and verb phrase embedding.

The results in section demonstrate that despite the fact that the model has no mechanism by which words are combined to form higher level units, nonetheless a notion of constituency can be extracted from the flat probability distributions that are paradigmatically associated with each word. Constituents may form not because words are connected directly to each other, but rather because they are associated with common patterns of activation. While it remains to be seen whether this notion of constituency is sufficient to satisfy all of the purposes to which linguists have put the concept, Dennis (2005) has shown how this a similar mechanism can help to explain a number of other sentence processing phenomena such as the emergence of transformational regularities and the unsupervised induction of semantic roles.

Processing Difficulty of Sentence Types: An Initial Comparison to the Psycholinguistic Data

In the introduction, we argued that the inability of the simple recurrent netork to scale to large vocabularies and large corpora was an important disadvantage as it prevents the direct comparison of the model against empirical sentence processing data. In this section, we describe a large scale simulation of our single layered model and an initial comparison to some sentence processing data reviewed by Gibson (1998).

We trained single layered archiecture outlined above on ≈ 1.2 million sentences of New York Times newswire text from the Gigaword corpus[1] (sections *nyt200001–nyt200003*). There were ≈ 27 million words, numerals and punctuation tokens in this corpus. We kept the 10,017 most frequent vocabulary items in the corpus[2] and replaced all others with tokens of the form *xxx*<suffix>, where <suffix> was the longest matching suffix from a small list of common suffixes such as *-ed*, *-tion*, etc. All tokens that did not have a matching suffix were left as *xxx*. In this way, we were able to accumulate additional information about infrequent vocabulary items.

We tested the sensitivity of our model to syntactic phenomena that have been identified as important test cases for models of sentence processing (Gibson, 1998) by measuring the relative perplexities it assigned to sentences exhibiting these phenomena. More precisely, using only in-vocabulary words, we hand-constructed a small corpus of sentences exhibiting these phenomena (modelled on examples given in (Gibson, 1998)) and computed their perplexities as $2^{H[\hat{p},q](w_1,\ldots,w_n)}$, where $H[\hat{p},q](w_1,\ldots,w_n)$ is the cross-entropy of our model q relative to the empirical distribution $\hat{p}$ with respect to the words $w_1,\ldots,w_n$ in a sentence. These sentences were crafted to exhibit subject versus object extraction from relative clauses and questions, singly versus doubly embedded relative clauses and singly versus doubly versus triply embedded clausal modifiers. As noted in Gibson (1998), humans generally find subject extraction in relative clauses more difficult to process than object extraction, while the reverse is true for questions. Further, human sentence processing difficulty increases in proportion to the number of levels of relative clause or clausal modifier embeddings. Each sentence in a pair (or triple) under comparison was constructed using the same words as all other members of that comparison group — only the order of the words was varied.

We found that our model is sensitive to the relative difficulties of the syntactic patterns exhibited in such sentences and assigns higher perplexity to those sentences that are predicted to be harder to process. Table 2 shows some example perplexities among comparison groups.

[1] Distributed by the Linguistic Data Consortium.

[2] 10,017 as there were 18 vocabulary items with the same frequency at the end of this list.

The reporter who attacked the senator admitted the error.	1.96
The reporter who the senator attacked admitted the error.	2.07
The coach who hit the player admitted his error.	1.49
The coach who the player hit admitted his error.	1.60
Who did the senator say likes the reporter?	1.84
Who did the senator say the reporter likes?	1.76
The student who the nurse helped had bothered the administrator who lost the medical reports.	3.06
The administrator who the student who the nurse helped had bothered lost the medical reports.	3.31
If the mother gets upset when the baby is crying, the father will help, so the grandmother can rest easy.	2.07
If when the baby is crying, the mother gets upset, the father will help, so the grandmother can rest easy.	3.24
Because if when the baby is crying, the mother gets upset, the father will help, the grandmother can rest easily.	4.00

Table 2: Sentences representing various levels of processing difficulty paired with perplexities from our model.

Conclusions

While the simple recurrent network has proven a useful mechanism in demonstrating the in prinicple capabilitites of connectionist models of sentnece processing, we have argued that it faces a number of fundamental challenges. Firstly, recent corpus analysis by Karlsson (2007) suggests that people are capable of online processing a single level of embedding and single level only. While performance drops quickly in teh SRN as the level of embedding increases, it is not limited to a single level in the way in which our single layer netork is. Secondly, the SRN has no obvious mechanism by which to account for hierarchical phrase structure. By contrast, the representations formed in the single layered network are capable of both the type and level of embedding of the constituents in which an item appears. Finally, we have shown that unlike the SRN, the single layered netowrk can scale to large vocabularies and corpora and can be applied effectively to the kinds of sentences found in psycholinguistics experiments.

Acknowledgments

We would like to thank Defense Research and Development Canada for their support of this project.

References

Bridle, J. S. (1990). Probabilistic interpretation of feedforward classification network outputs with relationships to statistical pattern recognition. In F. F. Soulie & J. Herault (Eds.), *Neuralcomputing: Algorithms, architectures and applications* (pp. 227–236). Springer-Verlag.

Chalmers, D. J. (1990). Syntactic transformations on distributed representations. *Connection Science*, *2*, 53 62.

Chomsky, N. (1957). *Syntactic structures*. The Hague: Mouton.

Christiansen, M. H., & Chater, N. (1999). Towards a connectionist model of recursion in human lingusitic performance. *Cognitive Science*, *23*, 157-206.

Dennis, S. (2003a). An alignment-based account of serial recall. In R. Alterman & D. Kirsh (Eds.), *Twenty fifth conference of the cognitive science society* (Vol. 25). Boston, MA: Lawrence Erlbaum Associates.

Dennis, S. (2003b). A comparison of statistical models for the extraction of lexical information from text corpora. In R. Alterman & D. Kirsh (Eds.), *Twenty fifth conference of the cognitive science society* (Vol. 25). Boston, MA: Lawrence Erlbaum Associates.

Dennis, S. (2004). An unsupervised method for the extraction of propositional information from text. *Proceedings of the National Academy of Sciences*, *101*, 5206-5213.

Dennis, S. (2005). A memory-based theory of verbal cognition. *Cognitive Science*, *29*(2), 145-193.

Elman, J. L. (1991). Distributed representations, simple recurrent networks and grammatical structure. *Machine Learning*, *7*, 195-225.

Ervin-Tripp, S. M. (1970). Substitution, context and association. In L. Postman & G. Keppel (Eds.), *Norms of word association* (p. 383-467). New York: Academic Press.

Gibson, E. (1998). Linguistic complexity: Locality of sequential dependencies. *Cognition*, *68*, 1-76.

Gureckis, T. M., & Love, B. C. (2007). Behaviorism reborn? statistical learning as simple conditioning. In *Proceedings of the cognitive science society.* Mahwah, NJ: Lawrence Erlbaum Associates.

Harrington, M., & Dennis, S. (2003). Structural priming in sentence comprehension. In R. Alterman & D. Kirsh (Eds.), *Twenty fifth conference of the cognitive science society* (Vol. 25). Boston, MA: Lawrence Erlbaum Associates.

Karlsson, F. (2007). Constraints on multiple center-embedding of clauses. *Journal of Linguistics*, *43*, 365-392.

Klein, D., & Manning, C. D. (2001). Distributional phrase structure induction. In W. Daelemans & R. Zajac (Eds.), *Connl-2001* (p. 113-120). Toulouse, France.

Pollack, J. B. (1988). Recursive auto-associative memory: Devising compositional distributed representations. In *Proceedings of the tenth annual meeting of the cognitive science society* (p. 3339). Hillsdale, NJ: Lawrence Erlbaum Associates.

Radford, A. (1988). *Transformational grammar: A first course*. Cambridge: Cambridge University Press.

Reich, P. (1969). The finiteness of natural language. *Language*, *45*, 831-843.

Sleator, D., & Temperly, D. (1993). Parsing english with a link grammar. In *Third international workshop on parsing technologies.*

Wiles, J., & Elman, J. L. (1995). Learning to count without a counter: A case study of synamics and activation landcapes in recurrent networks. In *Annual meeting of the cognitive science society* (p. 482-487). Lawrence Erlbaum Associates.