ACADEMIC
PRESS

Available online at www.sciencedirect.com
SCIENCE DIRECT®

Brain and Language 86 (2003) 40–56

Brain and Language

www.elsevier.com/locate/b&l

Phonology and syntax in specific language impairment: Evidence from a connectionist model☆

Marc F. Joanisse[a,*] and Mark S. Seidenberg[b]

[a] *Department of Psychology, The University of Western Ontario, London, Ont., Canada N6A 5C2*
[b] *University of Wisconsin, Madison, WI, USA*

Accepted 23 July 2002

Abstract

Difficulties in resolving pronominal anaphora have been taken as evidence that Specific Language Impairment (SLI) involves a grammar-specific impairment. The present study explores an alternative view, that grammatical deficits in SLI are sequelae of impaired speech perception. This perceptual deficit specifically affects the use of phonological information in working memory, which in turn leads to poorer than expected syntactic comprehension. This hypothesis was explored using a connectionist model of sentence processing that learned to map sequences of words to their meanings. Anaphoric resolution was represented in this model by recognizing the semantics of the correct antecedent when a bound pronoun was input. When the model was trained on distorted phonological inputs—simulating a perceptual deficit—it exhibited marked difficulty resolving bound anaphors. However, many other aspects of sentence comprehension were intact; most importantly, the model could still resolve pronouns using gender information. In addition, the model's deficit was graded rather than categorical, as it was able to resolve pronouns in some sentences, but not in others. These results are consistent with behavioral data concerning syntactic deficits in SLI. The model provides a causal demonstration of how a perceptual deficit could give rise to grammatical deficits in SLI.

---

1. Introduction

Children with specific language impairment (SLI) exhibit marked deficits in acquiring multiple aspects of language. Observed deficits include grammatical morphology (Gopnik & Crago, 1991; Leonard, 1987), phonology (Bird & Bishop,

---

[☆] This research was supported by Grants PGS-B 208931 and RG 238543-01 from the Natural Sciences and Engineering Research Council of Canada to MFJ, and National Institute of Mental Health grants KO2 MH01188 and RO1 MH58723 to MSS.

[*] Corresponding author. Fax: +519-661-3961.
E-mail address: marcj@uwo.ca (M.F. Joanisse).

doi:10.1016/S0093-934X(02)00533-3

1992; Edwards & Lahey, 1998; Joanisse, Manis, Keating, & Seidenberg, 2000; Kamhi, Catts, Mauer, Apel, & Gentry, 1988) and syntax (Montgomery, 1995; van der Lely, 1996; van der Lely & Stollwerck, 1997). What is perhaps most interesting about children with SLI is that these language deficits do not appear to co-occur with hearing, cognitive and frank neurological deficits (Johnston, 1984). This apparent language-specificity raises important issues about the bases of the impairment, the nature of linguistic knowledge, and how this knowledge is acquired and represented in the brain. However, evidence suggests that these children do in fact exhibit other types of deficits that extend beyond language, including problems with working memory (Gathercole & Baddeley, 1990; Kirchner & Klatzky, 1985; Montgomery, 1995) and speech perception (Tallal & Piercy, 1974), visual imagery (Johnston & Ellis Weismer, 1983) and analogical reasoning (Nelson, Kamhi, & Apel, 1987; Ellis Weismer, 1985). Thus, in at least some children, 'SLI' is not specific to language. The extent to which these non-linguistic impairments can explain these children's grammatical deficits is a matter of some controversy in current psycholinguistic research.

This article focuses on the grammatical deficits that are characteristic of SLI. There are two opposing accounts of these deficits. One holds that SLI reflects anomalies in the neurobiological representation of grammar. The non-linguistic capacities that are impaired in SLI are thus thought to involve other causes or to be consequences of the primary linguistic impairment (van der Lely, 1996). The alternative view holds that the linguistic deficits are secondary to impairments in the perception of speech (Joanisse & Seidenberg, 1998; Leonard & Eyer, 1996; Tomblin & Pandich, 1999), the neurobiological basis of which also affects other cognitive capacities (Tallal, Miller, & Fitch, 1995). The exact nature of this impairment and the extent to which these impairments can explain the full range of language problems in these children have been the subject of considerable debate. A major question is how an impairment in the processing of speech could give rise to impaired use of grammar. The present work addressed this issue using computational modeling, with the goal of exploring whether grammatical deficits can emerge in a connectionist network as a consequence of a simulated perceptual deficit. This would provide a stronger causal link between speech perception, phonology and syntactic impairments.

In this work, we address data indicating that children with SLI have difficulty comprehending specific types of syntactic relationships, such as reversible passives and bound pronouns and reflexives (van der Lely, 1996; van der Lely & Stollwerck, 1997). These deficits appear to suggest that SLI is associated with difficulty in processing configural aspects of grammar. For instance, generative syntax explains anaphoric reference as being governed by *Binding Conditions* that describe how the syntactic structure of a sentence determines how pronouns and reflexives are bound to their antecedents (Chomsky, 1981). Binding conditions explain why *them* cannot refer to *the boys* in the sentence *The boys told stories about them*. Thus, the generative account of SLI holds that problems with pronouns and reflexives are due to the absence or impairment of this component of grammar (van der Lely & Stollwerck, 1997). Since syntax is not closely related to phonological processing or auditory perception within the generative framework, such a deficit appears to present a significant challenge to the theory that SLI is caused by a phonological or speech processing deficit.

The present work investigates the alternative theory that a speech processing deficit in SLI limits the ability to comprehend some aspects of grammar. The link between speech perception and grammar is provided by the idea that a perceptual deficit can lead to reduced working memory capacity, due to its reliance on a phonological code during sentence comprehension. As discussed below, the idea that impaired sentence comprehension is due to a working memory deficit in SLI is not a new one (Gathercole & Baddeley, 1990; Montgomery, 1995). The contribution of the present work is twofold: first, it seeks to explain working memory deficits in SLI as a consequence of phonological problems, which are in turn a consequence of a perceptual deficit (Joanisse & Seidenberg, 1998). Second, it tests the causal link between speech perception, phonology, working memory and sentence comprehension deficits using a connectionist model.

Below, we present a brief review of syntactic binding deficits in SLI, along with previous research supporting the theory that phonological working memory influences this type of sentence comprehension ability. Evidence is also presented implicating phonology and working memory impairments in SLI, and correlating these deficits with their language impairments. Next, a simplified model of syntactic processing is developed as a demonstration of how a perceptual deficit might lead to working memory problems and deficits in comprehending certain sentence types. In this model, sentence structure is learned as a result of the serial nature of this task, and is directly affected by the quality of the phonological inputs it receives. The results of this modeling work support the theory that syntactic comprehension deficits can indeed be understood from the perspective of a speech perception impairment.

2. Overview of empirical data

van der Lely and Stollwerck (1997) investigated anaphoric processing in 12 children with what they characterized as grammatical SLI. Children were asked to judge whether a sentence read aloud matched a given picture, using a yes or no answer. For example, subjects were shown pictures of animals and people performing specific actions, and were read a sentence containing a bound pronoun or reflexive in the object position (e.g., *Balloo bear says Mowgli is tickling him*, vs. *Balloo bear says Mowgli is tickling himself*). The authors found that children with SLI were significantly worse at this type of task, compared to younger normally developing children at a similar stage in grammatical or vocabulary development (Table 1).

The authors also investigated whether this deficit represented a deficit that was specifically syntactic, or a more general problem in understanding sentences, by also testing cases where pronoun reference could be inferred from the sentence context.

Table 1
Data from van der Lely and Stollwerck (1997) indicating poorer syntactic resolution of anaphors in children with SLI compared to controls

Sentence type	Group	
	Control	SLI
Pronouns	95.8	86.4
Pronoun-context	100.0	93.1
Reflexives	87.3	78.1
Reflexive-context	99.0	94.4

For the sake of clarity, data have been averaged across the match/mismatch and quantifier/noun conditions in Experiments 1 and 2 of the original study. Control data are taken from the LA-1 group in the study, and consist of younger normal children matched for grammatical language achievement. Data are percent correct.

For instance, the antecedent of *her* in *Peter Pan says Wendy is tickling her* can be inferred from the fact that Wendy is the only female entity mentioned in the sentence. As Table 1 shows, children with SLI performed much closer to normals in these cases, suggesting that their deficit is syntactic in nature, and leaves more general comprehension skills intact.

One complication for this conclusion is that the SLI group did not consistently perform at chance on all sentences requiring the use of syntactic knowledge. Instead, their level of performance suggests that they were using syntactic information during this task, though to a lesser degree than controls, van der Lely and Stollwerck (1997) analyzed subjects' accuracy data by presenting performance for the picture match and mismatch conditions separately (i.e., trials in which 'yes' was the target answer were analyzed separately from trials in which 'no' was the target). This did yield performance at or near chance in five of eighteen experimental subtasks, but only in sentence-picture mismatch conditions where the correct answer was 'no.' This instead suggests that subjects tended toward false positive responses, rather than producing incorrect responses across all sentences involving anaphors. When overall percent correct is considered (100*(hits + correct rejections/false alarms + incorrect rejections)) as in Table 1, it is clear that SLI subjects were able to provide the correct answer at a rate above chance on these sentences (for further discussion, see Bishop, 1997).

This result is important because it suggests language impaired children are in fact able to process syntactic relationships, though to a lesser degree than their peers. This presents a serious challenge to a grammatical deficit hypothesis, which claims that these children's syntactic comprehension deficit is due to an impaired or missing module of grammar. If this is indeed the case, it is unclear how children with SLI are able to perform above chance on sentences in which a pronoun or reflexive's reference can only be resolved syntactically (rather than contextually). We suggest that this deficit is instead best accounted for by the theory that the comprehension of some sentences is impaired due to the impact of a perceptual/phonological deficit on working memory for sentence comprehension. This theory has the advantage of being able to explain not only the syntactic deficits in SLI, but also a broader range of their grammatical impairments. For example, it is suggested that a phonological deficit can also explain the character and range of morphological impairments in SLI (Joanisse & Seidenberg, 1998). It also provides an account for the observation that grammatical deficits in SLI are rarely wholesale. For example, SLI children in the van der Lely and Stollwerck study seemed to be performing above chance on the pronoun comprehension task, which suggests that they were using syntactic information to a limited extent. We argue that the graded nature of language deficits in SLI is problematic for a grammar-based approach, and is instead better explained as a consequence of a deficit to more general processing mechanisms such as speech perception and working memory. This theory is explored in-depth below, where we review the literature on phonology and working memory in sentence processing and then develop a model that explores the effects of a phonological impairment on syntactic processing.

2.1. Phonology and working memory in syntax

Working memory is used during sentence comprehension to help determine the critical relationships among a sentence's constituents, and to resolve possible ambiguities (MacDonald & Christiansen, 2002; Waters & Caplan, 1985). There is also evidence suggesting that this working memory capacity relies on a phonological code (e.g., Daneman & Tardif, 1987). This raises the possibility that a deficit in perceiving words' phonological forms can have a negative impact on the ability to process sentences, especially in sentences that rely heavily on working memory.

One methodology for studying the impact of working memory on sentence recognition is to group normal subjects based on their reading span, a quantification of working memory capacity involving the ability to maintain a list of words in memory while reading sentences (Daneman & Carpenter, 1983). King and Just (1991) have found that low reading span subjects tend to show slower reading times for the main verbs of object-relative sentences (*The reporter that the senator attacked admitted the error*) compared to subject-relative sentences (*The reporter that attacked the senator admitted the error*), whereas high span readers do not. Likewise, Daneman and Carpenter (1983) have found that higher span subjects were better able to determine the referents of pronouns in sentences compared to lower span subjects.

Such findings suggest that working memory capacity can influence sentence comprehension, possibly due to its role in maintaining sentence constituents in memory for the purpose of resolving syntactic relationships. These experiments also illustrate that sentences can differ in the demands they place on working memory for a variety of reasons, including sentence length, syntactic complexity (the number of embedded clauses and how these clauses are arranged), the frequency with which a construction occurs in everyday usage, and the phonological complexity of words in a sentence (MacDonald & Christiansen, 2002; McCutchen, Dibble, & Blount, 1994; St. John & Gernsbacher, 1998). Thus, reduced working memory span will lead to weaker syntactic ability. We propose that SLI represents an extreme case of this type of limitation, in which speech perception deficits lead to impaired phonological representations that in turn have a negative impact on the capacity to maintain sentences in working memory.

Further evidence for this theory comes from studies of agrammatic aphasics, who typically demonstrate difficulties in syntactic comprehension. Parallel to what is argued about SLI, some authors have claimed that these deficits derive from an impairment to neural regions subserving core grammar (Grodzinsky, 1990; Hildebrant, Caplan, & Evans, 1987). However, the alternative theory holds that sentence comprehension deficits in aphasia derive from a processing deficit that inhibits the use of syntactic knowledge. Syntactic difficulty reflects the demands of some sentences on general processing systems that subserve sentence comprehension. Dick et al. (2001) have investigated this theory by assessing syntactic comprehension in both aphasic and normal adults, using a sentence–picture matching task similar to what is used when assessing SLI. The authors tested aphasic patients' comprehension of active, passive, object cleft and subject cleft sentences. Their study found that although these sentences share an identical underlying structure, aphasic patients are much poorer at resolving object cleft and passive sentences, whereas performance on subject cleft sentences and active sentences was much closer to normal.

One possible explanation is that these sentences place differential demands on general processing. To test this hypothesis, Dick et al. investigated sentence comprehension in normal adults when listening to speech stimuli under a number of 'stress' conditions that either modified the acoustic signal in a variety of ways (e.g., time-compression, low-pass filtering, and adding acoustic noise), or involved performing concurrent working memory tasks. The authors observed that several conditions that involved manipulating the speech signal resulted in clear sentence comprehension difficulties, especially when these were combined with a concurrent phonological working memory task. Moreover, the actual pattern of difficulty that subjects demonstrated on the different sentence types closely matched that of aphasics.

The results of these studies suggest that not all sentence structures are equally easy to process, and that syntactic comprehension deficits can result from degraded speech inputs. It further supports the theory that working memory for sentence comprehension relies on a phonological code, such that degraded phonological representations resulting from a speech perception deficit will lead to problems with sentence comprehension.

2.2. Phonology and working memory impairments in SLI

The perceptual deficit account suggests that the source of syntactic acquisition deficits in SLI lies in listeners' difficulty maintaining adequate phonological representations of sentences in working memory in order to comprehend them. Indeed, there is good evidence to support the theory of working memory limitations in SLI (Gathercole & Baddeley, 1990; Kirchner & Klatzky, 1985; Montgomery, 1995). For example, Gathercole and Baddeley (1990) used a non-word repetition task to test working memory in children with SLI. It was expected that children with a working memory impairment would perform more poorly on longer words (i.e., 3- and 4-syllables) compared to shorter ones (1- and 2-syllables), since longer non-words place an increased load on working memory. As predicted, the authors found that children with SLI were significantly worse than controls at repeating these longer non-words, compared to shorter ones. A variety of other studies have also demonstrated similar non-word repetition problems, suggesting that this is a reliable characteristic of SLI (Bishop, North, & Donlan, 1996; Edwards & Lahey, 1998; Kamhi et al., 1988; Montgomery, 1995).

Results from Montgomery (1995) suggest that this impairment is related to syntactic processing difficulties in SLI. This study replicated previous findings of non-word repetition and syntactic comprehension deficits in SLI. However, it was also found that these children's performance on the two tasks was strongly correlated ($r = .62$, $p < .001$), suggesting that the two impairments are linked. Further analyses revealed that the SLI group had greater difficulty comprehending longer sentences like *The big brown furry dog is quickly chasing the little yellow and black cat*, compared to shorter, syntactically similar sentences like *The big brown dog is chasing the little cat*. These results seem to suggest that children with SLI do have working memory difficulties, and that this deficit is related to sentence comprehension impairments.

The theory that emerges from the data presented above is that children with SLI have working memory limitations that lead to their sentence comprehension difficulties. The proximal cause of this working memory limitation is suggested to be their impairment in phonology, which is itself caused by a speech perception deficit. What is missing from this account is a more direct illustration of the causal link among speech perception, phonology, working memory, and syntax deficits. The next section addresses this issue with the help of two connectionist simulations that demonstrate how such a causal link might be drawn. In the first simulation, a model of normal syntactic acquisition was implemented within a distributed neural network simulation, to demonstrate how this architecture encodes the relevant aspects of syntax. In the second simulation, a speech perception impairment was implemented within this same architecture, in order to investigate whether such a deficit would lead to a sentence comprehension deficit consistent with what is observed in children with SLI.

3. Simulation experiment

The simulations used in this study were not intended to simulate all aspects of syntactic processing, which surely involve a great deal more knowledge and structure than what was implemented here. Instead, they implemented several characteristics of sentence comprehension that were relevant to pronoun resolution: the recognition of word meanings in sentence context, the acquisition of abstract phrase structure such as verb subcategorization, and the use of syntactic structure to resolve long-distance syntactic dependencies.

The first simulation represented a demonstration of how normal processing can be achieved within a connectionist network. It implemented several aspects of sentence comprehension through the task of mapping the phonological forms of words in sentences to their meanings. This model learned to resolve anaphoric relationships by binding the meanings of pronouns and reflexives to their antecedents. The dynamics of this task were accomplished using an architecture that encoded characteristics of grammatical sentences within its connection weights. Working memory was simulated by implementing a discrete feedback loop that allowed the network to maintain internal representations as it was exposed to successive inputs.

3.1. Model architecture and task

The network architecture used in this study is illustrated in Fig. 1. Each unit in the network used a logistic activation function (range: 0.0–1.0). The input layer consisted of 108 units that represented the phonological features of 6 phonemes, fitting a CVCCVC frame. Each phoneme slot consisted of 18 binary phonological features that captured the natural classes of phonemes;[1] an activation value of 1.0 represented the presence of a feature, and 0.0 represented the absence of a feature. This coding scheme was able to represent a number of 1- and 2-syllable English words. For example, the words *John* and *Mary* were encoded as [ǰ an _ _ _] and [mϵ_ri_]. Words that did not fit this frame were truncated, usually by deleting one of the consonants in a cluster.

The output layer represented word meanings using a system of 98 distributed semantic features taken from the WordNet database (Miller, 1990). This feature scheme encoded important characteristics of semantic representations, such as the ability to represent degrees of similarity among word meanings. For example, the word cat was represented by activating the [NEUTRAL-GENDER], [ANIMAL], and [FELINE] features; *dog* was represented as [NEUTRAL-GENDER], [ANIMAL], and [CANINE]. Verbs were represented similarly, for example *surmise* was represented as [INFER], [SPECULATE], [EXPECT], and [JUDGE]. Words with similar grammatical roles tend to resemble each other, a tendency that might represent a useful cue to learning such aspects of grammar as ‘noun’ and ‘verb’ (Seidenberg, 1997).

The network learned to recognize sentences by mapping sequences of phonologically encoded words to semantically encoded outputs. The training procedure was as follows: at the start of each training trial, a sentence was chosen at random from the corpus of training sentences described below. Each word in the sentence was presented to the network for two time steps (a time step is defined as the propagation of activation between two adjacent layers). When activation of a word had propagated to the output layer, the resulting output pattern was compared to the word’s target pattern; connection weights were adjusted using the backprop through time algorithm in a way that sought to minimize the difference between actual and desired outputs (Williams & Peng, 1990). The learning rate was set to 0.005; error radius (the tolerance to which the network calculated error) was 0.1.

---

[1] Voiced, voiceless, consonantal, vocalic, obstruent, sonorant, lateral, continuant, non-continuant, ATR, nasal, labial, coronal, anterior, high, distributed, dorsal, and radical.

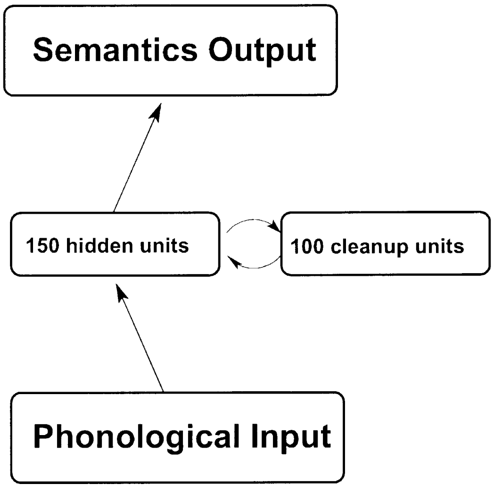

Fig. 1. Network used to simulate syntactic processing. Inputs represented words' phonological forms; outputs encoded word meanings. The network learned syntactic dependencies through exposure to grammatical sentences.

Each training trial ended when the activation of every word in the sentence had propagated through the network.

The input and output layers were connected through a set of 150 hidden units. These were in turn connected to and from a layer of 100 cleanup units (Plaut & Shallice, 1993), creating a recurrent architecture in which information could be maintained within the network by propagating activation back and forth between these two layers. This allowed the network to maintain a representation of previous states, in what could be described as the network's 'working memory.' For example, when the network was presented with the sentence *Laurie took* (*the*) *ball from Chuck*.[2] the network was able to retain an internal trace of the verb *took* and use it further on in the sentence to better recognize the preposition *from*, based on previous exposure to similar sentences.

The network learned to identify the antecedents of pronouns and reflexives by outputting the semantics of the correct referent when a bound pronoun or reflexive was present on the input. For instance, the word *himself* in *John says Bill likes himself* was encoded as [MALE] [REFLEXIVE-PRONOUN] [HUMAN] [BILL], whereas in the sentence *Bill says John likes himself* it was encoded as [MALE] [REFLEXIVE-PRONOUN] [HUMAN] [JOHN]. The same encoding was used for non-reflexive pronouns that were bound to a noun in the sentence, as in *John says Bill likes him*. The training set also contained some sentences containing unbound pronouns, as in *Carrie likes her*. In these cases, the semantics of the unbound pronoun were set to the appropriate entity type ([MALE], [FEMALE] or [NEUTER]) and [PRONOUN], with all other semantic features set to 0.

The network also represented temporal activation sequences using what are called delay lines (Pearlmutter, 1995). Previous research into the dynamics of neural systems has revealed that assemblies of neurons are able to respond to temporal orderings of stimuli by organizing themselves into order-sensitive groupings (e.g., Carpenter, Georgopoulos, & Pellizzer, 1999). This type of temporal order sensitivity was simulated in the present model by modifying how unit activation and error were

---

[2] These parentheses are used to indicate determiners that were not included in the training sentences, but are provided for the sake of clarity.

calculated. Typically, unit activation (x_i) is calculated as $x_i = \sum_j w_{ji} y_j$, where w_{ji} is the weight of the connection between nodes j and i, and y_j is the activation of nodes connected to it.

In this work, this equation was modified by adding time delays τ_{ji}, the lag along the connection from unit j to i: $x_i(t) = \sum_j w_{ji} y_j(t - \tau_{ji})$. This had the effect of allowing the activation of neuron i at time t to be influenced not only by the state of neuron j at time $t-1$, but also at times $(t-2)\ldots(t-\tau)$. In essence, this permitted the network to be more directly sensitive to inputs from layers at time steps earlier than $t-1$. This addresses a major limitation of standard backprop through time and simple recurrent networks, that activation propagating through the network at time n tends to obliterate activation from earlier time steps $n-1, n-2, \ldots$, making it difficult to encode temporal dependencies spanning across many time steps. Delay lines allow some connections within the network to be sensitive to earlier network states, which in turn allow the network to learn long-distance syntactic relationships. Delay lines were implemented in the present simulation by randomly assigning τ values of between 1 and 10 to each weight in the Hidden → Cleanup and Cleanup → Hidden connection sets. Connection weights in all other regions in the network were assigned a delay value of 1, effectively mimicking standard backprop through time activation propagation.

3.2. Training corpus and grammar

A simplified English grammar was created using a set of 23 verbs with various subcategorization structures, 26 proper nouns of either gender, and 9 gender-neutral nouns.[3] Sentences were created by algorithmically assigning nouns to each verb's thematic roles and adding any necessary prepositions. In the case of verbs with more than one subcategorization structure (*sleep*, *take*, *give*, *throw*, *run* and *walk*), sentences of each subcategorization structure were created (e.g., *Mike threw the ball to Chuck*, *Mike threw himself at Chuck*).

Sentences with pronouns were created by adding the pronouns *him*, *her*, *it*, *himself*, *herself*, and *itself* as direct and indirect object complements of verbs. The result was a set of 396,191 active voice, declarative simple sentences, some of which contained pronouns. Complex subordinate sentences were also created by adding a subject noun and psych verb (*think*, *guess*, *say*, and *surmise*) to a simple sentence, as in *Bob says Mary put* (*the*) *tree on* (*the*) *island*. The result was an additional 1,265,019 sentences in the training set, some of which contained pronouns. Together, this produced a total of approximately 1.5 million sentences.

The resulting sentences captured several important aspects of English syntax by encoding structural dependencies beyond simple linear relationships. In order to learn this grammar, the network had to acquire generalizations about grammatical classes of words, and facts about the structural organization of words in sentences. For instance, words group naturally into grammatical categories such as nouns and transitive verbs. Thus, a learner that is exposed to sentences like *Bob likes Harry*, *Bob likes Mary* and *Mary likes Bob* should also be able to recognize *Mary likes Harry* as grammatical, in spite of having never seen that specific sentence before. Likewise, the learner should not overgeneralize all members of a broader category to a context

---

[3] Verbs: *cry*, *sing*, *hit*, *console*, *threaten*, *grab*, *mock*, *tickle*, *slap*, *bite*, *approach*, *touch*, *hold*, *sleep*, *take*, *give*, *put*, *throw*, *dance*, *look*, *run*, *step*, and *walk*. Nouns: *Abe*, *Dot*, *Goose*, *Bill*, *Emma*, *Hat*, *Chuck*, *Fran*, *Island*, *Tony*, *Kim*, *Turkey*, *Yoshi*, *Carrie*, *Dog*, *Bailey*, *Laura*, *Cat*, *Pablo*, *Kate*, *Cow*, *Bob*, *Laurie*, *Turtle*, *Joe*, *Mary*, *Tree*, *Mark*, *Karen*, *Dan*, *Suzanne*, *Mike*, *Elaine*, *David*, and *Alex*.

that requires a narrower one; the learner should know that not all verbs take double-objects, and thus sentences like **Tony looked at Yoshi to Emma* are ungrammatical.

The grammar also required the network to learn structural relationships between words extending beyond bigram dependencies (e.g., what word can or cannot directly follow another). For example, the grammar encoded the fact that certain verbs require more than one NP or PP complement, as in *Bob put Mike on* (*the*) *cat* (vis. **Bob put Mike*). These types of dependencies extend beyond N-gram rules; they require the learner to encode configural patterns such as "VP = V(ditransitive) + NP + PP."

With respect to pronouns and reflexives, the grammar underlying the training sentences encoded a fairly complex set of principles governing antecedent–anaphor relationships. It was not sufficient for the network to learn N-back rules such as 'bind a pronoun to the fifth previous word' for a sentence like (*the*) *dog guessed* (*the*) *cow danced with it*, because this rule would not generalize to syntactically similar sentences like *Mary said Elaine mocked her*. Instead, the relationships between pronouns and their referents were determined by statistics holding over what could be described as abstract grammatical structures, such as the non-governing subject of a verb phrase.

For similar reasons, reflexives also required the development of structural (rather than finite-state) dependencies. While the generalization 'bind the first word in a sentence to a reflexive' may hold for some cases (*Mike put himself on* (*the*) *island*), it was not useful for other sentences (*Abe said Mike put himself on* (*the*) *island*). Similarly, a rule such as 'bind a reflexive to the second previous word' is also insufficient; while this would hold for *Dot showed Sally to herself* and *Dot showed herself to Sally*, it would not for the equally grammatical sentence *Dot showed Sally herself* or *Dot gave* (*the*) *turkey to herself*. Given these facts, it is suggested that a mechanism that successfully learns to recognize the sentences in the training set, and is able to accurately generalize to syntactically similar novel sentences, has arguably learned a grammar that reflects many of the important facts about human grammars in general.

3.3. Obtaining a training set

The network training set consisted of 40,000 sentences randomly selected from the corpus of 1.5 million sentences, as indicated in Table 2. A subset was used for several reasons. First, it was important to simulate how children learn syntax, which is by generalizing syntactic structures from exposure to a small subset of the possible sentences in their language. Second, limitations on computational hardware made it impractical to train the network on the entire corpus of sentences; the amount of RAM and hard disk space required for such a training pattern would have been considerably larger than what is currently available.

Sentences were presented to the network at random, using probability weighting such that some sentences were presented more frequently than others. Each sentence's probability estimate was obtained by first calculating the mean frequency of each word in the sentence (Francis & Kučera, 1982). For simple sentences, this value

Table 2
Breakdown of sentence types used in network training

Sentence type	Pronoun status		
	No pronouns	Pronoun	Reflexive
Simple	10,000	5000	5000
Complex-transitive	5000	2500	2500
Complex-ditransitive	5000	2500	2500

was then multiplied by 10 to reflect their greater likelihood of occurrence in everyday usage. Finally, actual presentation probabilities were obtained by calculating the log of the raw probability. This frequency manipulation served to increase the realism of the training procedure by increasing the network's exposure to higher frequency words and more common sentence structures. This scheme might also effectively simulate the process of 'starting small' in acquiring grammar, which has been shown to be useful for connectionist networks learning syntax (Elman, 1993).

3.4. Speech impaired network

To investigate the theory that impaired speech perception leads to impaired syntactic processing, a second network was trained in which subtle distortions were introduced in its phonological input. This network was identical to the one described above, except as follows: a random value was added to each phonological input unit's activation value at each time step during training. This noise had a Gaussian distribution centered on 0.0 with a standard deviation of ±0.4. Activation values falling outside the range of the network's logistic activation function were trimmed such that negative values were set to 0.0 and values greater than 1.0 were set to 1.0.

The effect of this Gaussian noise was to expose the network to a slightly different phonological form each time a given word was presented. This variation was not so great as to change the identity of a word's phonological form in most cases; it simply made it more difficult for the model to develop consistent phonological representations, since it was required to learn word forms that changed from one exposure to the next. This simulated a perceptual deficit typified by an increased tendency to miscategorize speech sounds and the development of imperfect or incomplete phonological representations of phonemes and words. This deficit was thus similar to what is observed in studies finding poorer than expected identification and discrimination of speech sounds in tests of categorical perception in language impaired children (Elliott, Hammer, & Scholl, 1989; Joanisse et al., 2000; Tallal & Piercy, 1974; Thibodeau & Sussman, 1979), and generally poor phonological processing (Bird & Bishop, 1992; Edwards & Lahey, 1998; Joanisse et al., 2000; Kamhi et al., 1988).

This impairment was predicted to have a negative impact on the network's performance due to the importance of phonology on the ability to maintain representations of words in a sentence over time. In contrast, such a deficit should not have a significant impact on syntactic abilities that are not working memory intensive. For instance, the network should display relatively normal word recognition and grammaticality judgments not involving long-distance syntactic dependencies.

3.5. Results

Both networks learned the training corpus until training error reached asymptote (unimpaired network: 3 million trials; speech impaired network: 3.5 million trials). Networks were first evaluated by presenting them with all the sentences in the training set and allowing them to output the appropriate semantic output for each word. Performance was assessed using a nearest-neighbor criterion, where the Euclidean distance between the network's output and each word in the network's vocabulary was computed, and the word with the smallest Euclidean distance was determined to be the winner. Sentences were scored as incorrect when the network's output for one or more words did not match the target word on this criterion. By this strict standard, the unimpaired network was able to correctly recognize 93% of the sentences in the training set. In contrast the speech impaired network correctly recognized 74% of sentences in the training set.

3.5.1. Grammaticality judgments

Grammaticality judgments are central to Generative linguistics, serving as a primary method for assessing speakers' grammatical knowledge (Chomsky, 1965). The present work diverges from the Generative view in many ways, including how it conceptualizes the representation of mental grammars; nevertheless, it remains important to explain this competence within the present model of syntactic knowledge. Grammaticality judgments were assessed using a methodology first proposed in Allen and Seidenberg (1999). This technique involves comparing how the networks treat novel grammatical and ungrammatical sentences by assessing the accuracy with which they can compute their correct semantic forms. Allen and Seidenberg observed important differences in how their model treated grammatical and ungrammatical sentences, with much poorer performance on ungrammatical ones compared to novel grammatical sentences. This differential treatment of grammatical and ungrammatical sentences suggests that their network had learned important generalizations about English sentence structure.

The networks in the present study were tested in a similar way; grammatical sentences were obtained by randomly selecting 40 sentences that were not used in training from the corpus of simple transitive sentences. A set of 40 ungrammatical sentences was obtained by modifying the verb or a verb complement in each novel grammatical sentence so that it was no longer grammatical (e.g., *Dot took Joe from Emma* was changed to **Dot look Joe from Emma*).

The networks' ability to compute the meanings of the sentences was assessed by measuring the mean sum-squared error (SSE) of the semantic output across all words in the sentence. This measure calculated the degree to which each unit in the Semantics layer deviated from the desired target output. For the unimpaired network, the mean SSE for the sentences in the grammatical testing set was very low compared to the ungrammatical sentences (grammatical: $M = 0.0306$, $SD = 0.0572$; ungrammatical: $M = 1.1831$, $SD = 1.9513$). An independent samples t test confirmed this difference to be significant ($t(78) = 3.7340$, $p < .001$). This is further illustrated in Fig. 2a, which compares how the unimpaired network recognized a grammatical sentence to an ungrammatical sentence. The region of interest is the point at which the network computed the output for *to*. In the grammatical sentence, the network produced a very low error across each word; in contrast, the ungrammatical use of the prepositional phrase *to Carrie* resulted in a higher error rate, suggesting that the network found the ungrammatical sentence more difficult. Fig. 2b illustrates a similar effect in the case of a verb-complementizer mismatch, where (*The*) *cat gave* (*the*) *turkey to Elaine* is contrasted with *(*The*) *cat gave* (*the*) *turkey at Elaine*. Here again, the network showed higher than expected error at the point of ungrammatically *at*, compared to *to*.

The speech impaired network was tested on the same sets of sentences. This network demonstrated the same pattern of lower SSE's for grammatical than ungrammatical sentences (grammatical: $M = 0.5951$, $SD = 0.9254$; ungrammatical: $M = 2.4949$, $SD = 1.9517$). An independent samples t test confirmed this difference was significant ($t(78) = 3.60$, $p < .001$). A two-way Anova was used to compare the two networks' mean SSE, with Sentence type (grammatical, ungrammatical) and Model (unimpaired, speech impaired) as factors. Both yielded significant main effects (Model: $F(1,39) = 31.569$, $p < .0001$; Sentence type: $F(1,39) = 31.733$, $p < .0001$). The Model × Sentence type interaction was marginal ($F(1,39) = 3.171$, $p = .08$).

These analyses indicate that the two simulations did differ insofar as the speech impaired network learned the task more poorly. However, the two were similar to the extent that both networks were able to distinguish grammatical and ungrammatical sentences, though this difference was less strong in the impaired network.

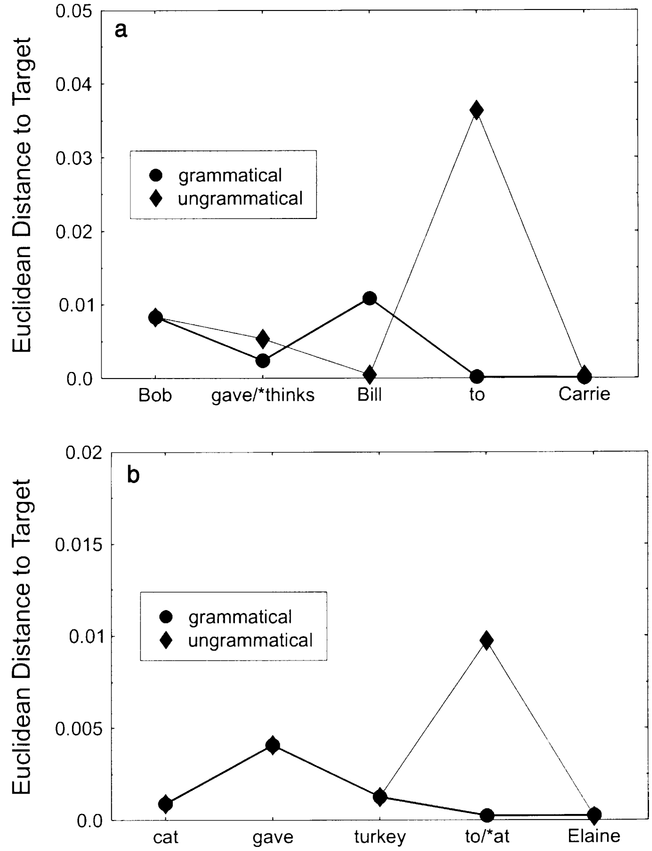

Fig. 2. Comparison of output error for grammatical and ungrammatical sentences. In (a), the network had greater difficulty computing the correct semantics of the ungrammatical pronoun *to* when the verb *thinks* was used. In (b), *at* was not a grammatical complementizer for the verb *gave*, thus the network had greater difficulty computing the correct semantics for it. Error was calculated as Euclidean distance to target.

3.5.2. Pronoun resolution

To investigate the networks' ability to resolve pronouns and reflexives, three sets of novel testing sentences were devised. The *Pronoun* testing set consisted of 24 complex-subordinate clause sentences containing bound pronouns, as in *Bob thinks Stan likes him*. The *Reflexive* testing set consisted of 24 testing sentences identical to the *Pronoun* set, but with each pronoun switched to the equivalent reflexive (e.g., *Bob thinks Stan likes himself*), in addition to 24 simple transitive sentences with a reflexive in the direct or indirect object position (e.g., *Stan likes himself*). The sentences in both the *Reflexive* and *Pronoun* sets were designed so that both nouns in the sentence were of the same gender as the pronoun or reflexive pronoun in the complement position; as such, these sentences tested the networks ability to resolve anaphoric expressions in cases where contextual information like gender could not be used. To test the opposite situation, in which gender information was useful for resolving bound pronouns, the *Gender* testing set was devised. This set consisted of 24 sentences used in the *Pronoun* and *Reflexive* sets, but with the nouns or pronouns modified so that the anaphors could be resolved with the help of gender information, (e.g., *Bob thinks Sally likes him*).

Testing was again performed using a nearest-neighbor decision method and a Euclidean distance metric. Sentences were scored as correct or incorrect on the basis of whether the correct output semantics was assigned to the referent of the pronoun

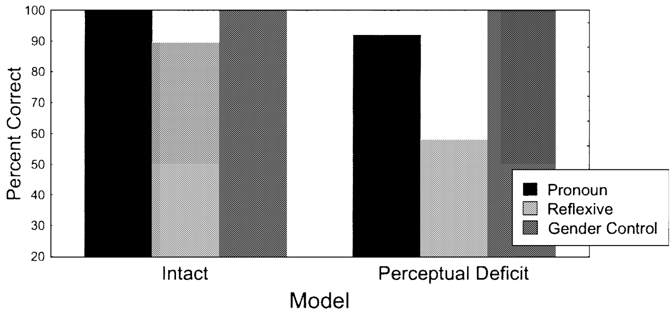

Fig. 3. Pronoun resolution in the speech impaired and intact networks.

in the sentence. The unimpaired network was tested at 2.2 million trials, a point that was intended to provide the analogue of a younger normal comparison control group; the speech impaired network was tested at asymptote (3 million trials). Results of testing showed performance consistent with the behavioral data; first, the speech impaired network was poorer on both the *Pronoun* and *Reflexive* sets, compared to the unimpaired network (Fig. 3). In addition, both networks performed more poorly on the *Reflexive* set than on the *Pronoun* set. Most interestingly, however, the speech impaired network's difficulty with pronouns did not extend to all sentence types; instead, it demonstrated the same degree of proficiency as the unimpaired network on the *Gender* set. This indicates that the speech impairment was not resulting in a wholesale degradation in performance, but instead lead to a specific difficulty in using syntactic information.

3.5.3. Impairment vs. developmental delay

One question is whether the speech impaired model was simply showing a developmental delay relative to the unimpaired model. A major claim about SLI is that it does not merely represent a delay in normal development, but is instead an aberrant developmental profile. To investigate whether this was the case here, network training was extended to 5 million training trials for both simulations, to determine whether the speech impaired network's performance would eventually reach a normal level of performance. The networks were tested on the *Pronoun* and *Reflexive* test sets at 100,000-trial intervals, over the entire course of training. The results showed that the speech impaired network's performance peaked at approximately 70% correct on the combined test sets at around 3.6 million training trials, and remained at that level until training was stopped at 5 million trials. In contrast, the unimpaired network exceeded 95% correct at 2.6 million trials, and remained at or near 100% for the remaining training intervals. A one-way ANOVA performed for the effect of model type (impaired and unimpaired), with training iteration as the random variable, was significant ($F = 99.923$, $p < .001$), confirming that performance differences between the two networks were significant over the entire course of training.

4. Conclusions and discussion

Behavioral studies have suggested that the morphosyntactic deficits in SLI are evidence of an impaired or missing module of linguistic competence (Berwick, 1997; Gopnik & Crago, 1991; Rice & Wexler, 1996). This in turn has lent support to the notion that an innate linguistic mechanism is used to learn and process language, and that normal language acquisition and processing is impossible in its absence (Chomsky, 1965; Pinker, 1989). The present work explored a different view that explains language acquisition impairments as a consequence of impaired speech perception and phonology (Joanisse & Seidenberg, 1998). We hypothesized that syntactic processing is impaired in SLI due to the impact of a perceptual deficit on phonology and working memory.

The acquisition of pronoun comprehension was modeled within a connectionist network that encoded grammar as statistical regularities that were learned through exposure to a sentence recognition task. This model was then used to test the influence of a speech perception impairment on sentence comprehension. This impairment yielded a degradation in the network's overall performance during training, and resulted in pronounced deficits in syntactically resolving bound anaphors. In addition, this deficit did not appear to be wholesale; as is the case in SLI, the network showed *some* degree of proficiency on these sentences.[4] This gradient effect represents a distinct prediction of the perceptual deficit theory, insofar as grammar-based theories explain these children's language deficits as a consequence of a missing module of grammar. Finally, the speech impaired network also demonstrated performance on some aspects of sentence comprehension that was closer to normal. For example, it was able to accurately recognize sentences in which pronoun reference could be inferred contextually. Thus, it seems that in spite of its problems with pronouns, the network had learned a number of important properties of syntactic structure.

The explanation for this behavioral pattern is that pronoun resolution is a dynamic task that relies on the retention of a sentence's phonological form over time. The connectionist architecture in these simulations used recurrent connections between the hidden and cleanup layers to maintain previously presented words for later retrieval. This mechanism differs from typical models of working memory in that it encodes intermediate results in a form that is distributed both in time and across multiple connected units. A consequence of this mechanism is that the addition of noise to the network's phonological inputs weakened its ability to maintain accurate representations of words over time. The conceptualization of working memory and phonology as two closely related mechanisms reflects the theory that the two are, in fact, inseparable and indistinct components of cognitive processing. This connectionist view of phonology and working memory is similar to that of MacDonald and Christiansen (2002), who propose that "neither *knowledge* nor *capacity* are primitives that can vary independently in theory or computational models; rather, they emerge from the interaction of network architecture and experience." The present work represents a clear illustration of this theory, by demonstrating how syntactic impairments can be directly tied to limitations in working memory capacity and phonological knowledge.

References

Allen, J., & Seidenberg, M. S. (1999). The emergence of grammaticality in connectionist networks. In B. MacWhinney (Ed.), *The emergence of language* (pp. 115–152). Mahwah NJ: Lawrence Erlbaum.

Berwick, R. C. (1997). Syntax facit saltum: Computation and the genotype and phenotype of language. *Journal of Neurolinguistics, 10*(2/3), 231–249.

---

[4] We assume that chance performance on the pronoun resolution task was between 6 and 11%, since network's task was to select the correct referent of a given pronoun or reflexive from the entire range of nouns of the appropriate gender.

Bird, J., & Bishop, D. V. M. (1992). Perception and awareness of phonemes in phonologically impaired children. *European Journal of Disorders of Communication, 27*, 289–311.

Bishop, D. V. M. (1997). *Uncommon understanding: Development and disorders of language comprenension in children*. Hove, East Sussex, UK: Psychology Press.

Bishop, D. V. M., North, T., & Donlan, C. (1996). Nonword repetition as a behavioral marker for inherited language development: Evidence from a twin study. *Journal of Child Psychology and Psychiatry, 37*, 391–403.

Carpenter, A. F., Georgopoulos, A. P., & Pellizzer, G. (1999). Motor cortical encoding of serial order in a context-recall task. *Science, 283*, 1752–1757.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1981). *Lectures in government and binding*. Dordrecht: Foris.

Daneman, M., & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*, 561–584.

Daneman, M., & Tardif, T. (1987). Working memory and reading skill re-examined. In M. Coltheart (Ed.), *Attention and performance XII* (pp. 491–508). Mahwah, NJ: Lawrence Erlbaum.

Dick, F., Bates, E., Wulfeck, B., Utman, J., Dronkers, N., & Gernsbacher, M. A. (2001). Language deficits, localization, and grammar: Evidence for a distributive model of language breakdown in aphasic patients and neurologically intact individuals. *Psychological Review, 108*, 759–788.

Edwards, J., & Lahey, M. (1998). Nonword repetitions in children with specific language impairment: Exploration of some explanations for their inaccuracies. *Applied Psycholinguistics, 19*, 279–309.

Elliott, L. L., Hammer, M. A., & Scholl, M. E. (1989). Fine-grained auditory discrimination in normal children and children with language-learning problems. *Journal of Speech and Hearing Disorders, 54*, 112–119.

Ellis Weismer, S. (1985). Constructive comprehension abilities exhibited by language-disordered children. *Journal of Speech and Hearing Disorders, 28*, 175–184.

Elman, J. L. (1993). Learning and development in neural networks: The importance of starting small. *Cognition, 48*, 71–99.

Francis, W. N., & Kučera, H. (1982). *Frequency analysis of English usage*. Boston: Houghton-Mifflin.

Gathercole, S. E., & Baddeley, A. D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory and Language, 29*, 336–360.

Gopnik, M., & Crago, M. (1991). Familial aggregation of a developmental language disorder. *Cognition, 39*, 1–50.

Grodzinsky, Y. (1990). *Theoretical perspectives on language deficits*. Cambridge, MA: MIT Press.

Hildebrant, N., Caplan, D., & Evans, K. (1987). The man$_i$ left t_i, without a trace: A case study of aphasic processing of empty categories. *Cognitive Neuropsychology, 4*(3), 257–302.

Joanisse, M. F., Manis, F., Keating, P., & Seidenberg, M. (2000). Language deficits in dyslexic children: Speech perception, phonology, and morphology. *Journal of Experimental Child Psychology, 77*(1), 30–60.

Joanisse, M. F., & Seidenberg, M. S. (1998). Specific language impairment: A deficit in grammar or processing? *Trends in Cognitive Sciences, 2*(7), 240–247.

Johnston, J. (1984). Specific language disorders in the child. In N. Lass, J. Northern, L. McReynolds, & D. Yoder (Eds.), *Handbook of speech–language pathology and audiology* (pp. 685–715). Philadelphia: B.C. Decker.

Johnston, J., & Ellis Weismer, S. (1983). Mental-rotation abilities in language-disordered children. *Journal of Speech and Hearing Research, 26*, 397–403.

Kamhi, A., Catts, H., Mauer, D., Apel, K., & Gentry, B. (1988). Phonological and spatial processing abilities in language- and reading-impaired children. *Journal of Speech and Hearing Disorders, 53*, 316–327.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language, 30*, 580–602.

Kirchner, D., & Klatzky, R. L. (1985). Verbal rehearsal and memory in language-disordered children. *Journal of Speech and Hearing Research, 28*, 556–565.

Leonard, L. B. (1987). Specific language impairment in children: A cross-linguistic study. *Brain and Language, 32*, 233–252.

Leonard, L. B., & Eyer, J. A. (1996). Deficits of grammatical morphology in children with specific language impairment and their implications for notions of bootstrapping. In J. L. Morgan & K. Demuth (Eds.), *Signal to syntax* (pp. 233–248). Mahwah, NJ: Lawrence Erlbaum.

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: A comment on Just & Carpenter (1992) and Waters & Caplan (1996). *Psychological Review, 109*, 35–54.

McCutchen, D., Dibble, E., & Blount, M. M. (1994). Phonemic effects in reading-comprehension and text memory. *Applied Cognitive Psychology, 8*(6), 597–611.

Miller, G. A. (1990). WordNet: An on-line lexical database. *International Journal of Lexicography, 3*, 235–312.

Montgomery, J. W. (1995). Sentence comprehension in children with specific language impairment: The role of phonological working memory. *Journal of Speech and Hearing Research, 38*, 187–199.

Nelson, L., Kamhi, A., & Apel, A. (1987). Cognitive strengths and weaknesses in language-impaired children: One more look. *Journal of Speech and Hearing Research, 28*, 556–564.

Pearlmutter, B. A. (1995). Gradient calculation for dynamic recurrent neural networks. A survey. *IEEE Transactions on Neural Networks, 6*(5), 1212–1228.

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

Plaut, D. C., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology, 10*(5), 377–500.

Rice, M. L., & Wexler, K. (1996). A phenotype of specific language impairments: Extended optional infinitives. In M. L. Rice (Ed.), *Toward a genetics of language* (pp. 215–237). Mahwah, NJ: Lawrence Erlbaum.

Seidenberg, M. S. (1997). Language acquisition and use: Learning and applying probabilistic constraints. *Science, 275*, 1599–1603.

St. John, M. F., & Gernsbacher, M. A. (1998). Learning and losing syntax: Practice makes perfect and frequency builds fortitude. In A. F. Healy & J. L. E. Bourne (Eds.), *Foreign language learning: Psycholinguistic experiments on training and retention* (pp. 231–255). Mawah, NJ: Erlbaum.

Tallal, P., Miller, S., & Fitch, R. H. (1995). Neurobiological basics of speech—a case for the preeminence of temporal processing. *Irish Journal of Psychology, 16*, 194–219.

Tallal, P., & Piercy, M. (1974). Developmental aphasia: Rate of auditory processing and selective impairment of consonant perception. *Neuropsychologia, 12*, 83–94.

Thibodeau, L. M., & Sussman, H. M. (1979). Performance on a test of categorical perception of speech in normal and communication disordered children. *Journal of Phonetics, 7*, 379–391.

Tomblin, J. B., & Pandich, J. (1999). Lessons from children with specific language impairment. *Trends in Cognitive Sciences, 3*(8), 283–285.

van der Lely, H. K. J. (1996). Specifically language impaired and normally developing children: Verbal passive vs. adjectival passive sentence interpretation. *Lingua, 98*, 243–272.

van der Lely, H. K. J., & Stollwerck, L. (1997). Binding theory and grammatical specific language impairment in children. *Cognition, 62*(1), 245–290.

Waters, G. S., & Caplan, D. (1985). The capacity theory of sentence comprehension: Critique of Just and Carpenter (1992). *Psychological Review, 103*, 761–772.

Williams, R. J., & Peng, J. (1990). An efficient gradient-based algorithm for on-line training of recurrent network trajectories. *Neural Computation, 2*, 490–501.