COGNITIVE SCIENCE Vol 23 (4) 1999, pp. 569–588 ISSN 0364-0213

A Probabilistic Constraints Approach to Language Acquisition and Processing

Mark S. Seidenberg

Maryellen C. MacDonald

University of Southern California

This article provides an overview of a probabilistic constraints framework for thinking about language acquisition and processing. The generative approach attempts to characterize knowledge of language (i.e., competence grammar) and then asks how this knowledge is acquired and used. Our approach is performance oriented: the goal is to explain how people comprehend and produce utterances and how children acquire this skill. Use of language involves exploiting multiple probabilistic constraints over various types of linguistic and nonlinguistic information. Acquisition is the process of accumulating this information, which begins in infancy. The constraint satisfaction processes that are central to language use are the same as the bootstrapping processes that provide entry to language for the child. Framing questions about acquisition in terms of models of adult performance unifies the two topics under a set of common principles and has important consequences for arguments concerning language learnability.

I. INTRODUCTION

Issues about acquisition have played a central role in the generative approach to language. That language involves innate, domain-specific types of knowledge is taken as one of the major findings to have emerged from the generative framework (Lightfoot, 1982). The “poverty of the stimulus” argument (Chomsky, 1965, 1986) is a cornerstone of the approach; for many years it has guided researchers toward certain types of explanations for how language is acquired (e.g., Universal Grammar) and away from others (e.g., learning). Over the past few years, an alternative approach to language acquisition has begun to emerge as an outgrowth of several developments outside of the linguistic mainstream. The sources for this new approach include the renewed interest in the statistical and probabilistic aspects of language on the part of many language researchers;

Direct all correspondence to: Mark S. Seidenberg, Neuroscience Program, University of Southern California, Los Angeles, CA 90089-2520; E-mail: marks@gizmo.usc.edu.

the connectionist approach to knowledge representation, learning, and processing; and research on learning in infants and young children. We will call this the "probabilistic constraints" approach, because its central idea is that acquisition and processing involve the use of multiple, simultaneous, probabilistic constraints defined over different types of linguistic and nonlinguistic information. This article will provide an overview of the approach, focusing on the relationship between acquisition and processing. Instead of asking how the child acquires competence grammar, we view acquisition in terms of how the child converges on adult-like performance in comprehending and producing utterances. This performance orientation changes the picture considerably with respect to classic issues about language learnability and provides a unified approach to studying acquisition and processing.[1]

II. GENERATIVE GRAMMAR AND PROBABILISTIC CONSTRAINTS

The questions that have been the focus of modern research on language are what is knowledge of a language, how is this knowledge acquired, and how is this knowledge used in comprehension and production (Chomsky, 1986).[2] The generative paradigm entails assumptions about how to find the answers to these questions and provides provisional answers to them (see Lightfoot, 1982, Atkinson, 1992 for reviews):

- Knowing a language involves knowing a grammar–a domain-specific form of knowledge representation that permits the creation of a nearly infinite set of well-formed utterances. The grammar specifies how language is structured at different levels of representation and provides a basis for distinguishing well-formed from ill-formed sentences.
- The grammar is a characterization of the knowledge of an idealized speaker-hearer. This "competence" assumption means that grammatical theory abstracts away from many of the factors that govern language use, including memory limitations, individual differences, facts about the computational system that implements the grammar, and so on.
- Poverty of the stimulus arguments suggest that this knowledge cannot be derived solely or even largely from experience. The input is impoverished insofar as children's knowledge of language eventually extends far beyond the range of utterances to which they are exposed; the input is overly rich insofar as it affords incorrect inductive generalizations that children never make; hence the input cannot be the source of core grammatical knowledge. Results from Gold (1967) and others suggest that grammar identification cannot be achieved unless there are strong constraints on the possible forms of grammatical knowledge.
- The view that children are born with knowledge of Universal Grammar is compatible with the results of behavioral studies indicating that various types of knowledge (e.g., empty categories; the binding principles) are present in children as young as they can be tested. It is also compatible with the observation that languages exhibit structures

for which there is simply is no evidence in the input (see Crain & Thornton, 1998, for review of both claims).

- This view will simultaneously account for facts about linguistic universals and it is compatible with other converging evidence (concerning, e.g., creolization; language acquisition under atypical circumstances; see Bickerton, 1984, and Landau & Gleitman, 1985, respectively).

III. AN ALTERNATIVE FRAMEWORK

The probabilistic constraints framework provides an alternative perspective on all of these issues.

An Alternative To Grammar

In the approach that we are advocating, knowing a language is not equated with knowing a grammar. Rather, we adopt the functionalist assumption knowledge of language as something that develops in the course of learning how to perform the primary communicative tasks of comprehension and production (see, e.g., Bates & MacWhinney, 1982). As a first approximation we view this knowledge as a neural network that maps between form and meaning and vice versa. Other levels of linguistic representation (e.g., syntax, morphology) are thought to be emergent structures that the network develops in the course of learning to perform these tasks. Allen and Seidenberg (1999) provide an illustration of this approach (for related studies, see Christiansen & Chater, 1999; Elman, 1990). They implemented an attractor network that was trained on two tasks: it was given a sequence of words as input and had to compute the words' semantics ("comprehension") or it was given semantic patterns as input and had to compute a sequence of words ("production"). The network was trained on sentences that exhibited a fairly broad range of syntactic structures; these sentences were based on ones that had been used by Linebarger, Schwartz, and Saffran (1983) in a classic study of aphasic language. The model was trained on 10 examples of each of 10 sentence types using a vocabulary of 97 words. It was then tested on novel examples of these structures using the same words. Allen and Seidenberg show that the model generalized to these novel forms quite accurately.

Even though this network's coverage of the grammar of English is quite limited, it serves to illustrate several components of the probabilistic constraints approach. The network developed a representation of (some aspects of) the structure of English in the course learning to produce and comprehend utterances, on the basis of exposure to well-formed examples. The representation of this knowledge—by the weights on connections between units—is not a grammar; it does not have the form of rewrite rules or constraints on tree structures that are seen in standard symbolic grammars or compute the same types of representations. Moreover, it encodes statistical aspects of the input that are excluded from standard generative grammars, information that on our view plays critical roles in language acquisition and use.

Grammars also provide a basis for distinguishing well- from ill-formed sentences, a task that assumes considerable importance because it has provided the main data for constructing grammatical theories. Allen and Seidenberg's (1999) model illustrates how this task can be construed within our framework. As part of their assessment of the model's capacity to generalize, Allen and Seidenberg tested it on novel grammatical and ungrammatical sentences such as "He came to my house at noon" and "He came my house at noon," respectively. The model performed qualitatively differently on these types of sentences, computing the semantic representations for words less accurately in ungrammatical ones. Allen and Seidenberg also report the results of testing the model on sentences that are comparable to Chomsky's classic "Colorless green ideas sleep furiously." The network's responses to such sentences patterned more closely with performance on other grammatical sentences than with ungrammatical ones.[3] Thus, within the limited fragment of English grammar on which the model was trained, it treated "grammatical" and "ungrammatical" sentences differently. These differences provide a basis for labelling sentences as well- or ill-formed even though the network itself is not a grammar.

An Alternative To "Competence"

Our approach does not share the competence orientation that is central to generative linguistics. We think that the competence-performance distinction encourages disregarding data that are actually essential to understanding basic characteristics of language. The competence approach uncontroversially excludes performance mishaps such as false starts, hesitations, and errors from the characterization of linguistic knowledge. However, it also excludes aspects of linguistic performance that are more central to the structure of utterances. It is assumed, for example, that language should be characterized independently of the perceptual and motor systems employed in language use; memory capacities that limit the complexity of utterances that can be produced or understood; and reasoning capacities used in comprehending text or discourse. Following the "Colorless green ideas sleep furiously" example, the competence theory also systematically excludes information about statistical and probabilistic aspects of language. How often particular structures are used and in what combinations are not seen as relevant to characterizing the essential nature of language.

It is clear from recent studies, however, that these aspects of language play enormously important roles in acquisition and processing (see, for example, MacDonald, Pearlmutter, & Seidenberg, 1994; Tanenhaus & Trueswell, 1995; Kelly, 1992; Saffran, Newport, & Aslin, 1996; Morgan & Demuth, 1996). The apparent complexity of language and its uniqueness vis á vis other aspects of cognition, which are taken as major discoveries of the standard approach, may derive in part from the fact that these "performance" factors are not available to enter into explanations of linguistic structure. Partitioning language into competence and performance and then treating the latter as a separate issue for psycholinguists to figure out has the effect of excluding many aspects of language structure and use from the data on which the competence theory is developed. What is left as a basis for characterizing language are the various abstract, domain specific, theoryinternal kinds of knowledge structures characteristic of grammatical theory. Linguists then ask how a grammar of this sort could be acquired. Given this approach to theory construction, it is not surprising that the resulting characterization of linguistic knowledge seems only remotely related to the child's experience.

The competence approach also promotes systematic overattributions about the nature of linguistic knowledge. It is well known that competence grammar is overly powerful relative to people's actual capacities to comprehend and produce utterances. The classic example is that the grammar permits unlimited amounts of center-embedding, as in the examples in (1). People's capacities to comprehend or produce such structures are quite limited: Only a single embedding (1a) is easily processed, and multiple levels of embedding are difficult or impossible to comprehend or produce.

1a. The cat [that the dog chased] is now sitting on the window sill.
1b. The cat that [the dog [that the girl bought] chased] is now sitting on the window sill.
1c. The cat that [the dog [that the girl [who Willard loves] bought] chased] is now sitting on the windowsill.

How to address the discrepancy between what the grammar allows and what people can comprehend or produce has been an issue since the earliest days of generative grammar (Chomsky, 1957). The standard solution is to allow such structures to be generated by the grammar but introduce extrinsic "performance constraints" to account for the difficulty of (1b–c) (see Gibson, 1998, for a recent example). Our approach is different: We are attempting to characterize a performance system that handles all and only those structures that people can. Performance constraints are embodied in the system responsible for producing and comprehending utterances, not extrinsic to it (MacDonald & Christiansen, 1999; Christiansen & Chater, 1999). This approach obviates the paradox created by a characterization of linguistic knowledge that generates sentences that people neither produce nor comprehend.

The fact that we are trying to account for peoples' attested capacities rather than the ones assumed by the competence approach is important. Our methodology involves implementing connectionist models that simulate detailed aspects of performance. In the past, some of the criticism of such models has focused on whether they can capture one or another aspect of grammatical competence, characterized in the usual idealized manner (e.g., Fodor & Pylyshyn, 1988). However, the appropriate way to assess our models is in terms peoples' performance, not the idealized characterization of linguistic knowledge that is competence grammar.

An Alternative to Grammar Identification

The probabilistic constraints approach also differs from generative grammar with regard to how the task confronting the language learner is characterized. The generative approach assumes that the task is *grammar identification*: the child has to converge on the knowledge structures that constitute the grammar. Questions about acquisition are framed in terms of problems in acquiring grammar, e.g., how the child could set parameters of the grammar based on limited experience, or how the child could recover from grammatical overgeneralizations in the absence of negative evidence. The alternative view is that the task the child is engaged in is learning to use language (see Seidenberg, 1997; see Bates & MacWhinney, 1982, for an earlier statement of this view). In the course of mastering this task, children develop various types of knowledge representations. The primary function of this knowledge is producing and comprehending utterances. A derived function is allowing the child to distinguish grammatical from ungrammatical sentences.

Consider the analogous problem of learning to read. The beginning reader must learn to recognize printed words; there are detailed theories of how this skill is acquired (e.g., Seidenberg & McClelland, 1989; Harm & Seidenberg, 1999). Once acquired, this knowledge can be used to perform many other tasks. One task that has been studied in scores of experiments is "lexical decision," which is judging whether a string of letters is a word or not. Even young readers can reliably determine that book is a word but NUST is not. Note, however, that the task confronting the beginning reader is not acquiring the capacity to make lexical decisions. By the same token, the task confronting the language learner is not acquiring the knowledge that supports making grammaticality judgments. In both cases, knowledge that is acquired for other purposes can eventually be used to perform these secondary tasks. Such tasks provide a useful way of assessing peoples' knowledge; however, it would be a mistake to define the acquisition problem in terms of the capacity to make such judgments.

This change in orientation from grammar identification to learning to use language has important consequences for standard poverty of the stimulus arguments. There is not sufficient space in this article to review all of the forms of argument and types of evidence that have been taken as evidence for the innateness of grammar. In brief, it turns out that many of the classic arguments rest on the assumption that the child's task is grammar identification, and these arguments simply no longer apply if the task is instead acquiring the performance system underlying comprehension and production. For example, the fact that the input contains both grammatical and ungrammatical utterances that are not labeled as such is far more relevant to grammar identification than it is to learning to comprehend and produce utterances. Given that the target is no longer construed as a grammar, our approach does not entail the assumptions that underlie the analyses of learnability by Gold (1967) and others. Other classic arguments concern how the child copes with noisy or variable input, issues for which connectionist learning principles provide a potential solution. One of the important properties of the algorithms used in training such models is their capacity to derive structural regularities from noisy or variable input. Similarly, the claim that "negative facts cannot be learned" (Crain, 1991) is contradicted by the behavior of even very simple networks such as the Allen and Seidenberg model described earlier. The network is trained on the basis of exposure to examples. Adjustments to the weights in response to these attested forms simultaneously provide evidence against structures that have not been observed because they do not occur in the language (e.g., because they are ungrammatical). The network nonetheless supports generalization because novel grammatical sentences share structure with previously encountered sentences. Thus, the Allen and Seidenberg (1999) model treated ungrammatical sentences such as "John came my house" differently from grammatical sentences such as "John came to my house" the first time it was exposed to them.

Evidence from Studies of Young Children

Two aspects of children's performance have been taken as particularly strong evidence for the innateness of grammatical principles. One is that children exhibit sensitivity to grammatical constraints at the youngest ages at which they can be tested. The other is that children show sensitivity to structures for which there is no evidence in the input (see Crain, 1991, for summaries of both arguments). With our change in theoretical orientation there is a need to re-examine these kinds of arguments. The validity of both claims depends on the theory of what is being learned (i.e., the target) and the theory of how children learn. The notion of what constitutes important evidence for learning a particular structure is not theory-neutral. Poverty of the stimulus arguments have been developed within a framework in which the target was grammar and the learning mechanism involved testing hypotheses about grammatical rules. Because these arguments have been very persuasive, research on language acquisition within the generative tradition has not examined relevant aspects of children's learning capacities very closely.

We have already suggested that our approach entails a different perspective on the nature of the target: The characterization of the target changes if we abandon the assumption that it is a grammar and various performance factors are permitted to enter into explanations of linguistic structure. Our approach also entails a different account of children's learning. Rather than involving hypothesis testing about grammatical rules, learning involves accumulating information about statistical and probabilistic aspects of language. We now know that learning about these aspects of language is very robust and begins at a very early age. For example, Saffran et al. (1996) demonstrated such learning in 8-month-olds. Other studies suggest that this kind of learning begins in utero: Newborns already show a preference for listening to speech in their mother's language (Moon, Panneton–Cooper, & Fifer, 1993). Results such as these raise questions about the extent to which children could have learned various aspects of language by the time they reach the age at which experimenters are able to test them (in studies such as Crain's, around 2.6–3.0-years-old). Claims about what cannot be learned by the child need to be assessed in terms of the kinds of statistical information available to the child and learning mechanisms that are able to extract non-obvious regularities from it.

In pointing out the potentially important role for statistical learning in language acquisition, we do not mean to assert that claims about the innateness of all aspects of linguistic knowledge are necessarily wrong. Our assertion is only that the probabilistic constraints framework provides strong motivation for re-examining claims about acquisition that are taken as givens within the generative approach. The same holds for other sorts of converging evidence thought to support the conclusion that grammar is innate. Linguistic universals, for example, are standardly explained by placing universal grammar in the brain of the child. However, there are other sources of constraint on the forms of languages that need to be considered. Languages may exhibit the properties they do because otherwise they could not be processed (given the nature of human perceptual and memory capacities), could not be learned (given the nature of human learning), would not fulfill particular communicative functions, or would not have evolved in the species. Together these considerations suggest a need to re-examine the both the nature of the biological endowment relevant to language and the evidence for it (see also Elman, Bates, Johnson, Karmiloff-Smith, Parisi & Plunkett, 1997, who reach similar conclusions).

IV. CONTINUITY BETWEEN ACQUISITION AND PROCESSING

In the generative approach, the target for the acquisition process is competence grammar, the idealized characterization of linguistic knowledge. In our approach, the target is provided by adult performance. Skilled users of language possess knowledge of how to comprehend and produce language, not just constraints on the well-formedness of utterances. The probabilistic constraints approach emphasizes the essential continuity between acquisition and processing, and the goal is to develop an integrated theory in which the same principles apply to both. This view suggests that it will be important for acquisition researchers to understand the nature of the adult processing system in order to understand how the child converges on it (Trueswell, Sekerina, Hill & Logrip, in press, provides an excellent example). Similarly, research on adult performance should benefit from an understanding of the acquisition process, particularly how constraints on learning shape the nature of the adult system. The observation that there is continuity between acquisition and processing may seem obvious, but in fact the generative approach has tended to treat the issues separately and to explain them by different principles. Acquisition is explained in terms of concepts such as the "maturation" of grammar (Wexler, 1990) or the setting of grammatical parameters (Lightfoot, 1991), whereas processing is explained in terms of a "parser" endowed with heuristics to aid ambiguity resolution (e.g., Frazier, 1987). This division is also seen in the interpretation of empirical results. Whereas accuracy on a task such as pointing to a picture that matches an auditory sentence is typically thought to reflect grammar development in the child, the identical task, when performed by adults, is generally taken as reflecting performance mechanisms rather than grammatical competence.

We do not have a comprehensive theory of all aspects of adult performance in hand, but a picture of the general character of the system is provided by recent constraint-based theories of adult language comprehension (MacDonald et al., 1994; Tanenhaus & Trueswell, 1995). The essential idea of the constraint-based approach is that comprehending or producing an utterance involves interactions among a large number of probabilistic constraints over different types of linguistic and non-linguistic information.[4] In much of this work, knowledge of language is represented in terms of a complex lexical network. MacDonald et al. (1994) review arguments for an enriched conception of the lexicon in which it encodes not merely the spellings, sounds, and meanings of words, but also information such as the argument structures associated with verbs. This view of the lexicon is broadly consistent with developments in syntactic theory, in which the lexicon has also assumed increasing importance. However, the use of a network representation also permits the encoding of information about the frequencies of occurrence for different types of information as well as higher-order statistics concerning combinations of elements. The representation that the network computes in comprehending a sentence is that which best satisfies the constraints of the language, which are encoded by the weights on connections between units. Comprehensive computational models are not yet available, but a number of researchers have developed simulations that implement small parts of the system (McRae, Spivey–Knowlton & Tanenhaus, 1998; Pearlmutter, Daugherty, MacDonald & Seidenberg, 1994; Tabor & Tanenhaus, this issue; Elman, 1990; Kim, Bangalore, & Trueswell, in press). In an interesting parallel development, recent work in computational linguistics has emphasized the importance of probabilistic lexical information in the construction of machine parsers (e.g., Charniak, 1997; Collins & Brooks, 1995). Although these parsers differ in fundamental ways from the kind of parallel constraint satisfaction networks that we take as the model for the human sentence processor, the shared interest in the role of probabilistic lexical constraints on the part of researchers holding very different theoretical orientations underscores the importance of this type of information.

To illustrate the general character of the constraint based approach, we will focus here on a few important aspects of it that are particularly relevant to acquisition. First, a network of this sort is both a representation of linguistic knowledge and a processing mechanism. As in other connectionist models, there is no strong distinction between the two. As a result, "performance constraints" (e.g., limits on the complexity of sentences that can be processed) are encoded by the same machinery as knowledge of language itself. This situation contrasts with other approaches in which knowledge of the language (grammar) is represented separately from the performance systems that make use of this knowledge (e.g., the parser, working memory). Second, the constraints that the model encodes are probabilistic rather than absolute. For example, a noun phrase (NP) at the start of a sentence is typically the agent of the action specified by the verb, but not always. Third, the interactions among constraints are nonlinear. Types of information that are not very constraining in isolation becomes highly constraining when considered together. For example, the probability that a sentence-initial NP is an agent goes up substantially if the NP is animate. Fourth, the levels of linguistic representation over which these computations occur are thought to emerge in the course of acquisition. For example, we think of morphological structure as an interlevel representation that emerges in a multilayer connectionist network that is computing relations among semantics, phonology, and the contexts in which words occur. Similarly, grammatical categories derive from several sources of correlated information, including meanings, phonological structure, and syntactic context (Kelly, 1992; Mintz, Newport, & Bever, 1998; Redington, Chater & Finch, 1998; Schuetze, 1997).

These principles have been explored in a number of recent studies of syntactic ambiguity resolution during language comprehension (see MacDonald et al., 1994, for review). Much of this research has focused on how semantic and syntactic properties of individual verbs affect the production and comprehension of syntactic structure. A simple example concerns the interpretation of prepositional phrases (PPs) in sentences such as

The spy saw the cop with the binoculars, in which the PP *with the binoculars* can be interpreted as modifying either the verb, such that it is an instrument of seeing, or the direct object NP, such that the spy saw the cop who had the binoculars rather than some other cop. This ambiguity is syntactic because the uncertainty of interpretation focuses not on the individual words but rather on the relationships between phrases, namely whether the phrase *with the binoculars* modifies a noun or verb.

Traditional accounts of syntactic ambiguity resolution have assumed that the fact that language comprehension is essentially immediate is incompatible with the use of multiple probabilistic constraints, which were thought to be too weak or not available early enough in processing to yield useful information. Instead, an initial interpretation of an ambiguity was thought to be computed based on a single metric, such as the complexity (e.g., Frazier, 1987) or frequency (Mitchell, Cuetos, Corley, & Brysbaert, 1995) of the alternative structures. These simple decision processes predict that people's initial interpretations of the PP ambiguity will frequently be wrong, as corpus data suggest that about 60% of the PPs modify the direct object NP, and the remaining 40% modify the verb (Hindle & Rooth, 1993; Collins & Brooks, 1995). Thus any simple structural metric will make errors in the initial interpretation of this ambiguity on the order of 40% of the time at best, and the Minimal Attachment metric favored by the best known structural account, the Garden Path model (Frazier, 1987), initially chooses the verb attachment and therefore makes about 60% errors. It is hard to reconcile these predictions with the observation that comprehenders resolve this ambiguity quite easily. It is also hard to understand why the human sentence processing mechanism would have converged on so inefficient a strategy.

Research within the constraint-based approach suggests that comprehenders avoid making errors 40–60% of the time by using other information that sentences provide. For example, many prepositions do not modify nouns and verbs with equal frequency (e.g. *of* almost exclusively modifies noun phrases, as in *glass of wine*, *president of the company*), and Collins and Brooks (1995) report that choosing the verb vs. noun attachment based solely on the frequency bias of the preposition will resolve the ambiguity correctly about 72% of the time. This is substantially better than the accuracy of any purely structural decision, and when several of these probabilistic lexical constraints are combined, the ambiguity resolution problem narrows greatly. Some other key lexical information comes from the verb in the sentence. Action verbs, which typically co-occur with a PP expressing the instrument or manner of the action (e.g., *cut with a knife*, *ate in a hurry*) promote the interpretation in which the PP modifies the verb, whereas perception verbs (see, hear, etc.) are not typically modified with manners or instruments, and they promote the noun-modification interpretation of the PP (Spivey–Knowlton & Sedivy, 1994; Taraban & McClelland, 1988). Spivey–Knowlton and Sedivy showed that an additional, weaker constraint is provided by the direct object NP: Indefinite NPs (e.g. *a dog*) promote the noun-modification interpretation of the PP, whereas definite NPs (*the dog*) promote the verb-modification interpretation. The verb and noun constraints combine in a nonlinear manner: the weaker noun definiteness constraint has little effect when a verb strongly promotes the verb-modification interpretation, but its effects can be clearly seen when the verb is one of perception. Finally, information from the surrounding discourse context also exerts some constraint, so that when the discourse creates a context in which more information is needed about the direct object NP (as when several cops were previously mentioned so that *The spy saw the cop*. . . does not pick out a unique referent in the discourse), then the noun-modification interpretation is promoted (Altmann & Steedman, 1988). Thus the system is one that is strongly guided by verb- and preposition-based constraints, which are in turn modulated by constraints from other lexical items and from the broader discourse context.

Other syntactic ambiguities have been shown to have a similar character. For example, the Main Verb/Reduced Relative ambiguity, which has a strong asymmetry in the frequencies of its alternative interpretations, is shown in (2).

2a. Temporary Main Verb/Reduced Relative Ambiguity: The three men arrested. . .
2b. Main Verb Interpretation: The three men arrested the fugitives who had escaped from the county jail.
2c. Reduced Relative Interpretation: The three men arrested in the parking lot had just escaped from the county jail.

The ambiguity centers on the role of an ambiguous verb, in this case *arrested*. In one interpretation, shown in (2b), *arrested* is the main verb of the sentence, and the preceding NP (*The three men*) is the agent of the action. In (2c), however, this NP is the patient of the action, and the verb *arrested* introduces a passive relative clause (called a "reduced" relative because the optional words *who were* are omitted from the start of the relative clause). A number of researchers have observed that interpretation of this ambiguity is strongly constrained by the frequency with which the ambiguous verb occurs in transitive and passive structures, of which reduced relative clauses are a special type (MacDonald, 1994; MacDonald et al., 1994; Trueswell, 1996). Interpretation of this structure is also constrained by combinatorial lexical information, such as the plausibility of the prenominal noun phrase (e.g. *The three men* in [2]) filling the agent or patient role of the verb (MacDonald, 1994; McRae, Spivey–Knowlton, & Tanenhaus, in press; Pearlmutter & MacDonald, 1992; Tabossi, Spivey–Knowlton, McRae & Tanenhaus, 1994; Trueswell, Tanenhaus & Garnsey, 1994). Additional constraints are provided by discourse context (Spivey–Knowlton, Trueswell, & Tanenhaus, 1993; Trueswell & Tanenhaus, 1991; Altmann & Steedman, 1988). Again, combination of constraints is nonlinear in that manipulations of noun agency or discourse context can successfully promote the rarer reduced relative interpretation only when properties of the ambiguous verb make this interpretation a viable one (MacDonald et al., 1994).

The picture that emerges from these and other studies of adult language comprehension is that adults have a vast amount of statistical information about the behavior of lexical items in their language, and that at least for English, verbs provide some of the strongest constraints on the resolution of syntactic ambiguities. Comprehenders know the relative frequencies with which individual verbs appear in different tenses, in active vs. passive structures, and in intransitive versus transitive structures, the typical kinds of subjects and objects that a verb takes, and many other such facts. This information is acquired through experience with input that exhibits these distributional properties. A verb's behavior is also closely related to its semantics and other properties specific to it. For example, part of the essence of the verb *arrested* includes the fact that it is frequently used in passive contexts in which the person arrested is the subject of the sentence. In our approach, this information is not some idiosyncratic fact in the lexicon isolated from "core" grammatical information; rather, it is relevant at all stages of lexical, syntactic and discourse comprehension.

"Bootstrapping" versus "Constraint Satisfaction"

Given this view of the adult system, one of the central questions in acquisition concerns how knowledge of probabilistic constraints is acquired. In fact, there is a significant body of acquisition research relevant to this question: studies of children's acquisition of verbs. This research (e.g., Gleitman, 1990; Pinker, 1989) has not been framed in terms of adult performance or connectionist models but in fact it quite compatible with our approach; a recent paper by Gillette, Gleitman, Gleitman, & Lederer (1998) makes the connection between this acquisition research and constraint satisfaction mechanisms explicit. Children learning English eventually come to know both the meanings of verbs and a complex set of conditions governing their occurrence in sentence structures (in some theories these are termed conditions on subcategorization). Consider for example some differences among the three common, semantically related verbs LOAD, POUR, and FILL. LOAD can appear in both of the locative constructions given in (3–4). POUR and FILL, in contrast, are each limited to one of the alternatives (5–8).

3. I loaded the bricks onto the truck.
4. I loaded the truck with bricks.
5. I poured the water onto the ground.
6. *I poured the ground with water.
7. *I filled the bricks onto the truck.
8. I filled the truck with bricks.

The behavior of these verbs is not idiosyncratic; each of them is representative of a class of semantically similar verbs that behave alike with respect to their participation in the two locative constructions:

load: pile, cram, scatter . . .
pour: drip, slop, spill, slosh . . .
fill: blanket, cover, flood, clog, coat . . .

Thus, there are systematic relationships between the syntax and semantics of verbs. Errors such as "fill the salt into the bear" (Bowerman 1982) make it clear that it takes time for the child to learn how verbs can be used. Exactly how the child exploits the correlations between syntax and semantics in acquiring this knowledge has been the subject of considerable controversy, however.

Pinker's (1989) theory emphasizes hypothesis testing and observational learning, in which the child observes the consistencies across situational contexts in which a verb is used to derive the verb's meaning. Aspects of a situation that are not consistent across uses of the verb would be expunged from a verb's lexical entry. The verb pour, for instance, would be observed in many contexts, including those in which a substance moves from a container into another container (pouring water into a bucket or a glass) as well as those in which a liquid moves from a container but not into another container (pouring water onto the ground). Since the endpoint of the liquid's motion is inconsistent across situations, the child would expunge any concrete specification of the endpoint of pouring events from the lexical entry for pour. Innate "linking rules," which specify the relation between an argument's place in a semantic representation and its place in a syntactic representation, would then allow a verb's syntactic privileges to be deduced directly from the resulting lexical entry. What will be consistent about events described by pour in the child's environment will be a liquid that moves from a container by the force of gravity, and this notion will form the basis of the word's lexical entry. Since this description includes an element (the liquid) that can be mapped to the innate notion "undergoer" or "patient," innate linking rules that map undergoers to grammatical direct object position will tell the child that the direct object of pour will be the liquid that undergoes movement. In this way, the child comes to know how to use pour (as pour and not as fill) because of what it means. This scenario assumes a probabilistic learning mechanism capable of keeping track of the similarities across situations, allowing the child to deduce which aspects of the local situation are relevant to the meaning of the verb, and relies on the existence of innate linking rules to explain how the child knows how to use the verb once its meaning is known.

Gleitman and her colleagues (e.g., Fisher, Gleitman, & Gleitman, 1991; Naigles, Gleitman & Gleitman, 1993) have argued that this observational procedure is too weak. Many verbs with opposite meanings can be used to describe the same situation, so paying attention only to the situation in which a verb is used will not tell the child which argument is subject and which object. For example, because every "chasing" event is also a "fleeing" event, the child has no way of knowing, from the situation alone, whether chase means "chase" or "flee." With pouring events there may be only one element of the situation that undergoes a change of location, and that will be the argument mapped onto direct object position. With chase and flee, however, there are two moving entities. Which will be the patient and thus linked to direct object? It could be the chaser, if the child views the event as one in which the fox is attempting to move to the location of the rabbit, or it could be the chasee, in which case the rabbit is attempting to distance himself from the fox.

Gleitman and colleagues' syntactic bootstrapping theory is that the forms in which a verb appears direct the child toward the correct meaning. Hearing a verb in a transitive frame, the child utilizes the linking rule that maps undergoers to direct object positions in reverse, such that the knowledge that an argument is the direct object tells the child that the argument is a patient, and that the verb's meaning is one that includes an undergoer of the type expressed by the argument. Other constructions provide other aspects of meaning. Hearing a verb in an intransitive frame, the child can deduce that the verb describes an action on the part of the subject. Hearing a ditransitive or to dative construction provides information that the verb contains the meaning of transfer. In other words, the child learns what a verb means because of how it is used.

Criticism of Gleitman's proposal has focused on the sufficiency of the syntactic bootstrapping mechanism. Pinker (1994), for example, has observed there are many more verbs than there are distinct syntactic contexts (subcategorization frames). Many aspects of verb meaning are not associated with difference in syntactic structure. For example, walking, strolling and sashaying differ semantically but behave the same way syntactically. This critique is based on a reading of Gleitman's proposal in which the child's use of syntactic information to bootstrap meaning restricts the use of other available sources of information, such as observation; however, the theory clearly does not limit the child in this way. Rather, it holds that the conjunction of information provided by the syntactic and discourse contexts in which verbs occur provides a basis for establishing their meanings. As Naigles et al. observed,

> We hold, with many others, that the information available from inspection of real-world contingencies for verb use, taken alone, is too variable and degenerate to fix the construals. Therefore there must be an additional convergent source of evidence that the child exploits in constructing a mental lexicon. One such potential source resides in the structural privileges of occurrence for the meaningfully distinct verb items. (p. 135).

Thus, the syntactic contexts in which a verb occurs constrain its meaning to a limited range of alternatives that can then be further disambiguated by observational data concerning the discourse context. In this way the use of syntactic information provides the necessary basis for overcoming the limits of observational learning, rather than replacing it in toto.

Similarly, Grimshaw (1990) has argued that because adjuncts cannot be distinguished from arguments, using syntax to learn the semantics of verbs will be an unreliable procedure. However, this kind of observation overlooks perhaps the central characteristic of constraint satisfaction mechanisms, which is that they rely on conjunctions of multiple sources of information that may be indeed be only partially constraining when taken in isolation.

Our view is that the "bootstrapping" mechanisms that have figured centrally in discussions of verb learning are components of a more general constraint satisfaction processing system that exploits correlations between different sources of information wherever they are found in the input. The view that lexical acquisition involves multiple kinds of bootstrapping is gaining currency among child language researchers (e.g., Morgan, 1986; Jusczyk, 1997; Morgan & Demuth, 1996; Gillette et al., 1998; Mintz et al., 1998) and there is as well a growing body of work that examines the extent to which children are attending to and learning from various statistical and probabilistic aspects of the input (see papers in Morgan & Demuth, e.g.). It is becoming quite clear that acquiring multiple types of probabilistic linguistic information is not an impossible task for the child; rather, even very young infants are picking up considerable information concerning distributional and sequential aspects of linguistic signals. The extent to which this type of learning provides the basis for various aspects of linguistic structure is unknown, but the similarity between these studies of learning and studies of adult processing is striking: "Bootstrapping" in acquisition is "constraint satisfaction" in adult processing; the former involves using correlations among different types of information to infer structure; the latter involves using such correlations to comprehend sentences.[5]

An illustration of the close theoretical linkage between the acquisition and adult processing can be found in a connectionist model of verb acquisition developed by Allen (1997). This model acquired verb semantics from the pairing of child-directed speech taken from the CHILDES corpus, particularly verb argument structure information, and information concerning the set of events accompanying the speech. For example, for the transitive sentence *Cathy broke the cup*, the model received the argument structure information that there were two arguments, one before the verb and one after, and that the verb was *break*. This information was paired with the interpretation that there were two participants, one the agent and one the patient, and that the event consisted of a breaking event. The goal of the modeling effort was to use knowledge acquired from exposure to these pairings to activate both appropriate argument role interpretations and the verb semantics for each utterance, including constructions on which the model had not been trained, such as *the toy broke*. Allen's model performed well, exhibiting both the ability to supply role interpretations for novel constructions and to activate appropriate verb semantics for novel verbs when given information about both the argument structure and the semantics of the noun arguments in the utterance. In other words, the model encodes the partial regularities of the system and "bootstraps" from incomplete data, enabling it to generalize to novel instances.

Allen showed that the model took advantage of a great deal of distributional information in the input to acquire its verb representations. This information included, in approximately descending order of importance, the frequency with which a verb was used, the set of constructions the verb appeared in, the frequency with which a verb was used in particular constructions, the semantic relation between a verb and other verbs used in similar constructions, the combined frequencies of related verbs, and the size of the set of semantically related verbs. These factors combined to form neighborhoods of verbs with semantically mediated privileges of co-occurrence. Allen's model therefore does not simply use syntactic information to acquire semantic information or information from the world to acquire syntactic information; rather, it simultaneously performs syntactic and semantic bootstrapping, using soft constraints from multiple sources to converge on syntactic and semantic representations of verbs. Moreover, though this process might be termed "bootstrapping" in the acquisition literature, it is equally an example of constraint-based language processing; the distributional information that the model acquired looks very similar to the constraints that are used in adult performance (MacDonald, 1999). This relationship is not accidental, because the task of the model was not acquisition per se but rather a simplified version of what human adult comprehenders do, namely assign a representation to each input sentence. In the course of assigning these representations, Allen's model passes activation across various levels of representation, and each utterance affects the weights between connections in the network. In this system, encoding of distributional information does not stem from some specialized acquisition mechanism but is rather an inevitable consequence of this kind of processing architecture when applied to the task of comprehension or production.

In summary, the characteristic of the adult system we take to be crucial is the rapid integration of multiple probabilistic constraints involving different types of information. This aspect of processing has been explored extensively in studies of syntactic ambiguity resolution and other aspects of "on-line" processing. The acquisition problem is how the child acquires this system. Specific issues that need to be addressed include how structures ("levels of linguistic representation") emerge in the course of acquisition; the properties of these representations, which may deviate from classical accounts; and which kinds of statistical information are acquired from experience.

We have described one area, verbs and their argument structures, in which this approach has been explored in some detail. Related work includes studies of the word segmentation problem (i.e., how children identify spoken words in continuous speech; Aslin et al., 1996; Christiansen et al., 1998) and how children identify the grammatical categories of words (Kelly, 1992; Mintz et al., 1998; Redington et al., 1998). All of these problems share important characteristics. In each case, researchers have identified a variety of sources of information in the input that might contribute to solving the problem. Using these cues to linguistic structure raises classic learnability questions: How does the child know which cues are the relevant ones? The cues are probabilistic rather than absolute; if a cue is not entirely reliable, how could it be useful? Moreover, what benefit could arise from combining several unreliable cues?

Answers to these questions are provided by coupling these observations about probabilistic cues with computational mechanisms that explain how cues can be identified and combined. Connectionist networks provide a useful tool in this regard. A network that is assigned a task such as identifying the boundaries between words acts as a discovery procedure: it will make use of whatever information in the input facilitates mastering the task This property is simply a consequence of using a learning algorithm that minimizes a cost function. From our perspective, the question "how does the child know which statistics to compute" is ill-formed. Certain information will be acquired given the nature of the architecture, the learning rule, and the input data. We can attempt to quantify this information statistically but the best characterization of what the child knows will be provided by examining the behavior of a network that simulates the child's behavior in appropriate detail. Examining the behavior of such networks also provides insights about the cue reliability question. These networks are not restricted to using a single type of information in solving a problem; their power derives from the capacity to combine multiple probabilistic cues efficiently. Individual cues that are not highly informative when taken in isolation may be highly constraining when taken in conjunction with each other.

V. PROSPECTS

People from the generative tradition often assume that the goal of this work is to show that language is not innate; thus it represents a return to a form of empiricism that most linguists thought was conclusively refuted years ago (see, e.g., Pinker, 1991; Wexler, 1991). Our assumption is that modern cognitive neuroscience and developmental neurobiology provide alternatives to both the Chomskian version of nativism and *tabula rasa* empiricism. It seems obvious that children are born with capacities to think, perceive, and learn that allow them to acquire language whereas chimpanzees do not. However, the hypothesis that children are born with knowledge of Universal Grammar may not be the only way to explain the facts about acquisition. At this point we do not have confidence in the validity of various arguments in support of innate Universal Grammar, but this does not mean the hypothesis is necessarily wrong. Our assumption is that we cannot validly infer what is innate without thoroughly investigating what can be learned. It is already clear that much more can be learned on the basis of the input available to the child, using simple and general learning mechanisms, than the generative approach assumes. How far these mechanisms take the child toward the adult steady-state is not known but needs to be investigated thoroughly.

Acknowledgments: Supported by grants NIMH MH PO1-47566, NIMH MH KO2 01188, and NSF SBR95-1270. We thank Morten Christiansen and Joseph Allen for helpful discussions.

NOTES

1. This article provides a brief overview of a theoretical framework that has emerged out of the work of many people. Space limitations preclude a comprehensive review of the literature; see MacWhinney (1998); Chater and Christiansen (1999); and Christiansen and Chater (this issue). We have attempted to develop an integrated framework that, while not sui generis, casts the issues in a particular way, one that derives in large part from considering language acquisition from the perspective provided by what is known about adult performance.
2. To this list we would now add "what is the neurobiological basis of language" and "how did language evolve in the species."
3. Allen and Seidenberg discuss why this result obtained. In brief, the crucial fact is that although the words in a sentence like "Colorless" have very low transition probabilities, they contain familiar sequences of semantic types such as <property property thing action manner>. At this level they are comparable to grammatical sentences with higher word transition probabilities. In contrast, ungrammatical sentences such as "Ideas colorless sleep furiously green" do not contain such sequences (thus, <thing property action manner property> does not occur in the language). The model was sensitive to both types of sequential information and therefore treated grammatical and ungrammatical sentences differently.
4. For early statements of this general idea see Marslen–Wilson and Welsh (1978) and Rumelhart (1977).
5. Pinker (1987) discussed bootstrapping mechanisms in language acquisition in terms of constraint satisfaction principles drawn from artificial intelligence. This is a symbolic notion of constraint satisfaction somewhat different than the one discussed here (see Macworth, 1977). However, Pinker's description of the verb learning problem as one involving multiple simultaneous constraints was rather accurate. He closed his article by noting that connectionist models were said to be able to efficiently compute complex interactions among constraints but that a connectionist researcher had told him that this was a difficult problem.

REFERENCES

Allen, J. (1997). Probabilistic constraints in acquisition. In A. Sorace, C. Heycock, & R. Shillcock (Eds.), *Proceedings of the GALA '97 Conference on Language Acquisition* (pp. 300–305). Edinburgh: University of Edinburgh Human Communications Research Center,.

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

Altmann, G. T. M., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition, 30,* 191–238.

Aslin, R.N., Woodward, J.Z., LaMendola, N.P., & Bever, T.G. (1996). Models of word segmentation in fluent maternal speech to infants. In J. L. Morgan, & K. Demuth (Eds.), *From signal to syntax*. Mahwah, NJ: Erlbaum.

Atkinson, M. (1992). *Children's syntax: An introduction to principles and parameters theory.* Oxford: Blackwell.

Bates, E., & MacWhinney, B. (1982). Functionalist approaches to grammar. In E. Wanner, & L. Gleitman (Eds.), *Language acquisition: The state of the art.* New York: Cambridge University Press.

Bickerton, D. (1984). The language bioprogram. *Behavioral and Brain Sciences, 7,* 173–221.

Bowerman, M. (1982). Evaluating competing linguistic models with language acquisition data: Implications of developmental errors with causative verbs. *Quaderni di Semantica, 3,* 5–66.

Charniak, E. (1997). Statistical techniques for natural-language parsing. *AI Magazine, 18,* 33–43.

Chater, N., & Christiansen, M. H. (in press). Connectionism and natural language processing. In S. Garrod & M. Pickering (Eds.), *Language processing*. Cambridge, MA: MIT Press.

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

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

Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. New York: Praeger.

Christiansen, M.H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes, 13,* 221–268.

Christiansen, M.H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science, 23,* 157–205.

Christiansen, M.H., & Chater, N. (1999). Connectionist natural language processing: The state of the art. *Cognitive Science, 23*, 417–437.

Collins, M., & Brooks, J. (1995). Prepositional phrase attachment through a backed-off model. In *Proceedings of the Third Workshop on Very Large Corpora*.

Crain, S. (1991). Language acquisition in the absence of experience. *Behavioral and Brain Sciences, 14,* 597–650.

Crain, S., & Thornton, R. (1998). *Investigations in universal grammar*. Cambridge, MA: MIT Press.

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

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff–Smith, A., Parisi, D., & Plunkett, K., *Rethinking innateness*. Cambridge, MA: MIT Press.

Fisher, C., Gleitman, H., & Gleitman, L. (1991). On the semantic content of subcategorization frames. *Cognitive Psychology*, *23,* 331–392.

Fodor, J., & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: A critical analysis. *Cognition, 28,* 3–71.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Eds.), *Attention and performance XII: The psychology of reading* (pp. 559–586). Hillsdale, NJ: Erlbaum.

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

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1998). Human simulations of vocabulary learning. Ms. submitted for publication.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition, 1,* 3–55.

Grimshaw, J. (1990). *Argument structure*. Cambridge, MA: MIT Press.

Gold, E. M. (1967). Language identification in the limit. *Information and Control, 16,* 447–474.

Harm, M., & Seidenberg, M. S. (1999). Phonology, reading, and dyslexia: A connectionist modeling perspective. *Psychological Review, 106,* 491–528.

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

Jusczyk, P. W. (1997). *The discovery of spoken language*. Cambridge, MA: MIT Press.

Kelly, M. H. (1992). Using sound to resolve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review, 99,* 349–364.

Kim, J., Bangalore, S., & Trueswell, J.C. (in press). The convergence of lexicalist perspectives in psycholinguistics and computational linguistics. In P. Merlo, & S. Stevenson (Eds.), *Lexical basis for sentence processing*. Philadelphia: John Benjamins.

Landau, B., & Gleitman, L. (1985). *Language and experience: Evidence from the blind child*. Cambridge, MA: Harvard University Press.

Lightfoot, D. (1982). *The language lottery*. Cambridge, MA: MIT Press.

Lightfoot, D. (1991). *How to set parameters*. Cambridge, MA: MIT Press.

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. (1999). Distributional information in language comprehension, production, and acquisition: Three puzzles and a moral. In B. MacWhinney (Ed.), *The emergence of language*. Mahwah: NJ: Erlbaum.

MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes, 9,* 157–201.

MacDonald, M. C., & Christiansen, M. H. (1999). Reassessing working memory: A comment on Just & Carpenter and Waters & Caplan. Manuscript submitted for publication.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101,* 676–703.

MacWhinney, B. (1987). *Mechanisms of language acquisition*. Hillsdale, NJ: Erlbaum.

MacWhinney, B. (1995). Connectionism and language learning. In N.C. Alexander (Ed.), Basic and applied perspectives on learning, cognition, and development. *The Minnesota Symposia on Child Psychology* (Vol. 28, pp. 33–57). Mahwah, NJ: Erlbaum.

MacWhinney, B. (1998). Models of the emergence of language. *Annual Review of Psychology, 49,* 199–227.

Macworth, A. (1977). Consistency in networks of relations. *Artificial Intelligence, 8,* 99–118.

Marslen–Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10,* 29–63.

McRae, K., Spivey–Knowlton, M., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language, 38,* 283–312.

Mintz, T. H. (1996). *The roles of linguistic input and innate mechanisms in children's acquisition of grammatical categories,* unpublished Ph.D. dissertation. Rochester, NY: University of Rochester.

Mintz, T. H., Newport, E. L., & Bever, T. G. (1998). The distributional structure of grammatical categories in speech to young children. Manuscript submitted for publication.

Mitchell, D. C., Cuetos, F., Corley, M. M. B., & Brysbaert, M. (1995). Exposure-based models of human parsing: Evidence for the use of coarse-grained (nonlexical) statistical records. *Journal of Psycholinguistic Research, 24,* 469–488.

Moon, C., Panneton-Cooper, R. P., & Fifer, W. P. (1993). Two-day infants prefer their native language. *Infant Behavior Development, 16,* 495–500.

Morgan, J. (1986). *From simple input to complex grammar*. Cambridge, MA: MIT Press.

Morgan, J., & Demuth, K. (1996). *Signal to syntax: Bootstrapping from speech to grammar in early acquisition*. Mahwah, NJ: Erlbaum.

Naigles, L.G., Gleitman, H., & Gleitman, L. (1993). Children acquire word meaning components from syntactic evidence. In E. Dromi (Ed.*), Language and cognition: A developmental perspective*. Norword, NJ: Ablex.

Pearlmutter, N. J., & MacDonald, M. C. (1992). Plausibility and syntactic ambiguity resolution. In *Proceedings of the 14th Annual Conference of the Cognitive Society*, (pp. 498–503). Hillsdale, NJ: Erlbaum.

Pearlmutter, N. J., Daugherty, K. G., MacDonald, M. C., & Seidenberg, M. S. (1994). Modeling the use of frequency and contextual biases in sentence processing. In *Proceedings of the 16th Annual Conference of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA: Harvard University Press.

Pinker, S. (1987). The bootstrapping problem in language acquisition. In B. MacWhinney (Eds.), *Mechanisms of language acquisition*, (pp. 339–441). Hillsdale, NJ: Erlbaum.

Pinker, S. (1989). *Learnability and cognition*. Cambridge, MA: MIT Press.

Pinker, S. (1991). Rules of language. *Science, 253,* 530–534.

Pinker, S. (1994). How could a child use verb syntax to learn verb semantics. In L. Gleitman, & B. Landau (Eds.), *The acquisition of the lexicon* (pp. 377–410). Cambridge, MA: MIT Press.

Redington, M., Chater, N., & Finch (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science, 22,* 426–469.

Rumelhart, D. E. (1977). Toward an interactive model of reading. In S. Dornic (Ed.), *Attention and performance VI* (pp. 220–235). Hillsdale, NJ: Erlbaum.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Statistical learning by 8-month-old infants. *Science, 274,* 1926–1928.

Schuetze, H. (1997). *Ambiguity resolution in language learning: Computational and cognitive models*. Stanford, CA: CSLI Publications.

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

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review., 96,* 523–568.

Seidenberg, M. S., Allen, J., & Christiansen, M.H. (1998). Probabilistic constraints in language acquisition. In A. Sorace, C. Heycock, & R. Shillcock (Eds.), *Proceedings of the GALA '97 Conference on Language Acquisition* (pp. 382–387). Edinburgh: University of Edinburgh Human Communications Research Center.

Spivey–Knowlton, M. J. &. Sedivy, J. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition*, *55,* 227–267.

Spivey–Knowlton, M. J., Trueswell, J. C., & Tanenhaus, M. K. (1993). Context and syntactic ambiguity resolution. *Canadian Journal of Experimental Psychology, 47,* 276–309.

Tabor, W., & Tanenhaus, M. (1999). Dynamical models of sentence processing. *Cognitive Science*, *23,* 000–000.

Tabor:W., Juliano, C., & Tanenhaus, M. K. (1997). Parsing in a dynamical system: An attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Language and Cognitive Processes, 12,* 211–271.

Tabossi, P., Spivey–Knowlton, M. J., McRae, K., & Tanenhaus, M. K. (1994). Semantic effects on syntactic ambiguity resolution: Evidence for a constraint-based resolution process. In C. Umilta, & M. Moscovitch (Eds.), *Attention and performance XV* (pp. 589–616). Cambridge, MA: MIT Press.

Tanenhaus, M., & Trueswell, J. (1995). Sentence Processing. In P. Eimas, & J. L. Miller (Eds.), *Handbook of perception and cognition: Language*. New York: Academic Press.

Taraban, R., & McClelland, J. L. (1988). Constituent attachment and thematic role assignment in sentence processing: Influences of content-based expectations. *Journal of Memory and Language, 27,* 597–632.

Trueswell, J. C. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language, 35,* 566–585.

Trueswell, J.C., Sekerina, I., Hill, N. & Logrip, M. L. (in press). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic disambiguation. Manuscript submitted for publication. *Journal of Memory and Language, 33,* 285–318.

Wexler, K. (1990). Innateness and maturation in linguistic development. *Developmental Psychobiology, 23,* 645–660.

Wexler, K. (1991). On the argument from the poverty of the stimulus. In A. Kasher (Ed.), *The Chomskyan turn*. Oxford: Blackwell.