CHAPTER

10

Learning and Losing Syntax: Practice Makes Perfect and Frequency Builds Fortitude

Mark F. St. John
Pacific Science and Engineering Group, Inc.

Morton Ann Gernsbacher
University of Wisconsin–Madison

ABSTRACT

Passive and cleft-object constructions are harder to comprehend and break down more easily under stress or brain damage than active and cleft-subject constructions. We contend the reason is that they are simply less frequent and therefore less well practiced. We model this frequency effect using a simple recurrent network architecture (St. John & McClelland, 1990). The model was trained on four sentence constructions (simple active, simple passive, cleft-subject, and cleft-object), with one construction trained more frequently than the others. Generalization to new sentences was high, demonstrating mastery of the syntax rather than memorization of the training instances. The high-frequency construction was mastered first and proved more robust under simulated brain damage. Frequency is a powerful phenomenon in other areas of cognition; we demonstrate its role in language learning.

Some types of sentences are harder than others. Passive-voice sentences (e.g., *The doctor was admired by the lawyer*) are learned later than active-voice sentences (e.g., *The lawyer admired the doctor*; de Villiers, 1985; Horgan, 1978) and break down more severely due to brain damage (e.g., Caplan & Futter, 1986; Schwartz, Saffran, & Marin, 1980) or stressful or noisy environments during normal comprehension (Kilborn, 1991). Similarly, object-relative clauses (e.g., *The doctor that the lawyer admired*) are harder than subject-relative clauses (e.g., *The lawyer that admired the doctor*), center-embedded sentences (e.g., *The lawyer that advised the scientist admired the doctor*) are harder than right-branching sentences (e.g., *The lawyer admired the doctor that the scientist advised*), and right-branching sentences are harder than conjoined sentences (e.g., *The lawyer admired the doctor and advised the scientist*), both to learn (de Villiers, 1985; Ford, 1983) and to comprehend following brain damage (Bates, Friederici, & Wulfeck, 1987; Caplan, Baker, & Dehaut, 1985; Caplan & Futter, 1986) or stress (Miyake, Carpenter, & Just, 1994). What accounts for these differences in speed of acquisition in children, ease of comprehension in normal adults, and errors of comprehension in aphasics and stressed adults?

Researchers have suggested several factors that make some syntactic constructions more difficult to comprehend than others. Sentences containing more noun phrases are presumed to be more difficult than those containing fewer noun phrases. For example, sentences containing a relative clause have three noun phrases, whereas simple, one-clause sentences usually contain only two noun phrases; therefore, sentences containing relative clauses should be more difficult to comprehend than sentences containing only one clause. Sentences containing two verbs are presumed to be more difficult than sentences containing only one verb. A third factor is noncanonical word order. The typical (some would say "preferred") word order for English sentences is subject-verb-object (SVO), as represented in active-voice sentences and subject-relative clauses. In contrast, passive-voice sentences and object-relative clauses present their constituents in noncanonical word order.

Structuralists, like Grodzinsky (1986) and Caplan (Caplan & Futter, 1986; Caplan & Hildebrandt, 1988), argue that constructions with noncanonical word orders are harder to comprehend and more prone to break down because they require special processes to compute an underlying normal form. For example, in terms of trace and gap-filling models (Grodzinsky, 1986; Zurif & Swinney, 1994), the passive-voice construction contains a gap where the object used to be before it was moved to the front of the sentence (e.g., *The lawyer was admired [gap] by the doctor*). A special process is called upon to fill the gap by locating and assigning the noun phrase that fills it. In normal adults, this gap-filling process adds time to the comprehension process. In aphasics, the gap-filling process may be degraded so that passive sentences cannot be computed correctly. A slightly more complex explanation holds for the difference between cleft-subject constructions (e.g., *It was the doctor that [gap] admired the lawyer*) and cleft-object constructions (e.g., *It was the lawyer that the doctor admired [gap]*). Though both constructions contain a gap, in the cleft-subject construction the gap-filling process is trivial because the filler is immediately available, whereas in the cleft-object construction the gap-filling process is more elaborate because the filler resides in a distant part of the parse tree.

What differs among structuralists is whether the brain damage in aphasia degrades these extra processes directly or indirectly. According to physical structuralists like Caplan and Hildebrandt (1988), the damage is direct. Syntactic parsing is accomplished through a constellation of very specific parsing processes distributed around the pari-sylvian fissure between the temporal and frontal lobes. Caplan and Hildebrandt (1988) report single-subject dissociations between filling gaps in subordinate clauses and co-indexing pronouns and reflexives with their referents, between co-indexing pronouns and co-indexing reflexives, and between filling gaps in subordinate clauses (NP-trace) and filling gaps in *wh*-questions (*wh*-trace, as in *What did Johnny eat [gap]?*). The implication is that these specific deficits arise from damage to specific brain areas.

Resource structuralists (e.g., Kolk & van Grunsven, 1985; Linebarger, Schwartz, & Saffran, 1983; Miyake et al., 1994; Zurif & Swinney, 1994), on the other hand, posit a breakdown or decrement of a generic linguistic workspace. These researchers propose that certain constructions, such as the noncanonical passive construction, require more storage, workspace, or time for comprehension than other constructions, such as the canonical active construction. Damage in aphasia reduces this capacity or slows processing to the degree that there is no longer enough space or time to compute the extra processes required to comprehend some constructions. For example, there may not be enough time to perform the gap-filling process needed to comprehend a passive. One piece of evidence in line with this view is Linebarger et al.'s (1983) finding that Broca's aphasics could still perform grammaticality judgments. If we assume that judging grammaticality in a way that is not time pressured is easier than comprehending sentences, often in a time-pressured way, then a decrement in a general resource might well disrupt time-oriented comprehension but allow grammaticality judgment to proceed.

A number of researchers have expanded this view by suggesting that linguistic processing capacity could be exceeded and comprehension fail not only because of brain damage but because of added stress to the processing system, such as an overriding memory load, noise in the environment, or a very rapid presentation. Miyake et al. (1994) presented written sentences very rapidly to normal adults and observed that comprehension broke down more on the more complex sentences. They argued that the very rapid presentation rate precluded subjects from fully processing each word of the sentences. In effect, the speeded sentences overwhelmed the subjects' processing capacities. The result was a comprehension deficit that grew worse as the constructions grew more difficult, just as found for aphasics (Caplan & Hildebrandt, 1988). Most relevant to our present concerns, noncanonical word order constructions, such as passives, were more vulnerable to disruption than comparable canonical word order constructions, such as actives (see also Kilborn, 1991; King & Just, 1991; MacDonald, Just, & Carpenter, 1992).

THE ROLE OF FREQUENCY

In each of the preceding theories, there is no mention of frequency, and yet frequency plays a key role in human memory, skill acquisition, and learning. Moreover, many aspects of language processing are affected by frequency. For example, in word recognition, the frequency with which a word has previously been read or heard greatly affects how rapidly that word can be recognized: More-frequent words are read and understood considerably faster and more accurately than less-frequent words (e.g., Forster & Chambers, 1973; Gernsbacher, 1984; Just & Carpenter, 1980). Frequency also plays an interesting role in the time to name regularly spelled words versus irregularly spelled words. Regularly spelled words can be named faster and more accurately than irregularly spelled words. However, this effect pertains to only lower frequency words; the effect of regular spelling disappears for high-frequency, well-practiced words (see Seidenberg & McClelland, 1989).

At the level of sentence comprehension, the frequency of occurrence of words in different constructions affects processing (see Mitchell, 1994, and MacDonald, Pearlmutter, & Seidenberg, 1994, for reviews). For instance, MacDonald et al. (1994) reviewed research that shows that the frequency with which verbs occur in different constructions affects parsing decisions. These effects show up most clearly in sentences containing syntactic ambiguities. For instance, some verbs, like *raced,* are more frequently active than passive and more frequently intransitive than transitive. These frequencies make sentences in which the verbs occur in passive, transitive constructions difficult to parse correctly, as illustrated by the classic, *The horse raced past the barn fell.* However, other verbs, like *carried,* are more balanced in their frequency of occurring in active versus passive sentences, making a passive, transitive interpretation such as *The child carried to safety cried* relatively easy to comprehend. The implication is that the frequency with which specific verbs occur in different constructions is collected and utilized during early parsing processes.

Most relevant to the topic of the current volume, we propose that the effects of frequency derive from actual practice on individual forms. A central-processing-capacity theory might account for frequency effects by claiming that lower frequency words or even lower frequency constructions consume more capacity than do higher frequency words or constructions; when capacity is exceeded, processing slows (or even breaks) down.

An alternative conception, and the one that we offer here, is that frequency of practice directly affects acquisition, processing speed, and vulnerability—in other words, frequency of practice directly affects syntactic learning, use, and loss. For word recognition and lexical access, we generally think of a distributed memory where practice on individual words makes them individually faster and easier to process. For sentence comprehension, practice on individual verbs makes them suggest some syntactic roles more than others. Similarly, might not the frequency of individual constructions, such as actives and passives, affect the speed and ease of their acquisition and processing, and their vulnerability to loss? If so, the effects of speeded presentation, noise, memory load, or brain damage might not result from the loss of special processes or the diminution of central capacity. Instead, less-frequent constructions would be weaker and therefore more vulnerable to loss. Again, recent work in word reading (e.g., Hinton & Shallice, 1991; Marchman, 1993) provides an analogy by showing the effects of word frequency on performance following real and simulated brain damage. Performance on frequent forms was relatively well preserved, whereas performance on less-frequent forms was disrupted.

Frequency, of course, cannot account for all syntactic differences. The amount of information contained in a sentence surely affects comprehension. Sentences containing multiple propositions, such as center-embedded and right-branching clauses, are more difficult than simple one-clause constructions. Sentence length must also play an obvious role in the time and processing that it takes to read a sentence and in the amount of storage required to remember and respond to a sentence. Exploring the role that sentence length plays in learning and loss was another goal of the research we present in this chapter.

The major goal of the research presented in this chapter was to demonstrate the role that the often neglected factor of frequency plays in syntactic learning and loss. Our strategy was first to marshal evidence of frequency differences and processing differences between actives and passives and second to demonstrate, using a computational model, how a difference in frequency can produce a difference in learning and loss. Of course, computational modeling cannot prove that people work this way, but it can underscore the coherence and plausibility of a theory.

A PARALLEL DISTRIBUTED PROCESSING MODEL

Our model is a parallel distributed processing neural network. Although it is unclear how symbolic models could account for effects of frequency in normal or abnormal processing, connectionist models are designed to produce frequency effects. Connectionist models learn from experience and produce stronger weights with experience. Strong weights in turn process stimuli faster and more reliably, and strong weights are more invulnerable to damage.

In our model, the words of an input sentence are presented sequentially over time to a simple recurrent network (Elman, 1990; St. John & McClelland, 1990). The model is probed with comprehension questions asking, *Who did what to whom?* for each sentence. The model is trained to produce the correct responses to these questions. Critically, the model is not provided with any predetermined linguistic representations. The model is required to learn whatever representations and processing that it requires to compute correct answers. The model simply learns to map surface forms to meanings, and more-frequent mappings gain more strength (Bates & MacWhinney, 1987). The model allows us to investigate the learning, comprehension, and loss of comprehension of various syntactic constructions.

Materials

In the work reported here, we focused on the four single-clause constructions employed by Miyake et al. (1994) and illustrated in Table 10.1. Sentences of each construction type were composed from 20 possible participants (e.g., *scientist* and *governor*) for the first noun, 15 possible verbs (e.g., *advised* and *admired*), and 19 remaining possible participants for the second noun. Thus, for all four constructions, there were 22,800 ($4 \times 20 \times 15 \times 19$) sentences. The four syntactic constructions were assigned various frequencies of occurrence during training depending on the specific simulation.

The model was asked questions about who was the agent or patient of the action in each training sentence. For example, for the sentence, *It was the banker that the lawyer advised,* the questions were "advised-agent?" ("Who advised?") and "advised-patient?" ("Who was advised?"). We believe these are plausible sorts of questions and therefore a plausible training task because the questions match the gist of the comprehension questions employed by Miyake et al. (1994) and the questions represent the sort of information a reader or listener is likely to obtain from a sentence.

TABLE 10.1
Syntactic Constructions Used in Simulation

Syntactic Construction	*Example*
Simple-Active	The scientist advised the governor.
Simple-Passive	The doctor was admired by the lawyer.
Cleft-Subject	It was the banker that praised the singer.
Cleft-Object	It was the dancer that the artist insulted.

Procedure

On each training trial, a syntactic construction was selected according to its assigned frequency, and a sentence of that type was generated randomly. Each word of the sentence was presented to the network one at a time. After each word, the model was asked one question about the agent or the patient of the action in the sentence. The model's response was compared to the correct response (i.e., target), and the errors were used to modify the model. Table 10.2 shows the steps for comprehending (and training) an example sentence.

Of course, early in a sentence, the model may not have the information needed to answer one or more of the potential questions correctly. At such points, the model minimizes its error by activating all potential answers according to their probability of being correct. At the beginning of a sentence, this probability is just the a priori probability of the occurrence of each participant. As more of a sentence is presented, more information is available to constrain the probabilities and allow better guesses. By the end of a sentence, enough information is known to answer all questions with certainty. This training procedure forces the network to eke out as much information as it can from each word of a sentence to learn the probabilities and minimize its guessing errors.

Network

The network, which is illustrated in Fig. 10.1, is essentially a simple recurrent network with some additional architecture to accomplish the question asking and answering training procedure. The input layer consists of three parts: word units, the recurrent context units, and the comprehension question units. One unit is used to represent each possible word, including a unit for "period." No semantic or syntactic features of any words are represented. Any knowledge of words, word class, and syntax is acquired through

TABLE 10.2
Processing a Simple-Active Sentence

			Network Output			
Word	*Question*	*Target*	*Scientist*	*Governor*	*Lawyer*	*Doctor*
the	advised-agent	scientist	.05	.05	.05	.05
scientist	advised-agent	scientist	.50	.03	.03	.03
advised	advised-agent	scientist	1.00	.00	.00	.00
the	advised-patient	governor	.00	.05	.05	.05
governor	advised-patient	governor	.00	1.00	.00	.00
period	advised-agent	scientist	1.00	0.00	.00	.00

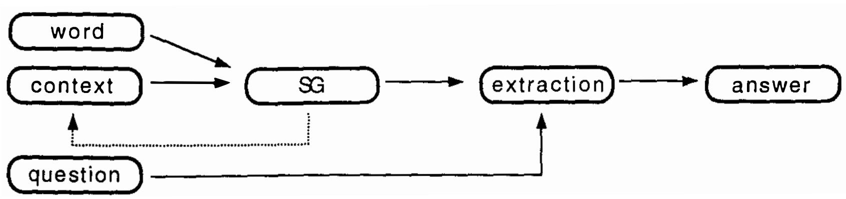

FIG. 10.1. Architecture of the network. SG = sentence gestalt.

training as the model discovers the effects of each word on sentence interpretation. The question units encode the question that is randomly picked to be asked at that point in the sentence (e.g., *admired-agent*). There is a unit for each possible verb and a unit for each thematic role (i.e., agent and patient). The output units represent the possible answers to the questions. There is one output (answer) unit for each participant (i.e., there are 20 output question-answering units).

To begin comprehending and training a sentence, the unit that corresponds to the first word of the sentence is activated in the word part of the input layer, the context units are set to 0.0, a question is generated, and the units that correspond to that question—a verb and a thematic role—are activated in the question part of the input layer. Activation from the word, context, and question feeds forward through the network to activate the sentence gestalt (SG) layer. The SG layer is a set of hidden units (because they lie between the input and output layers). Their purpose is to represent the meaning of the sentence as a whole as it develops (i.e., the sentence is input and processed). Activation then feeds forward from the SG layer to the output layer. The network's output is compared to the correct question answer, and error is used to change the weights of the model via back propagation (Rumelhart, Hinton, & Williams, 1986). The cross-entropy error measure (Hinton, 1989) is used because its optimum is the probability of a particular answer given a particular sequence of inputs.

The word units and question units are then reset to zero, and the second word of the sentence is processed. The appropriate word unit is activated, a new question is activated, and the activation values from the sentence gestalt layer are copied over to the context units. This copy action creates a recurrent loop in the network and thereby allows the network to maintain an internal representation of the developing sentence meaning as each word is processed in turn. Activation feeds forward again through the network, the new output is compared to the new question answer, and any error is back propagated. The model continues through the sentence word by word.

For all simulations, the model had 80 hidden units in the sentence gestalt layer and the extraction layer. The learning rate was 0.02, and the momentum was 0.0. These parameters modulate the size of weight changes as the network learns. Correct answers, the targets used for training, were set to

1.0, and incorrect answers were set to 0.0. All network weights were initialized with small random values.

Brain damage (or stress) was modeled using McClelland's (1993) noisy neuron procedure. Under brain damage or stress, the activation value of each sentence gestalt unit is perturbed by a small random number on each feed-forward cycle. A random number, drawn from a uniform distribution centered at 0.0, is added to the net input of each unit.[1] The upper and lower bounds of the distribution can be manipulated. Larger bounds cause more damage because larger random numbers more strongly perturb the activation of each unit.

The noisy neuron procedure, rather than random removal of weights (e.g., Hinton & Shallice, 1991; Marchman, 1993), was used here because of its closer analogy with stress, either memory load or speeded presentation. The idea is that neural representations are always somewhat noisy and that brain damage exacerbates this condition. The analogy between the noisy neuron model and Miyake et al.'s (1994) speeded presentation manipulation is that if neural processes are noisy, more time might be required to establish clean and strong representations. Under speeded processing conditions, there is little time to clean up and strengthen inherently noisy representations.

Results

To begin, we illustrate the processing of a simple active sentence. Following this illustration we consider the effects of sentence length, syntactic construction frequency, and simulated brain damage on learning and comprehension.

Active Sentence Illustration. For this illustration, we trained the network with equal frequencies of all four syntactic constructions; then we presented a simple active sentence for processing. Table 10.2 shows the ideal activations of a selection of output units as each word of the sentence is processed (the trained model comes very close to matching these values). Toward the beginning of the sentence, the model does not have enough information to answer any question correctly; activation is distributed equally lightly across all 20 participants. Even after the first noun has been processed, the model

[1]The net input is the sum of products of weights and activations leading into each unit. The net input value is then turned into an activation value between 0.0 and 1.0 by applying the logistic activation function

$$net_i = \Sigma_j W_{ij}A_j \qquad activation_i = \frac{1}{1 + e^{-net_i}}$$

where W_{ij} is the weight to unit i from unit j, and A_j is the activation of unit j.

does not know whether that participant is the agent or patient of the sentence because both active and passive constructions are equally frequent in this particular training corpus. The model's response to the agent question at this point is to activate "scientist" 0.5 because the model knows that "scientist" will be either the agent or the patient; it activates the other possible participants equally lightly.

As more words are processed, the model is able to answer more questions correctly. For example, after the verb is presented, because the verb was not preceded by "was," the model knows the sentence is in active voice, and it can then answer questions about the agent of the sentence. Similarly, if "was" had preceded the verb, the model would know that the sentence was in passive voice and would be able to answer the patient question correctly. Similar clues apply to each sentence type. After the second noun phrase, the model can answer questions about the patient role as well. At the end of the sentence, to answer both questions correctly, the model must be maintaining information about early sentence constituents although they are now absent from the input; the information is maintained internally in the sentence gestalt representation.

Sentence Length. Active-voice sentences, in addition to being more frequent than passive-voice sentences, are shorter. Does this difference in length also affect training time; does the longer passive-voice construction require more training than the active-voice construction, regardless of frequency? To answer these questions, we trained three identical models on three different length constructions. Each model was trained on only one construction, and we compared the number of training trials of the three models to reach 100% correct performance. One model was trained on the simple active construction, which contained five words (e.g., *The scientist advised the governor*). An identical model was trained on the simple passive construction, which contained seven words (e.g., *The governor was advised by the scientist*). A third model was trained on a third syntactic construction we created for this experiment, the passive + 2 construction. The passive + 2 construction was a normal, simple passive construction with two additional syntactically neutral words added to make the construction nine words long (e.g., *The governor <u>from California</u> was advised by the scientist*). The two extra words were always positioned after the first noun phrase, and they were the same for each sentence. Therefore, all three syntactic constructions were identical in terms of number of participants and the complexity of the mapping from surface form to meaning.

The experiment was repeated five times using models starting from different initial random weight values. Training continued until a model achieved correct performance on 500 consecutive, randomly generated sentences. Correct performance was defined as answering the sentence-final question correctly, that is, activating the correct output unit more strongly than every other output unit for the sentence-final question. All $20 \times 15 \times 19$ (5,700) sentences for a syntactic construction were used during training.

Averaged over the five repetitions, the five-word construction reached mastery after 146,000 sentence training trials. The seven-word construction reached mastery after 177,000 training trials, and the nine-word construction reached mastery after 261,000 training trials. The five-word construction trained significantly faster than the seven-word construction, $F(1, 4) = 9.6$, $p < .04$, and the seven-word construction trained significantly faster than the nine-word construction, $F(1, 4) = 83.9$, $p < .001$. In sum, longer syntactic constructions required more training despite their similarity in terms of the number of participants and complexity.

Generalization Performance. Next, we trained a fresh model on all four syntactic constructions (simple active, simple passive, cleft-subject, and cleft-object) with equal frequencies simultaneously to demonstrate that the network could learn all four syntactic constructions simultaneously and generalize to novel sentences. For each of the four syntactic constructions, two different participants and two different verbs were set aside from training for later generalization testing. This situation left 18 possibilities for the first participant, 17 remaining possibilities for the second participant, and 13 possibilities for the verb, resulting in a total of 15,912 possible training sentences, or 70% of the total corpus.

The network mastered the training corpus after 311,000 trials, where mastery was again defined as correctly answering the sentence-final question for 500 consecutive, randomly generated sentences. A second network, starting from different random initial weight values, mastered the training corpus after 263,000 trials. Interestingly, these numbers are substantially lower than the sum of training times for each of the individual syntactic constructions; the number of training trials for the short simple active construction alone was 146,000. Consequently, there was a mutual benefit to training all four syntactic constructions together, perhaps in part because they share the same participants and verbs.

Table 10.3 shows the generalization performance for both networks (Network 1 and Network 2) on each syntactic construction following 400,000

TABLE 10.3
Generalization Performance, Equal Frequency Corpus

Network	*Simple Active*	*Simple Passive*	*Cleft Subject*	*Cleft Object*
1	95 +/- 3	99 +/- 2	90 +/- 3	92 +/- 2
2	93 +/- 3	94 +/- 3	94 +/- 2	92 +/- 2

training trials. Every syntactic construction showed substantial, although not perfect, generalization. Generalization was tested by presenting sentences generated from the entire corpus, including both trained and untrained sentences. Because the training sentences represent 70% of the total corpus of sentences, correct performance above 70% represents generalization. Each cell shows the average percentage correct and standard deviation for five sets of 100 sentences.

Frequency. What happens when the training frequencies are not equal? New simulations made either the simple active or the simple passive construction eight times more frequent during training than the other three constructions. It is difficult to know what ratio best matches reality. One estimate (Goldman-Eisler & Cohen, 1970) held that the ratio of active to passive constructions in oral English lay between 35:1 and 10:1, depending on the discourse context. Conversations produced the highest ratios, and speeches in the House of Commons produced the lowest ratios. Ratios in written text have been found to be lower, for example, about 6:1 (Taylor & Taylor, 1983). We chose a ratio of 8:1 as a ballpark figure and as a reasonably high ratio. We also report simulations with 24:1 and 1:1 ratios for comparison (see Table 10.7).

By our analysis, simple active and simple passive sentences are similarly complex. They both contain two participants mapped to two thematic roles. Our point is that differences in frequency (and to some extent length) cause the difference in training time and comprehension. To underscore this point, we first discuss the simulation in which the simple passive was the high-frequency construction. Separate simulations in which the simple active was the high-frequency construction showed completely comparable results (see Table 10.7 on p. 245).

Again, in the simulations two different participants and verbs for each syntactic construction were set aside for generalization testing. Training sentences were generated randomly with the constraint that the passive construction occur 8 times more frequently than each of the other three constructions. The simulation mastered the whole training corpus after 514,000 training trials. Therefore, this unequal frequency simulation required more training trials to reach mastery than the equal simulation. The reason, of course, is that most of the training trials are taken up with the frequent passive construction, leaving few training trials for the less-frequent constructions.

Interestingly, however, the low-frequency constructions receive a large boost from the high-frequency passive construction. Whereas in the equal-frequency simulation, each construction reached mastery-level performance after roughly 77,800 training trials, in the unequal-frequency simulation, the low-frequency constructions reached mastery-level performance after

roughly 46,700 training trials. In fact, the low-frequency constructions derived a benefit from the high-frequency passive construction, again at least in part due to sharing participants and verbs.

Table 10.4 shows the amount of generalization at four points during training. For each cell in the table, five sets of 100 random sentences were tested, and the average percentage correct and standard deviations are reported. Critically, the high-frequency passive-voice construction was mastered first. In the end, all four syntactic constructions were mastered, generalization was very high, and the effect of frequency was hidden. However, under noise (our simulation of loss) the effect of frequency returned.

The effects of noisy neurons on performance, to simulate brain damage or stress, are shown in Table 10.5. We tested the fully trained (800,000 training trials) network where the passive construction was trained 8 times more frequently than the other three syntactic constructions. We randomly generated test sentences from the four syntactic constructions, and as each word of a test sentence was processed, uniformly random noise was added to the sentence gestalt units. Noise varied between −1.0 and +1.0 or between −1.5 and +1.5. Correct comprehension was again defined as a correct answer to the sentence-final question. Each cell in Table 10.5 shows the average percentage correct and standard deviation for five sets of 100 sentences for a given syntactic construction. No generalization sentences were included in these test sets; we only examined sentences on which the network had been trained.

Overall, greater noise damage produced more breakdown; for each syntactic construction, correct performance dropped when the noise level was

TABLE 10.4
Generalization Performance, Frequent Passive Corpus

Training	*Simple Active*	*Simple Passive*	*Cleft Subject*	*Cleft Object*
200k	62 +/- 6	85 +/- 6	56 +/- 3	42 +/- 3
400k	92 +/- 1	91 +/- 2	81 +/- 3	72 +/- 4
600k	87 +/- 4	94 +/- 1	88 +/- 3	88 +/- 2
800k	98 +/- 1	98 +/- 2	96 +/- 3	91 +/- 3

TABLE 10.5
Performance Under Noise, Frequent Passive Corpus

Noise Level	*Simple Active*	*Simple Passive*	*Cleft Subject*	*Cleft Object*
-1.0 to +1.0	93 +/- 2	95 +/- 1	82 +/- 2	84 +/- 2
-1.5 to +1.5	69 +/- 3	77 +/- 4	64 +/- 3	66 +/- 3

increased from the −1.0 to +1.0 range to the −1.5 to +1.5 range. Most importantly, the more-frequent passive-voice construction was preserved relative to the less-frequent constructions. Further simulations showed that higher training ratios of frequent:infrequent constructions produced even larger differences between syntactic constructions in vulnerability to noise (see Table 10.7).

Both cleft constructions broke down worse under noise damage than the simple active or simple passive construction. The reason for the low performance of the cleft constructions is partly due to their length. Both cleft constructions are eight words long, whereas the simple active is only five and the simple passive is seven words long. We trained a fresh model on a modified corpus in which the cleft constructions were abbreviated to six words by removing the initial *it was* (e.g., *the scientist that advised the governor* for a cleft-subject and *the governor that the scientist advised* for a cleft-object). The simple active and simple passive constructions were not modified. All four syntactic constructions were trained with equal frequency. Following training, we evaluated the comprehension performance of the model on all four syntactic constructions under both normal and noise damage conditions. The model performed as well on the abbreviated cleft-subject construction as on the simple active construction, but the model performed poorly on the abbreviated cleft-object construction. The unusual noun–noun–verb word order of the cleft-object construction makes it different from the other three constructions and may account for its slower learning and easier breakdown.

We also compared the effects of noise damage on comprehension on the equal frequency training simulations to the effects of noise damage on the frequent passive simulation. We tested the two fully trained equal-frequency simulations (Network 1 and Network 2) following 400,000 training trials on the same two levels of noise. The results are shown in Table 10.6. The pattern of vulnerability to noise damage was different: Most important, the simple passive construction was not better preserved than the equally frequent simple active construction. Clearly, the better preservation of the passive construction in the high-frequency passive simulation was due to its higher training frequency and greater practice in that simulation. Practice

TABLE 10.6
Performance Under Noise, Equal Frequency Corpus

Network	*Noise Level*	*Simple Active*	*Simple Passive*	*Cleft Subject*	*Cleft Object*
1	-1.0 to +1.0	91 +/- 1	94 +/- 2	86 +/- 5	86 +/- 2
1	-1.5 to +1.5	76 +/- 5	76 +/- 6	57 +/- 5	68 +/- 3
2	-1.0 to +1.0	95 +/- 1	92 +/- 3	83 +/- 3	83 +/- 5
2	-1.5 to +1.5	78 +/- 4	73 +/- 4	60 +/-11	62 +/- 4

enhances robustness to noise damage. The normal eight-word cleft constructions were again damaged more than the simple constructions. We trained two fresh networks, making the simple active construction eight times more frequent than each of the other three syntactic constructions. The results are reported in Table 10.7. With the simple active construction most frequent, the simple active construction reached mastery first. Again, after sufficient training, all four syntactic constructions were mastered and achieved high levels of generalization performance. Finally, under noise damage, the high-frequency simple active construction sentences were comprehended better than were the simple passive construction sentences.

We compared the effect of a variety of frequency training ratios on the number of training trials required to master the corpus and the difference in breakdown between the higher and lower frequency constructions. Larger ratios, such as 1:24 simple actives to simple passives, require more training trials and create larger differences in how much each construction breaks down under noise damage. Table 10.7 shows the number of training trials to reach 100% correct (i.e., mastery) performance for each simulation. Table 10.7 also shows the differences in comprehension under noise between the simple active and simple passive constructions. The differences were computed from the average percentage correct of 1,500 sentences of each syntactic construction. Two networks were trained at each ratio of frequencies using different starting weights and random orders of trained sentences. The average results of both networks for each ratio are reported along with the ratio of training.

The total amount of training affects the size of the frequency difference under noise damage. Early in training, before the less-frequent constructions are fully mastered, the effect of the training frequency ratio is large. As all the syntactic constructions are mastered and generalization surpasses 90%, the frequency effect decreases. As training continues further, the differences decrease further. The lower frequency construction eventually draws even with the higher frequency construction. Because training operates by error correction, weight changes occur only on errors. Consequently, as constructions are better mastered, fewer errors are made, and fewer and smaller

TABLE 10.7
Training and Breakdown Effects of Various Frequency Ratios

Active : Passive	*Trials to Mastery*	*Point of Evaluation*	*Noise Level*	*Percentage Correct Difference*
8:1	516k	600k	1.5	4.8 (active better)
1:1	287k	400k	1.5	2.1 (active better)
1:8	514k	800k	1.5	6.0 (passive better)
1:24	1141k	1600k	1.5	9.3 (passive better)

weight changes occur. Once the more-frequent construction is well mastered, the less-frequent construction can catch up. Such nonlinear effects of training are well known. For example, in a simulation of word recognition (Seidenberg & McClelland, 1989), the highest frequency items showed the smallest effects of spelling irregularity because they were all perfectly mastered.

What errors does the model make under noise damage? In the high-frequency passive simulation, approximately one fourth of the errors consisted of activating no participant at all in response to a sentence-final question. Another one fourth of the errors consisted of activating a participant not mentioned in the input sentence. Fully half the errors consisted of reversing the roles of the participants in the input sentence; the agent was assigned the patient role, and the patient was assigned the agent role. Furthermore, most errors occurred in low-frequency constructions. In other words, in simulations where the simple passive was most frequent, the most common error was to misassign the first noun in a simple active sentence to the patient role, as if it were a passive sentence. In simulations where the simple active was most frequent, the most common error was to misassign the first noun in a simple passive sentence to the agent role, as if it were an active sentence. We can think of this mismapping as a regularization error to fit a problem input sentence into the more-frequent construction's mapping of sentence constituents to agents and patients.

Interestingly, this role reversal is the most common error found in studies with aphasics. In these studies, patients typically hear a spoken sentence and are then required to point to one of several pictures depicting the relation described in the sentence. One picture depicts the correct relation, one picture depicts the reverse relation, and other pictures contain extraneous participants never mentioned in the spoken sentence. Patients rarely point to pictures depicting extraneous participants; instead, their errors mostly consist of pointing to the role-reversal picture. Furthermore, patients err more on low-frequency constructions, like the simple passive in English, than on high-frequency constructions, like the simple active.

A QUANTITATIVE COMPARISON TO APHASIC DATA

How well does the difference in breakdown between frequent and infrequent syntactic constructions in the simulation compare with data from aphasics? The simulation produces a difference of 5 to 10% between high- and low-frequency constructions, depending on the training ratio. The trace deletion hypothesis (Grodzinsky, 1986), however, predicts that simple active sentences should produce 100% correct responses, whereas simple passives should produce 50% correct responses. Essentially the theory predicts that

TABLE 10.8
Comprehension of Active and Passive Sentences

Study	*Subjects*	*Sentences*	*Active*	*Passive*	*Difference*
Goodglass et al., 1993	7	6	90%	80%	10%
Martin et al., 1989	4	16	80%	66%	14%
Caplan & Futter, 1986	1 (nf)[3]	6	100%	50%	50%
Caplan et al., 1985	37 (exp. 2)	5	88%	58%	30%
Pierce & Wagner, 1985	14 (10 nf, 4 f)	9	74%	63%	11%
Schwartz et al., 1980	5 (nf)	24	75%	51%	24%
Weighted ave.			81%	60%	21%

simple active sentences can be processed correctly using a default agent–verb–patient word ordering strategy. Under this strategy, however, simple passive sentences produce a conundrum because word order says the first noun is the agent whereas the *by*-phrase says the second noun is the agent. Given this conflicting information, aphasics should respond randomly (50% correct) on simple passive sentences. Yet, the simulation comes nowhere near producing a 50% difference between active and passive sentences.

The trace deletion hypothesis, however, is not supported by empirical results. Across a number of studies, the difference between aphasics' comprehension of simple active versus simple passive sentences averages 21%, not 50%. Table 10.8 reports the percentage correct on semantically reversible active and passive sentences—sentences where semantic information does not predict the assignment of agent or patient—for several studies.

Several studies reported in Table 10.8 combined scores from fluent and nonfluent patients (f and nf in Table 10.8), because the two groups perform quite similarly on this comprehension task. Weighted average comprehension scores and differences were calculated by weighing the percentage correct by the number of patients and the number of sentences in each study (Σ[#subjects $\times$ #sentences $\times$ score]).[2] The model's percentage correct difference is smaller than the moderate differences observed with aphasics. For the model, the difference is approximately 6% for a training ratio of 8:1 and 9% for a training ratio of 24:1. For aphasics, the difference is approximately 21%.

[2]For the Pierce and Wagner (1985) study, data were averaged from two conditions. Either the critical active or passive sentences were presented alone or they were preceded by a paragraph that introduced the characters in the critical sentence but did not bias the case-role assignments. Though this context manipulation enhanced comprehension, it did not interact with the type of syntactic construction; therefore, these conditions were combined for our purpose.

This discrepancy might be due to several factors. First, the discrepancy might simply be due to the small number of sentences used in some studies. In the model, any similarly small random sample of sentences might show a larger, or smaller, difference between comprehension of active and of passive sentences. Second, the discrepancy between the model and patient performance might be due to our poor estimate of the frequency difference between active and passive sentences in the language to which aphasics are, or were, exposed. The frequency of the passive construction varies across contexts. Although it is fairly common in academic writing, it is fairly rare in spoken conversation. Exact counts are difficult to obtain, and even if they were easily obtained, it is not clear which counts on which corpora are most appropriate for characterizing aphasics. For example, is overall frequency the critical count, or should spoken and written counts be separated? Finally, the experiences of different aphasics may vary substantially. It would be interesting, for example, to know whether patient comprehension performance varies with educational level.

In addition, a part of the discrepancy is very likely due to extralinguistic strategies used by patients. One such strategy that patients may use is to treat all sentences by default as the most frequent construction: the simple active. The strategy assigns the first noun of a sentence to the agent role and the first noun phrase after the verb to the patient role. This strategy can be valuable as a backup when the normal linguistic process fails to compute case-role assignments. In the simulation, roughly one fourth of the errors occur when no concept was activated in response to a case-role question. In these cases, the first-noun-agent strategy might be invoked. If the test sentence were active voice, this strategy would produce a correct response—a good bet in normal English spoken discourse. Using this strategy, therefore, can correct one fourth of the errors on the default (high-frequency) construction. On the other hand, if the test sentence were passive, this strategy would produce an error; the strategy chooses the incorrect interpretation, so these errors remain errors.

Another fourth of the simulation's errors comes from activating an incorrect, random participant in response to a question. In most aphasic studies, patients match a sentence to a picture. Because it is unlikely that a picture will correspond with the random participant generated by the linguistic process, the patient may again employ the first-noun-as-agent strategy. Using this strategy would boost the percentage correct score for the default construction by another one fourth of its error rate. For example, in the high-frequency passive simulation, the comprehension scores were 77% for the high-frequency construction and 69% for the low-frequency construction. The extralinguistic strategies would add 12% to the high-frequency construction score. The adjusted percentage correct scores would be 89% and 69%—a 20% difference. With these strategic corrections, the model would closely

match the effect size found for aphasic patients. Of course, the effect size in the model and presumably in individual aphasics will vary according to how thoroughly the extralinguistic strategies are applied, the severity of the brain damage (or the amount of noise damage in the model), and the training ratio of the constructions. The crucial point is that the model produces the same ballpark figures as aphasics.

How might these strategies be applied in the model? Because these are extralinguistic strategies, it is fair to place them outside the network. Perhaps some response-making process could examine the answers of the model. When no answer is activated or when an extraneous participant is activated, this response process could make the default thematic role assignments. Of course this process would have to know what the default was and which noun phrase preceded the verb, but these are simple matters.

THE FREQUENCY MECHANISM

What mechanism in the network is responsible for making the more-frequent syntactic construction more invulnerable to noise damage? Three mechanisms seem likely. First, frequency may produce bigger weights for the more-frequent construction, and bigger weights are more immune to small noise fluctuations. Second, frequency may produce more distributed representations for the more-frequent construction, and more distributed representations are more redundant and therefore more invulnerable to noise damage (French, 1992). Third, the more-frequent construction becomes a default mapping that can operate despite noise damage.

We tested these alternative frequency mechanisms by examining the fully trained high-frequency passive simulation. We examined the sentence-final sentence gestalt representations for 100 simple active sentences and 100 simple passive sentences. For each sentence, we calculated the number of units in the sentence gestalt representation with an activation of at least 0.3. Surprisingly, the simple active sentences (the lower frequency construction) had significantly more units active (27) than the simple passive sentences (25), $F(1, 198) = 74.4$, $p < .001$, and the units representing the simple active sentences had higher activations (on average, 0.91) than the units representing the simple passive sentences (on average, 0.88), $F(1, 198) = 26.3$, $p < .001$. Contrary to the first two hypothesized mechanisms, therefore, the low-frequency construction had more units active, and these units had higher activations.

To test the third hypothesis, that the more-frequent construction became a default, we performed a regression analysis on the ability of the sentence gestalt units to predict a simple active versus a simple passive construction. We examined the sentence-final activations of all 100 sentence gestalt units for 100 active and 100 passive sentences (i.e., we computed the ability of each unit to predict the voice of each sentence). The constant in the regression equation strongly predicted the high-frequency passive construction. It appears, then, that the high-frequency construction became the default mapping. The idea of a default is that the network weights are set up to produce a certain answer or compute a certain function, much like a bias term produces a certain default result. The extra units in the low-frequency active construction representation may be needed to override the default or bias.

What if no default is possible? To examine this situation we trained both the passive construction and the cleft-object construction with equally high frequency and the active construction and the cleft-subject construction with equally low frequency (8:8:1:1, respectively). Following training, we confirmed that the high-frequency constructions were more invulnerable to noise damage than the low-frequency constructions.

Then, we again examined the sentence gestalt representations for 100 sentences of each syntactic construction type. Now, both active- and passive-voice constructions activated the same number of units (passive = 23 units with activations above 0.3 vs. active = 23 units with activations above 0.3, $F(1, 198) = 0.2$, $p = 0.70$), but the units representing the high-frequency simple passive construction were more active than the units representing the low-frequency simple active construction (passive = average activation of 0.89 vs. active = average activation of 0.86, $F(1, 198) = 17.8$, $p < .001$). Similarly, units representing the high-frequency cleft-object construction were more active than the units representing the low-frequency cleft-subject construction, $F(1, 198) = 88.4$, $p < .001$.

To summarize, when one syntactic construction was more frequent than any other construction, it achieved robustness by becoming a default mapping. When more than one syntactic construction was high frequency, a default would not work, and the more-frequent constructions achieved robustness by using higher activations.

DISCUSSION

Training frequency had a reliable impact on the learning and robustness to loss during comprehension of different syntactic constructions. No complex structural machinery was required; all that was required was a difference in frequency. What are the origins of this frequency difference in natural language? We propose that the difference derives from two sources. The original source is some small, functionally based inclination. For example, consider the frequency advantage of active voice over passive voice. It could well arise from speakers' inclinations for making agents or causers of actions, rather than patients or themes, the topics of discourse. Then, the simple active construction fits this inclination by placing the agent in subject position. A second source is that as syntactic constructions become more frequent, better practiced, and easier, they become preferred. This preference leads to more frequent selection for expression of thought. In turn, the constructions become even more frequent and thereby easier. As this feedback cycle continues, the frequency difference grows.

Support for this frequency–practice hypothesis comes from the early acquisition of the passive construction in Sesotho, a Bantu language. In Sesotho, the passive construction is much more frequent than it is in English, and children acquire it earlier (Demuth, 1989)—just as our frequency–practice hypothesis predicts. The difference presumably lies in the fact that the passive voice plays a more important functional role in Sesotho than it does in English.

Even in English, some specific lexical constructions take only passive form. For instance, in the commentary on hockey games, the player with the puck is always the subject. If another player checks the player with the puck, the sentence must appear in passive form to retain the player with the puck as subject, for example, "Gretzky was checked by the aggressive Penguin." This inclination and its resulting frequency difference should make this particular passive easy to understand. It should also make the equivalent active form relatively difficult to understand within this specific context. In the same vein, academic readers encounter passive constructions more frequently in academic contexts than do casual readers in casual contexts. We suggest that due to this higher frequency of practice, academic readers should find passive constructions relatively easy to process.

A strong prediction of our frequency–practice hypothesis is that artificially increasing the frequency of a syntactic construction should make it easier for humans to process. For example, with enough practice, the passive voice could conceivably become easier than the active voice. It is just a matter of frequency and practice. We are currently investigating this empirical prediction. Subjects perform a sentence–picture verification task, and we manipulate the frequency of active and passive constructions. We took care to preclude the use of any extralinguistic strategies. Early results look promising for altering the relative speed of comprehending these constructions by changing their relative frequencies.

Critically, the model's simulation of the effects of frequency on learning and loss depend exactly on frequency and practice. The model is not provided with any predetermined syntactic or semantic processes or representations, and there is no central executive or central processing resource or capacity. Furthermore, frequency is not represented by anything like a counter that gets consulted during processing. Instead, frequency and practice effects are inherent in the normal functioning of the network. Frequency of practice leaves its mark by changing and strengthening the weights of the network, making more-frequent words and constructions easier to process and maintain.

The picture of language comprehension we are developing is much more like a memory than a structured program. Comprehension is viewed as mapping from surface forms to meaning, and the strength of this mapping affects the speed and accuracy of processing. Each instance of practice modifies the processor and the mapping and makes that specific sentence, those words, and that syntactic construction easier to process. Damage affects processing generally and hurts weaker mappings more than it hurts stronger mappings. Consequently, stronger mappings, those that have received more practice, are relatively preserved.

More-frequent syntactic constructions can achieve robustness to noise damage in several ways. If one syntactic construction is more frequent than any other, that syntactic construction can become a default. If more than one syntactic construction is high frequency, then these constructions can achieve robustness by cultivating more units with higher activations.

An obvious extension of the current work is to consider more complex syntactic constructions, such as relative clauses. The human data from Miyake et al.'s (1994) speeded sentence comprehension work and Caplan and Hildebrandt's (1988) aphasic sentence comprehension work show that more complex constructions are more prone to loss. There are, of course, a variety of reasons for increases in breakdown; for instance, more complex syntactic constructions contain more words and more participants. These experiments also show that syntactic constructions that follow a noncanonical word order, such as center-embedded object-relative clauses, break down more than syntactic constructions that follow a canonical word order, such as center-embedded subject-relative clauses. Because these noncanonical forms are far less common than the canonical forms, we can again appeal to simple frequency as the cause of the difficulties. Of course, alternative theories exist to explain these differences in terms of structures and special processes, but we propose that, once again, these ancillary mechanisms are unnecessary and that the simple difference in frequency and practice can account for the behavioral differences.

Although our research does not directly involve second language learning, some implications can be drawn. As we saw, the frequency of training of each construction is important for both time to mastery and robustness to breakdown. Frequency should be an equally important factor in second language learning: The frequency of exposure or training of a construction will have a strong impact on the strength of the construction's representation.

A second language program that ignored native frequencies, for example, by presenting all syntactic variations with equal frequency, would prove a

grave disservice to the language learners. In terms of cue strength (Bates & MacWhinney, 1987), learners would not acquire the correct relative strengths of different cues. They would not know which mappings of surface forms to meanings were the default, and they would not know when additional linguistic cues had the strength to override the default. Elizabeth Bates shares a classic example, translated here from the original Italian, "Which dish is best? The lasagna recommends Elizabeth." With appropriate Italian cue strengths, the sentence very naturally conveys Elizabeth's opinion. With inappropriate cue strengths, for example in English, in which noun–verb–noun order is crucial, the sentence conveys something quite extraordinary. Pity the second language learner who has not learned the appropriate cue strengths.

CONCLUSIONS

Why are noncanonical syntactic constructions, like the simple passive, harder to understand and more prone to break down than similar but canonical syntactic constructions? Our goal has been to uncover and demonstrate general information processing principles, such as length and frequency, to explain these differences in learning and loss. We posit a distributed processor where mappings improve with practice and where weaker mappings fall prey to noise damage as processing gets longer, harder, or faster. We suggest that these ideas and findings are relevant to second language learning.

ACKNOWLEDGMENTS

Preparation of this chapter was supported by grants from the National Institutes of Health (RO1 NS 29926) and the Army Research Institute (DASW0194-K-0004, DASW0196-K-0013 and DAAG55-97-1-0224) awarded to Morton A. Gernsbacher.

REFERENCES

Bates, E., Friederici, A., & Wulfeck, B. (1987). Sentence comprehension in aphasia: A crosslinguistic study. *Brain and Language, 32,* 19–67.

Bates, E., & MacWhinney, B. (1987). Competition, variation, and language learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition: The 20th annual Carnegie symposium on cognition* (pp. 157–193). Hillsdale, NJ: Lawrence Erlbaum Associates.

Caplan, D., Baker, C., & Dehaut, F. (1985). Syntactic determinants of sentence comprehension in aphasia. *Cognition, 21,* 117–175.

Caplan, D., & Futter, C. (1986). Assignment of thematic roles to nouns in sentence comprehension by an agrammatic patient. *Brain and Language, 27*, 117–134.
Caplan, D., & Hildebrandt, N. (1988). *Disorders of syntactic comprehension.* Cambridge, MA: MIT Press.
de Villiers, J. (1985). The acquisition of English. In D. Slobin (Ed.), *The crosslinguistic study of language acquisition, vol. 1* (pp. 27–139). Hillsdale, NJ: Lawrence Erlbaum Associates.
Demuth, K. (1989). Maturation and the acquisition of the Sesotho passive. *Language, 65,* 56–80.
Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14,* 179–212.
Ford, M. (1983). A method for obtaining measures of local parsing complexity throughout sentences. *Journal of Verbal Learning and Verbal Behavior, 10,* 203–218.
Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior, 12,* 627–635.
French, R. M. (1992). Semi-distributed representations and catastrophic forgetting in connectionist networks. Philosophical issues in connectionist modelling: Connection science [special issue]. *Journal of Neural Computing, Artificial Intelligence & Cognitive Research*, *4*, 365–377.
Gernsbacher, M. A. (1984). Resolving twenty years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General, 113,* 256–281.
Goldman-Eisler, F., & Cohen, M. (1970). Is N, P, and NP difficulty a valid criterion of transformational operations? *Journal of Verbal Learning and Verbal Behavior, 9,* 161–166.
Goodglass, H., Christiansen, J. A., & Gallagher, R. (1993). Comparison of morphology and syntax in free narrative and structured tests: Fluent vs. nonfluent aphasics. *Cortex, 29*, 377–407.
Grodzinsky, Y. (1986). Language deficits and the theory of syntax. *Brain and Language, 27,* 135–159.
Hinton, G. E. (1989). Connectionist learning procedures. *Artificial Intelligence, 40,* 185–234.
Hinton, G. E., & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review, 98,* 74–95.
Horgan, D. (1978). The development of the full passive. *Journal of Child Language, 5,* 65–80.
Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*, 329–354.
Kilborn, K. (1991). Selective impairment of grammatical morphology due to induced stress in normal listeners: Implications for aphasia. *Brain and Language, 41,* 275–288.
King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language, 30,* 580–602.
Kolk, H. H., & Van Grunsven, M. M. (1985). Agrammatism as a variable phenomenon. *Cognitive Neuropsychology, 2,* 347–384.
Linebarger, M. C., Schwartz, M. F., & Saffran, E. M. (1983). Sensitivity to grammatical structure in so-called agrammatic aphasics. *Cognition, 13,* 361–392.
MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology, 24,* 56–98.
MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101,* 676–703.
Marchman, V. A. (1993). Constraints on plasticity in a connectionist model of the English past tense. *Journal of Cognitive Neuroscience, 5*, 215–234.
Martin, R. C., Wetzel, F. W., Blossom-Stach, C., & Feher, E. (1989). Syntactic loss versus processing deficit: An assessment of two theories of agrammatism and syntactic comprehension deficits. *Cognition, 32,* 157–191.
McClelland, J. L. (1993). Toward a theory of information processing in graded, random, and interactive networks. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 655–688). Cambridge, MA: MIT Press.

Mitchell, D. C. (1994). Sentence parsing. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 375–409). San Diego, CA: Academic Press.
Miyake, A., Carpenter, P. A., & Just, M. A. (1994). A capacity approach to syntactic comprehension disorders: Making normal adults perform like aphasic patients. *Cognitive Neuropsychology, 11,* 671–717.
Pierce, R. S., & Wagner, C. M. (1985). The role of context in facilitating syntactic decoding in aphasia. *Journal of Communication Disorders, 18,* 203–213.
Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland, and the PDP Research Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1, pp. 318–362). Cambridge, MA: MIT Press.
Schwartz, M. F., Saffran, E. M., & Marin, O. S. M. (1980). The word order problem in agrammatism: I. Comprehension. *Brain and Language, 10,* 249–262.
Seidenberg, M. S., & McClelland, J. L. (1989). A distributed developmental model of word recognition and naming. *Psychological Review, 96,* 523–568.
St. John, M. F., & McClelland, J. L. (1990). Learning and applying contextual constraints in sentence comprehension. *Artificial Intelligence, 46,* 217–257.
Taylor, I., & Taylor, M. M. (1983). *The psychology of reading.* New York: Academic Press.
Zurif, E., & Swinney, D. (1994). The neuropsychology of language. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 1055–1074). San Diego, CA: Academic Press.