Modeling the Use of Frequency and Contextual Biases in Sentence Processing

Neal J. Pearlmutter
Beckman Institute
405 N. Mathews St.
University of Illinois
Urbana, IL 61801
npearlmu@psych.uiuc.edu

Kim G. Daugherty
Hughes Aircraft Co.
M/S 676/Z341
P.O. Box 3310
Fullerton, CA 92632
kimd@gizmo.usc.edu

Maryellen C. MacDonald
Hedco Neuroscience Building
Univ. of Southern California
Los Angeles, CA 90089-2520
mcm@gizmo.usc.edu

Mark S. Seidenberg
Hedco Neuroscience Building
Univ. of Southern California
Los Angeles, CA 90089-2520
marks@neuro.usc.edu

Abstract

MacDonald, Pearlmutter, and Seidenberg (1993) propose an alternative to the dominant view in sentence processing that syntactic ambiguities are resolved by heuristics based on structural simplicity. MacDonald et al. argue that such ambiguities can be defined in terms of alternatives associated with information in individual lexical items, and thus that syntactic ambiguities can be resolved by lexical disambiguation mechanisms relying on access to the relative frequencies of alternatives and to biases created by contextual constraints. We present evidence from a computer simulation of the use of frequency-based and contextual constraints in the processing of the main verb/reduced relative syntactic ambiguity, showing that frequency and relatively limited contextual information from a sample of natural language can interact sufficiently to model basic results in the literature.

Introduction

Until very recently, questions about the use of frequency information in language comprehension were considered for the processing of words (lexical items) but only rarely for the processing of more complex structures (sentences and whole discourses). Frazier's (1979) well-known garden path model of sentence processing relies on complexity-based rather than frequency-based heuristics to determine the resolution of syntactic ambiguities: The most intensively investigated of these heuristics, Minimal Attachment, forces the syntactic processor to construct the phrase structure tree with the fewest nodes. However, for many Minimal Attachment ambiguities, frequency is inversely related to complexity, so a system resolving ambiguities in favor of the most frequent interpretation would display the appropriate preferences. But constructing such a system is difficult: An infinite variety of frequencies might conceivably be maintained, and any frequency-based system must somehow limit what it counts.

A number of researchers have suggested that Minimal Attachment ambiguities in the garden path model are also ambiguities over lexical representations (e.g., Juliano & Tanenhaus, 1993; MacDonald, in press; MacDonald, Pearlmutter, & Seidenberg, 1993; Trueswell, Tanenhaus, & Kello, 1993), and that therefore the resolution of these ambiguities should depend on factors relevant to the disambiguation of other cases of lexical ambiguity. Duffy, Morris, and Rayner (1988; also Simpson & Krueger, 1991; Tabossi & Zardon, 1993) showed that the processing of lexical meaning ambiguity depends on both the relative frequency of alternative meanings and the degree to which the sentence context biases in favor of one meaning or another.

On this view, frequency and contextual biases are relevant to syntactic ambiguity resolution when syntactic ambiguities are described in terms of their lexical components. For example, MacDonald et al. (1993) presented an enriched view of the lexicon in which a wide variety of information about a word was included in the word's lexical entry: In addition to information about a word's phonology, orthography, and semantics, they suggested that knowledge of a variety of grammatical and morphological features of a word was also present, including grammatical category (noun, verb, etc.), agreement features (number, person, gender), active versus passive voice, inflectional morphology (in particular, the past tense vs. past participle distinction), and argument structure.

Because of the amount of information in the lexicon and the potential complexity of interactions between different sorts of information, it is important to understand how such a system would behave during sentence processing. We are therefore currently engaged in modeling interactions between lexical and contextual constraints, and this paper reports on the first phase of a project intended to provide an implementation of ideas discussed by MacDonald et al. Argument structures are particularly important for the lexical view of syntactic ambiguity, so we first briefly describe them. We then summarize this view of ambiguity resolution and discuss the best-known example of syntactic ambiguity, the main verb/reduced relative (MV/RR) ambiguity. Finally, we present a model of the processing of the MV/RR ambiguity in which the constraints described by MacDonald et al. interact to predict interpretation preferences.

Argument Structures

Argument structures provide a shorthand mechanism for integrating syntactic and semantic information about the arguments of a word (e.g., Pinker, 1989). Argument structures are annotated lists of *thematic roles*, where the annotations provide information about the syntactic properties of the arguments, and thematic roles provide semantic information. Examples for the verb *cook* are shown in (1).

(1)a. <Agent, Theme> Murray cooked the stew.
  b. <Agent> Murray cooked (for fun).
  c. <Theme> Murray cooked (as the cannibals waited).

The argument structure in (1a) specifies that *cook* takes two arguments, the first of which (the subject; *Murray*) is assigned the Agent thematic role. Agents are canonical performers of actions and are typically animate and imbued with intensions. *Cook*'s second argument in (1a) is a direct object and receives the Theme role (*the stew*). Themes typically have actions performed on them and/or undergo changes of state, position, or possession. In (1b) and (1c), *cook* takes only one argument; in (1b), the argument is an Agent, and thus Murray cooked something, but what it was is left unspecified. In (1c), however, *Murray* is assigned the Theme role, and thus he is being cooked rather than cooking something else.

Lexical Representation and Ambiguity Resolution

MacDonald et al. proposed that a word's grammatical feature alternatives are directly associated with the word in the lexicon. Thus information in lexical entries is specifically required by the syntactic component of the system: The appropriate structure for an input can only be determined by combining lexically-based constraints (e.g., that the subject noun is plural and the verb is plural) with constraints provided by the grammar itself (e.g., that verbs and their subjects must agree in number). As a result, the availability of different alternatives in the lexicon can determine how ambiguities are resolved.

The availability of alternatives depends on their frequencies. Thus, for example, the lexicon contains information about the relative frequencies of the various argument structures for *cooked*. Thus, the relative frequency of grammatical feature alternatives ought to affect disambiguation: More frequent alternatives become available more rapidly and therefore affect (syntactic) processing sooner. MacDonald et al. also argued that contextual biases affect ambiguity resolution, although on the average, frequency biases will tend to dominate: Contextual biases provided by natural language contexts tend not to be as constraining as frequency biases can be.

MacDonald et al. showed that when lexical factors are considered, a variety of results from the syntactic ambiguity literature, many of which were incompatible with the dominant view in the field, could be explained. They therefore concluded that a range of syntactic ambiguities could be described instead as ambiguities over lexical representations, and that they could be resolved by mechanisms of lexical rather than structural disambiguation. They further proposed constraint satisfaction as the relevant mechanism for both lexical and syntactic processing, where the constraints are the relative frequencies of alternatives in the lexicon, biases created by context in favor of one or another lexical alternative, and the specification of the grammar itself (see Juliano & Tanenhaus, 1993; and Trueswell et al., in press; for similar proposals). In the next section, we discuss the MV/RR ambiguity and some specific constraints relevant to its resolution.

The Main Verb/Reduced Relative Ambiguity

The most-studied syntactic ambiguity over the past decade has been the main verb/reduced relative (MV/RR) ambiguity, as illustrated in (2) (Trueswell et al., in press).

(2)a. The defendant examined by the lawyer was important.
  b. The evidence examined by the lawyer was important.

In (2), *examined* can be treated either as the main verb of the sentence or as part of a reduced relative clause modifying the preceding noun. If comprehenders adopt the main verb interpretation, they will display difficulty later, because *was* is in fact the main verb of the clause. In the reduced relative case, however, little difficulty should be encountered.

MacDonald et al. discussed three potential frequency biases relevant to the disambiguation of the MV/RR ambiguity, all associated with the verb: the frequencies of the verb's alternative argument structures, the frequency of the verb in active versus passive voice, and the frequency of the verb as a past tense versus as a past participle form. These frequencies are relevant because the reduced relative interpretation requires a past participle, passive voice, and an argument structure with a direct object; and thus to the degree that each of these alternatives is more frequent than its competitor(s), the reduced relative interpretation will be preferred. If instead competitors to these alternatives are more frequent, the main verb interpretation of the ambiguity will be preferred.

Certain contextual constraints have also been studied for these ambiguities. In (2), the contextual constraint comes from the plausibility of the initial noun as an Agent for *examined*: In (2a), *defendant* is a plausible Agent (defendants can examine things), but in (2b), *evidence* is inanimate and is an extremely implausible Agent. The only argument structure available for *examined* is that in (1a), so if *evidence* is not assigned the Agent role, it must be assigned the Theme role. However, this is only allowed by the grammar if *examined* is in passive voice, which will then force the reduced relative interpretation. Trueswell et al. (in press) found that subjects had less difficulty on items like (2b), containing helpful contextual biases, than on items like (2a), where the contextual help was not present. Thus we expect that if a model had access to information from unambiguous sentences about frequency and contextual biases relevant to the MV/RR ambiguity, it would display both effects of frequency and effects of contextual bias when presented with ambiguous sentences.

The Model

We chose a simple three-layer feedforward connectionist network to implement a part of the MacDonald et al. theory. Such networks are well-suited to performing constraint satisfaction tasks (e.g., Cottrell, Munro, & Zipser, 1987; McCelland, Rumelhart, & Hinton, 1986). The architecture of the model is shown in Figure 1. The model's task was to select an interpretation for each input token (sentence). The model was trained on unambiguous inputs to give it the opportunity to learn the constraints; the model was then tested on ambiguous versions of the input tokens to see what its interpretation preferences were in the face of the MV/RR ambiguity.

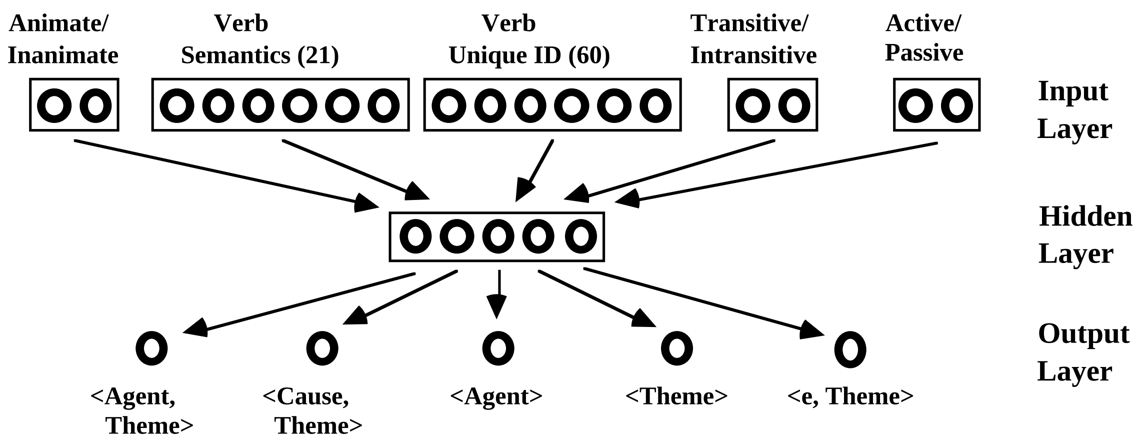

Figure 1: Schematic of the model. Arrows indicate feedforward connections.

Table 1: Example verbs and semantic features.

doze	bodily process	*glitter*	emission
	entity-specific		entity-specific
	manner		
		burn	motion
dissolve	motion		state
	state		resultant state
	resultant state		entity-specific
			emission
			existence

Input Representation

As shown in Figure 1, the 87 units in the input layer can be broken into five groups:

- Subject noun phrase animacy (2 units).
- Verb featural semantics (21 units), a simplified verb semantics with 21 features, based primarily on Levin (1989). Table 1 shows some examples of verbs from the training set with their semantic features.
- Verb unique identification (60 units). To differentiate verbs with identical semantic features, each verb had a unique pattern over these units. Patterns were generated randomly; each unit had a 25% chance of being set on.
- Transitive/intransitive (2 units), indicating the presence/absence of a direct object noun phrase.
- Active/passive (2 units), indicating the absence/presence of a form of *be* before the verb.

The units representing verb information (featural and unique ID units) had values of 1 or 0. For the other units (animacy, direct object presence/absence, and active/passive), the pattern (.6 .4) over two units represented one value (e.g., animate), and the pattern (.4 .6) represented the other value (e.g., inanimate). Values close together were chosen to force the system to rely more strongly on information about the verb in a sentence.

Output Representation

Each of the five units in the output layer represented a combination of argument structure and active/passive. Three of these units represented active voice and one of the argument structures in (1); one of the other units represented active voice and the <Cause, Theme> argument structure, which differs from (1a) only in that Causes are inanimate, whereas Agents are animate. The fifth unit in the output layer represented a passive interpretation, combining both passive voice and an argument structure containing a Theme direct object but no subject (shown as *e*). This last combination is required for the reduced relative interpretation, whereas the other four units correspond to various main verb interpretations. Thus when the input was ambiguous, the most highly activated output unit indicated the model's interpretation of the ambiguity.

Training

The model learned the constraints from a training set. Sixty past participle verbs were selected for this set, primarily from experiments involving the MV/RR ambiguity. All sentences containing these verbs were extracted from the *Wall Street Journal* (WSJ) electronic corpus, and each was coded for the animacy of the verb's subject, the presence of *be* in a passive, the presence of a direct object, and the argument structure of the verb. From these codings, a set of input tokens was created for each verb. Each input token consisted of a pattern on the input layer paired with a pattern on the output layer. The frequency of each input token in the set of WSJ sentences for the given verb was coded with the input token. From the 60 verbs, 176 input tokens were generated (only a few of the possible input tokens occurred for most verbs).

The 176 tokens were presented probabilistically (according to WSJ frequency) to the model in each epoch, and the model used a back-propagation learning algorithm to adjust its connection weights (initially set to small random values). The learning rate was set at 0.001 and the momentum value was 0.9. The model was trained for 5000 epochs. We ran four separate simulations with different initial random weights and different unique ID patterns; all results are averaged across the four simulations.

Testing the Training Set

Nothing about the model's learning of the constraints is theoretically crucial, but it is important that the model does actually learn on unambiguous inputs.[1] Thus after 5000 epochs of training, the model was tested on every token in the training set. The mean activation of the correct output unit for each token was .920, much greater than that for the next highest unit for each token, .061. Thus the model was able to learn the training set, and it developed strong, accurate preferences for unambiguous inputs.

Testing Ambiguous Inputs

To determine the model's response to ambiguous inputs corresponding to the MV/RR ambiguity, the training set tokens were used, but the input patterns for the units encoding active/passive and direct object presence/absence were set to (.4 .4) rather than (.6 .4) or (.4 .6). Thus the model no longer had any information about these properties of the input, and this should correspond to the point in human language processing when a string such as *The defendant examined* has been seen, but nothing is known yet about whether a direct object will follow, or whether *examined* is in active or passive voice. The activation levels of the argument structures in the output should therefore indicate the model's preferences, based on what it has learned from the (unambiguous) inputs about frequency- and context-based constraints.

Frequency-dependency. As a test of the model's learning of frequency constraints, we examined how the activation of each argument structure (output unit) for each verb was related to its frequency for that verb. Because the model was receiving no information about voice or the existence of a direct object, its responses depended on what it knew about argument structure preferences. Table 2 presents the correlation between frequency and activation across verbs for each argument structure. Activation is very closely related to frequency for all of the argument structures except <Agent, Theme>, indicating that the model was generally successful at using the argument structure frequency constraints in the input. The lack of effect in the latter case is under further investigation but probably resulted from many verbs (with a wide range of semantics) taking the <Agent, Theme> structure, with a very wide range of frequencies. Given the small number of hidden units in the model, it must rely on significant generalization across verbs, and in the <Agent, Theme> case, this is difficult, resulting in a loss of crucial information.

Verbs with a single argument structure. We also tested the model's knowledge of verb argument structure preferences by examining its responses to two particular classes of verbs. Ambiguous inputs for these verbs, although missing information about voice and presence of a direct object, should

Table 2: Correlation of argument structure frequency with activation across verbs.

Argument Structure	Frequency - Activation Correlation
<Agent, Theme>	.25
<Cause, Theme>	.80*
<Agent>	.97*
<Theme>	.68*
<e, Theme>	.75*

*$p < .0001$.

nevertheless be unambiguous, because these verbs (regardless of their environment) allowed only one argument structure. One of these classes, including verbs like *doze* and *glare*, permitted only the <Agent> argument structure and took only animate subjects. For these verbs, the mean activation of the <Agent> unit was .991; no other argument structure had an activation above .040. The second class of effectively unambiguous verbs included examples such as *disappear*, *sparkle*, and *wilt*. These verbs allowed only the <Theme> argument structure, but unlike the class of verbs like *doze*, they could take either animate or inanimate subjects. As in the case of verbs like *doze*, the activation of incorrect argument structures was always below .040 for these verbs. However, the activation of the correct <Theme> structure depended on the animacy of the subject: For animate subjects, its mean activation was .731, but for inanimate subjects, its mean activation was .963. Thus the model was quite effective at keeping track of frequency-based argument structure preferences for these verbs, but it also showed some sensitivity to animacy (a context-based constraint).

Fully ambiguous verbs. Finally, to see the model's preferences in fully ambiguous cases, which should correspond to typical cases of the MV/RR ambiguity, we examined performance on tokens for verbs which allowed multiple argument structures (i.e., those not included in the above two classes). The mean activation across verbs for all argument structures is shown in Figure 2, which shows that in addition to argument structure frequency constraints, the model was able to rely on constraints from animacy: When the subject was animate, the model preferred argument structures which could take animate subjects; when the subject was inanimate, the model's preferences changed appropriately to reflect this. Furthermore, for inanimate subjects, the model strongly preferred the passive argument structure, reflecting the selection of the reduced relative interpretation of the ambiguity, replicating Trueswell et al. (in press). This preference is not present when the subject is animate.

Discussion

The network model we have presented was trained on a reasonably realistic sample of English written text and was able to extract useful information about lexical frequencies. Testing on ambiguous inputs indicated that the constraints could be applied to generate predictions about preferences which

[1] In fact, the model will not be able to precisely learn the training set, because for five of the verbs, the same input pattern was trained to two different output patterns. For example, a token like *The passengers landed safely* would be coded as a <Theme> argument structure, whereas a token like *The pilot landed gingerly* would be coded as an <Agent> structure; but in the model's input, the subject noun phrase was identified only in terms of its animacy, so the two inputs appeared the same.

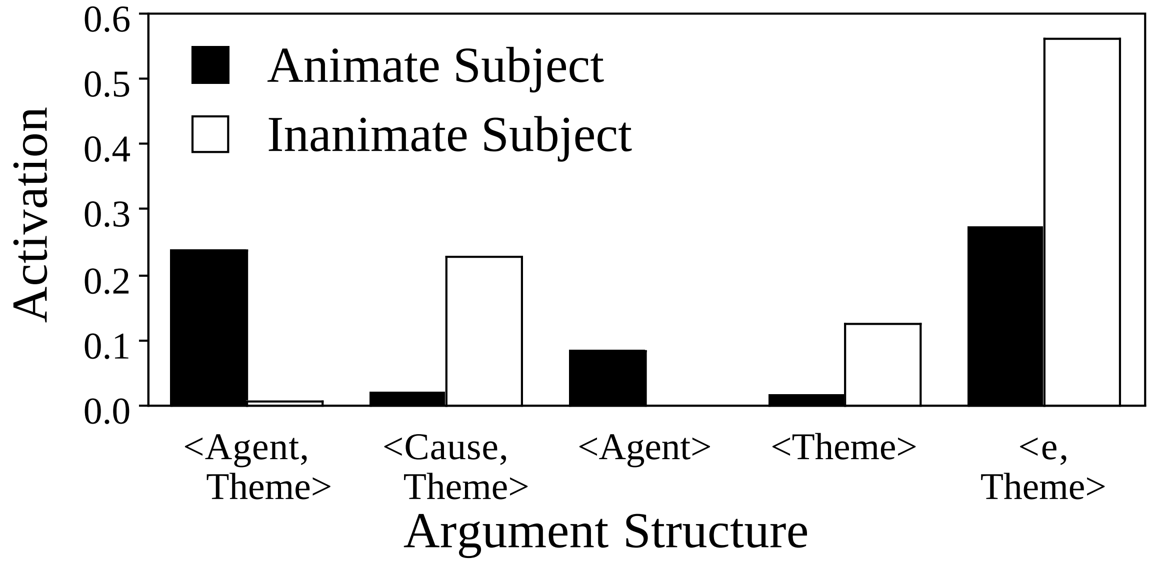

Figure 2: Argument structure preferences for ambiguous verbs.

supported those of MacDonald et al. (1993). The model also simulated the primary result in Trueswell et al. (in press): that the presence of an inanimate subject can create a preference for the reduced relative interpretation of the MV/RR ambiguity. However, the model had some difficulty handling the <Agent, Theme> argument structure: It should have been more strongly preferred for the ambiguous inputs with animate subjects, and its activation failed to correlate with its frequency across verbs. Part of the problem may be that this argument structure is not as frequent in our data set as in English. In addition, because it is so common, the semantics of verbs which allow it do not necessarily share many semantic features, and with only a limited number of verbs, it may be difficult for the model to learn much about it and to generalize about the verbs which allow or prefer it. We are currently doubling the number of verbs in the model, and this will increase the relative number of tokens involving this argument structure. This ongoing work will also further develop the verb semantics and investigate a wider range of ambiguities and constraints.

With respect to additional constraints in particular, it will eventually become important to increase the contribution of noun semantics in the model. As noted above, coding only animacy already creates minor problems, in that, for some verbs (e.g., *landed*), more complex properties of the subject determine the choice of thematic role and thus the choice of argument structure (*The passengers landed* vs. *The pilot landed*), so the model is trained to map the same input to two possible outputs. However, there is a more general issue to be considered, which is how more complex contextual constraints are to be represented in the model. At present, contexts (animacy) are treated as higher-order or more subtle cases of frequency effects in the model, but whether such a strategy can be extended to more complex contexts is an open question (cf. Pearlmutter & MacDonald, 1992; 1994). The strategy depends on being able to isolate specific features into which contexts can be decomposed, but as suggested by the *pilot/passengers landed* contrast, this may not always be possible, at least if the features are to remain applicable to more than a single lexical item (or noun-verb pair).

One alternative to this approach of identifying the relevant features in advance is to allow the model to develop representations it needs on its own. St. John and McClelland (1990) created a model of this sort by training a recurrent connectionist network with distributed representations to fill in thematic role information about sentences. Although the domain was quite limited, the model's interpretations eventually reflected the relative frequency, and, approximately, the relative plausibility of various competing thematic role assignments. Whether this type of model could also display sensitivity to syntactic constraints and an interaction between syntactic and semantic constraints, in a manner anything like humans (MacDonald et al., 1993; Trueswell et al., in press) is unclear: Connectionist models as a class continue to have difficulty handling structured representations of the sort required to state syntactic constraints (but cf. Elman, 1990). Part of our ongoing work related to the current model is an attempt to develop a framework combining the strength of connectionist models — simple, uniform processes which interact to create complex behavior — with the potential explanatory power of structured syntactic representations and theories.

Acknowledgements

Much of the work reported here was supported by an NSF pre-doctoral fellowship (NJP), a Hughes Aircraft Company Doctoral Fellowship (KGD), NSF Grant DBS-9120415 and a grant from the USC Zumberge Research Innovation Fund (MCM), and NIMH Grant MH47566 (MSS). We greatly appreciate the comments of three anonymous reviewers and the help of Josh Butler and Steve Kedrowski in coding WSJ tokens.

References

Cottrell, G. W., Munro, P., & Zipser, D. (1987). Learning internal representations from gray-scale images: An example of extensional programming. *Proceedings of the 9th Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, *27*, 429–446.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179–211.

Frazier, L. (1979). *On comprehending sentences: Syntactic parsing strategies*. Bloomington, IN: Indiana University Linguistics Club.

Juliano, C., & Tanenhaus, M. K. (1993). Contingent frequency effects in syntactic ambiguity resolution. *Proceedings of the 15th Annual Conference of the Cognitive Science Society*.

Levin, B. (1989). *English verbal diathesis* (Lexicon Project Working Paper No. 32). Cambridge, MA: MIT, Center for Cognitive Science.

MacDonald, M. C. (in press). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1993). The lexical nature of syntactic ambiguity resolution. Beckman Institute Technical Report UIUC-BI-CS-93-13 (Language Series), University of Illinois.

McClelland, J. L., Rumelhart, D. E., & Hinton, G. E. (1986). The appeal of parallel distributed processing. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing*. Cambridge, MA: MIT Press.

Pearlmutter, N. J., & MacDonald, M. C. (1992). Plausibility effects in syntactic ambiguity resolution. *Proceedings of the 14th Annual Meeting of the Cognitive Science Society*, 498–503.

Pearlmutter, N. J., & MacDonald, M. C. (1994). Individual differences and probabilistic constraints in syntactic ambiguity resolution. Manuscript submitted for publication.

Pinker, S. (1989). *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: MIT Press.

Simpson, G. B., & Krueger, M. A. (1991). Selective access of homograph meanings in sentence context. *Journal of Memory and Language*, *30*, 627–643.

St. John, M., & McClelland, J. L. (1990). Learning and applying contextual constraints in sentence comprehension. *Artificial Intelligence*, *46*, 217–257.

Tabossi, P., & Zardon, F. (1993). Processing ambiguous words in context. *Journal of Memory and Language*, *32*, 359–372.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (in press). Semantic influences on parsing: Use of thematic role information in syntactic disambiguation. *Journal of Memory and Language*.

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 528–553.