defined by the sensory weights, that is, the model will climb the surface defined by the value function V.

26. A. R. Damasio, *Descartes' Error* (Putnam, New York, 1994); A. Bechara, A. R. Damasio, H. Damasio, S. Anderson, *Cognition* **50**, 7 (1994).

27. S. P. Perrett, B. P. Ruiz, M. D. Mauk *J. Neurosci.* **13**, 1708 (1993); J. L. Raymond, S. G. Lisberger, M. D. Mauk *Science* **272**, 1126 (1996).

28. S. Grossberg, *Math. Biosci.* **15**, 253 (1972); R. L. Solomon and J. D. Corbit, *Psychol. Rev.* **81**, 119 (1974); S. Grossberg, *ibid.* **89**, 529 (1982).

29. W. Schultz and R. Romo, *J. Neurophysiol.* **63**, 607 (1990); T. Ljungberg, P. Apicella, W. Schultz, *ibid.* **67**, 145 (1992); J. Mirenowicz and W. Schultz, *Nature* **379**, 449 (1996).

30. W. Schultz, P. Apicella, T. Ljungberg, *J. Neurosci.* **13**, 900 (1993).

31. T. Aosaki *et al.*, *ibid.* **14**, 3969 (1994); A. M. Graybiel *Curr. Opin. Neurobiol.* **5**, 733 (1995); *Trends Neurosci.* **18**, 60 (1995). Recent models of sequence generation in the striatum use fluctuating dopamine input as a scalar error signal [G. S. Berns and T. J. Sejnowski, in *Neurobiology of Decision Making*, A.

Damasio, Ed. (Springer-Verlag, Berlin, 1996), pp. 101–113.

32. T. F. Freund, J. F. Powell, A. D. Smith, *Neuroscience* **13**, 1189 (1984); Y. Smith, B. D. Bennett, J. P. Bolam, A. Parent, A. F. Sadikot, *J. Comp. Neurol.* **344**, 1 (1994).

33. C. Cepeda, N. A. Buchwald, M. S. Levine, *Proc. Natl. Acad. Sci. U. S. A.* **90**, 9576 (1993).

34. J. R. Wickens, A. J. Begg, G. W. Arbuthnott, *Neuroscience* **70**, 1 (1996).

35. P. S. Goldman-Rakic, C. Leranth, M. S. Williams, N. Mons, M. Geffard, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 9015 (1989); T. Sawaguchi and P. S. Goldman-Rakic, *Science* **251**, 947 (1991); G. V. Williams and P. S. Goldman-Rakic, *Nature* **376**, 572 (1995).

36. R. J. Dolan *et al.*, *Nature* **378** 180 (1995).

37. P. R. Montague, C. D. Gancayco, M. J. Winn, R. B. Marchase, M. J. Friedlander, *Science* **263**, 973 (1994). The mechanistic suggestion requires that local cortical activity (presumably glutamatergic) increases the sensitivity of nearby dopamine terminals to differences from baseline in spike production

along their parent axon. This may result from local increases in nitric oxide production. In this manner, baseline dopamine release remains constant in inactive cortical areas while active cortical areas feel strongly the effect of increases and decreases in dopamine delivery due to increases and decreases in spike production along the parent dopamine axon.

38. We thank A. Damasio and T. Sejnowski for comments and criticisms, and C. Person for help in generating figures. The theoretical work received continuing support from the Center for Theoretical Neuroscience at Baylor College of Medicine and the National Institutes of Mental Health (NIMH) (P.R.M.). P.D. was supported by Massachusetts Institute of Technology and the NIH. The primate studies were supported by the Swiss National Science Foundation, the McDonnell-Pew Foundation (Princeton), the Fyssen Foundation (Paris), the Fondation pour la Recherche Midicale (Paris), the United Parkinson Foundation (Chicago), the Roche Research Foundation (Basel), the NIMH (Bethesda), and the British Council.

Language Acquisition and Use: Learning and Applying Probabilistic Constraints

Mark S. Seidenberg

What kinds of knowledge underlie the use of language and how is this knowledge acquired? Linguists equate knowing a language with knowing a grammar. Classic "poverty of the stimulus" arguments suggest that grammar identification is an intractable inductive problem and that acquisition is possible only because children possess innate knowledge of grammatical structure. An alternative view is emerging from studies of statistical and probabilistic aspects of language, connectionist models, and the learning capacities of infants. This approach emphasizes continuity between how language is acquired and how it is used. It retains the idea that innate capacities constrain language learning, but calls into question whether they include knowledge of grammatical structure.

Modern thinking about language has been dominated by the views of Noam Chomsky, who created the generative paradigm within which most research has been conducted for over 30 years (*1*). This approach continues to flourish (*2*), and although alternative theories exist, they typically share Chomsky's assumptions about the nature of language and the goals of linguistic theory (*3*). Research on language has arrived at a particularly interesting point, however, because of important developments outside of the linguistic mainstream that are converging on a different view of the nature of language. These developments represent an important turn of events in the history of ideas about language.

The Standard Theory

The place to begin is with Chomsky's classic questions (*4*): (i) what constitutes knowledge of a language, (ii) how is this knowledge acquired, and (iii) how is it put

to use? The standard theory provides the following answers (*1–5*).

In answer to the first question, what one knows is a grammar, a complex system of rules and constraints that allows people to distinguish grammatical from ungrammatical sentences. The grammar is an idealization that abstracts away from a variety of so-called performance factors related to language use. The Competence Hypothesis is that this idealization will facilitate the identification of generalizations about linguistic knowledge that lie beneath overt behavior, which is affected by many other factors. Many phenomena that are prominent characteristics of language use are therefore set aside. The clear cases that are often cited in separating competence from performance include dysfluencies and errors. In practice, however, the competence theory also excludes other factors that affect language use, including the nature of the perceptual and motor systems that are used; memory capacities that limit the complexity of utterances

that can be produced or understood; and reasoning capacities used in comprehending text or discourse. The competence theory also excludes information about statistical and probabilistic aspects of language—for example, the fact that verbs differ in how often they occur in transitive and intransitive sentences ("John ate the candy" versus "John ate," respectively), or the fact that when the subject of the verb "break" is animate, it is typically the agent of the action, but when it is inanimate, it is typically the entity being broken (compare "John broke the glass" with "The glass broke"). That this information should be excluded was the point of Chomsky's famous sentence "Colorless green ideas sleep furiously" and the accompanying observation that, "I think that we are forced to conclude that . . . probabilistic models give no particular insight into some of the basic problems of syntactic structure" (*6*). Finally, the competence theory also disregards the communicative functions of language and how they are achieved. These aspects of language are acknowledged as important but considered separable from core grammatical knowledge.

The grammar's essential properties include generativity (it can be used to produce and comprehend an essentially infinite number of sentences); abstractness of structure (it uses representations that are not overtly marked in the surface forms of utterances); modularity (the grammar is organized into components with different types of representations governed by different principles); and domain specificity (language exhibits properties that are not seen in other aspects of cognition; therefore, it cannot be an expression of general capacities to think and to learn).

The second question regarding language

Neuroscience Program, University of Southern California, Los Angeles, CA 90089–2520, USA. E-mail: marks@gizmo.usc.edu

acquisition is framed in terms of how the child acquires a grammar with these properties. This problem is said to be solvable only if the child possesses considerable grammatical knowledge in advance of experience, that is, from birth. The basis for this conclusion is a set of observations about the nature of acquisition termed the "poverty of the stimulus" argument (*1, 5*).

Several aspects of the child's experience make grammar learning a difficult inductive problem. The input to the child is degenerate, consisting of both grammatical and ungrammatical sentences that are not labeled as such. It is also variable: children are exposed to different samples of utterances but converge on the same grammar. The input does not include reliable negative evidence, that is, evidence about which structures are not allowed by the grammar; logical arguments suggest that in the absence of such evidence there must be strong innate constraints on the possible forms of grammars (*7*). Finally, languages exhibit properties for which there is no positive evidence in the input. The claim here is that there cannot be any overt evidence for the kinds of abstract underlying structures characteristic of grammatical theory (*8*). That essential aspects of grammar are innate—represented in the brain of the neonate—is said to be the only viable explanation for how languages could be learned so rapidly yet under such impoverished conditions. This hypothesis simultaneously accounts for universal properties of languages.

These classic learnability arguments are thought to severely limit the role of experience in acquisition. The input is said to be too poor in that what people know extends far beyond the sample of utterances to which they are exposed, but simultaneously, the input is said to be too rich insofar as it affords incorrect inductive generalizations that children never make. The input therefore cannot be the source of core aspects of linguistic knowledge. The major thrust of the "learnability" approach is that language is essentially unlearnable and therefore must be a kind of human instinct (*9*). The role of experience is merely to allow the child to acquire a lexicon and set some language-specific parameters (*10*).

The third question, regarding how language is used, is traditionally framed in terms of how the grammar is used in producing and comprehending utterances. Understanding performance is thought to presuppose having a substantive theory of the structure of language in hand, and, in fact, much more research has focused on working out the details of the competence grammar than on studying how it is used. The validity of this assumption about the

epistemological precedence of grammar is an issue to which I return later.

Other Developments

This elegant account of the nature of language is complicated by two major developments outside the linguistic mainstream.

Emergence of connectionism. The connectionist or neural network approach is being used to address many issues in the cognitive and neurosciences (*11*). Connectionist theories attempt to explain behavioral phenomena in terms of networks of simple, neuronlike processing units. Such networks are typically implemented as computational models that learn to perform tasks, such as recognizing words or faces on the basis of exposure to examples. Learning involves gradual changes to the weights on connections between units that determine patterns of activation in the network. The behavior of the system depends on its architecture (the number and types of units, the pattern of connectivity between them, and the equations governing the spread of activation), the structure implicit in the ensemble of training examples, and the learning rule used to adjust the weights. Connectionist concepts are also being used to develop theories of neuronal function, particularly how collections of neurons come to perform complex tasks (*12*). The use of the same theoretical vocabulary to explain both observable behavior and its neurophysiological basis is an important development, presenting the possibility of a unified account of the two.

Connectionism makes available a rich set of tools and ideas that are potentially relevant to understanding language. First, it incorporates a novel form of knowledge representation that provides an alternative to equating knowledge of a language with grammar. Whereas grammatical theory focuses on characterizing the structure of language, the connectionist approach focuses on how networks come to perform tasks such as comprehension and production (*13*). Such networks do not directly incorporate or implement traditional grammars; rather, grammatical theory can be seen as providing higher level generalizations about network behavior. These descriptions are only approximately correct because they abstract away from details of the underlying computational mechanism (*14*).

Second, the learning procedures used in training these networks (*15*) represent a significant advance over the simpler ones that were the focus of earlier criticism (*16*). An important property of these algorithms is their ability to derive structural regularities from relatively noisy input data. This

property is relevant to how the child acquires language under naturalistic conditions and has important implications for poverty-of-the-stimulus arguments, which are discussed below.

Third, these learning principles are typically used in training multilayer networks that provide a model of how people develop abstract representations of language and other complex phenomena (*17*). Such underlying representations are created as networks find solutions to the tasks they are learning to perform (*18*). This possibility was not afforded by earlier theories in which learning merely involved the creation of stimulus-response chains (*19*).

Fourth, such models provide an alternative account of generalization. The standard approach emphasizes the observation that the grammar, though finite, can be used to generate an infinite set of sentences. The capacity to generalize has provided the classic evidence that knowledge of a language involves rules (*20, 21*). Connectionist models provide an alternative account: although the weights are set on the basis of experience, they can be used to process novel forms. Networks trained on the pronunciations of written words in English, for example, can generalize to novel forms to which the network has not been exposed (*22*).

Finally, these networks incorporate a powerful processing mechanism. Processing involves the parallel satisfaction of multiple, simultaneous, probabilistic constraints (*23*). The network is trained through exposure to a large number of examples. The learning algorithm allows it to represent the statistical structure of the input. The weights can then be seen as encoding a large number of probabilistic constraints derived from prior experience. These constraints include simple and complex contingencies between different types of information. The network's output on a given trial simultaneously satisfies all of these intersecting constraints.

Renewed interest in statistical and probabilistic aspects of language. Although Chomsky's remarks in 1957 about the limits of statistical approaches to language largely extinguished interest in the topic for many years (*24*), several converging developments have led to a strong revival of interest in these aspects of language in the 1990s.

One factor has been the technical advances that permit more serious investigation of these issues. The large language samples and computational resources required for deriving robust estimates of language statistics have only recently become available to researchers. An important step was the creation of major archives of adult and child language corpora (*25*). Interest

has also been stimulated by significant advances in the use of statistical approaches in applied areas such as automatic speech recognition and machine translation (26).

Among psycholinguists, there is an increasing appreciation of the relevance of this kind of information to language acquisition and use. Studies of adult language comprehension have focused on processes that occur "on-line" in listening or reading (27). Sentences exhibit multiple ambiguities, creating a complex computational problem for the perceiver. In sentence 1 below, for example, the word "plane" can refer to an airplane, a geometric element, or a tool, and the word "left" can be either a direction or the past tense of leave. These lexical ambiguities coexist with syntactic ambiguities spanning several words. In sentence 1, the comprehender must determine that "left" is a verb, that it is in the active voice, that "the plane" is the entity doing the leaving, and that "for" introduces the destination (the east coast) of the leaving event. In sentence 2, "left" is again a verb but in the passive voice, "the plane" is the entity being left by some unspecified agent, and "for" indicates a beneficiary (the reporter) of the leaving event. Such ambiguities are a pervasive aspect of natural language and must eventually be resolved for comprehension to occur.

1) The plane left for the East Coast.
2) The plane left for the reporter was missing.
3) The note left for the reporter was missing.

Psycholinguistic experiments indicate that ambiguities are resolved by rapidly exploiting a variety of probabilistic constraints derived from previous experience with language and the world (28). Some constraints relevant to sentences 1 and 2 include the fact that the vehicle meaning of "plane" is more frequent than its other meanings, and that the verb "left" is used more often in active rather than passive constructions. In addition to this first-order frequency information, there are combinatorial constraints concerning the probable relations between words. For example, the verb interpretation of "left" is promoted in sentences 1 and 2 by the fact that "plane" is not a plausible modifier of the noun "left," so that "the plane left" is not a sensible noun phrase. Although sentence 3 above has the same syntactic structure as sentence 2, it is more easily comprehended because it is much more plausible for a note to be left than to leave (29).

Connectionist principles provide an account of how these kinds of information can be efficiently acquired, represented, and exploited. The simultaneous satisfaction of multiple probabilistic constraints provides a way of rapidly and efficiently combining different sources of information. Importantly, the interactions among constraints are nonlinear: bits of information that are not very constraining in isolation become quite informative when taken in conjunction with other bits of information. Both "the plane" and "left" are highly ambiguous insofar as both have multiple common meanings involving different parts of speech. Yet the conjunction of the two makes it very likely that "the plane" is an airplane and "left" is the past tense of leave. It is the capacity to exploit multiple sources of probabilistic information that allows the network, and by hypothesis the comprehender, to rapidly converge on correct interpretations (30).

Given this view of adult performance, the principal question about acquisition is how the child develops a system with these characteristics. Seen in this light, recent findings concerning the remarkable learning abilities of infants are enormously exciting (31). Such studies show that infants naturally and automatically encode statistical aspects of caregiver speech without overt guidance or reward (32). Much of the evidence has emerged from studies of infants' acquisition of speech (31). Current research focuses on how infants use such analyses of the input to solve problems such as identifying words in continuous speech and their grammatical functions. Such learning apparently begins in utero, because newborns prefer listening to speech in the mother's language as opposed to other languages (33). Learning based on the frequencies and distributions of environmental events is emerging as an essential aspect of cognitive development (34).

To summarize, theories of competence grammar have excluded various aspects of language use in pursuit of more fundamental generalizations. Facts about language are explained in terms of abstract, domain-specific knowledge structures that are remotely related to the child's experience. Language therefore gives the appearance of being unlearnable and unrelated to other aspects of cognition. More recent studies suggest that acquisition and processing are driven by exactly the kinds of information that competence grammar has traditionally excluded. On the newer view, the child's task is learning to use language, not grammar identification. This performance orientation emphasizes the continuity between acquisition and skilled processing: the same mechanisms are involved in acquiring language as in using it. In acquisition research, the informal notion of "bootstrapping" has been used to describe how children use correla-

tions between different aspects of language to infer structure (35). Connectionist networks provide a generalization and formalization of this notion in terms of the satisfaction of multiple probabilistic constraints. This mechanism plays a key role in the child's entry into language, providing the basis for identifying words, their meanings and grammatical functions, and the kinds of structures they participate in (36). The bootstrapping mechanisms that provide entry into language for the child are in the adult the constraint satisfaction mechanisms used in skilled comprehension and production.

Thus, the newer approach attempts to explain language in terms of how it is acquired and used rather than an idealized competence grammar. The idea is not merely that competence grammar needs to incorporate statistical and probabilistic information; rather, it is that the nature of language is determined by how it is acquired and used and therefore needs to be explained in terms of these functions and the brain mechanisms that support them. Such performance theories are not merely the competence theory plus some additional assumptions about acquisition and processing; the approaches begin with different goals and end up with different explanations for why languages have the properties that they have.

Current Issues

The framework that I have described is new, and very little research has as yet been published; much of what there is consists of conference papers and doctoral theses. Little of the work to date has addressed the kinds of phenomena that have been the focus of linguistic theorizing over the past several decades, so the range of phenomena that the framework will be able to explain is still to be determined. What is already clear, however, is that the approach provides a plausible basis for reopening classic questions about language acquisition. This can be seen by examining the well-studied problem of how children acquire knowledge of verbs. Consider sentences 4 through 9 (asterisks indicate ungrammatical sentences):

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

Speakers of a language eventually come to know both the meanings of verbs and a complex set of conditions governing their

occurrence in sentences. The three semantically similar verbs "load," "pour," and "fill" are associated with different syntactic privileges in English. "Load" can appear in both locative (sentence 4) and "with" (sentence 5) constructions; "pour" and "fill" each allow one of the two alternatives. How children acquire this knowledge has been the focus of considerable debate (37, 38). Standard poverty-of-the-stimulus arguments suggest that the problem is quite hard (39). Children are exposed to erratic samples of such structures yet rapidly converge on the right generalizations. Parents do not provide reliable negative evidence: They do not explicitly say that sentence 6 is allowed but sentence 7 is not. Thus, the child, in contrast to the linguist, does not have access to all of the structures in sentences 4 through 9. When ungrammatical utterances happen to occur, they are not labeled as such. Children nonetheless have to learn the conditions under which verbs do not enter into specific constructions. Although errors are made in the course of acquisition (for example, "Daddy, disappear the coin"), children's behavior seems highly constrained insofar as there are many incorrect generalizations that could be made but are not. The mechanisms that prevent children from accepting sentences 7 or 8 by analogy to sentences 4 and 5 have been a source of mystery.

One important hint is provided by the observation that there are "neighborhoods" of verbs with similar privileges of occurrence (38, 40): "load," for example, patterns with verbs such as "pile," "cram," "spray," and "scatter"; "pour" patterns with "drip," "slop," and "slosh"; "fill" with "blanket," "cover," and "coat." Being able to discern these subgroups would be quite beneficial to language learners because it would liberate them from having to be exposed to examples of every verb in every construction. A child might not have heard a sentence such as sentence 6 but would know that it is felicitous on the basis of exposure to sentences containing semantically related verbs that pattern the same way. Similarly, the fact that a group of related verbs all occur in the same syntactic structures might provide a kind of "indirect negative evidence" that other verbs do not behave this way (41).

Unfortunately, merely describing the conditions governing the use of verbs in English has proved quite difficult, and how the child would ever converge on the relevant generalizations on the basis of the limited information provided by parental speech has been unclear. The verbs within a group tend to overlap in meaning (for example, pile and cram); however, they also differ in some respects (for example, spray-

ing and scattering involve dispersion of the affected element but piling and cramming do not), and there are similarities between verbs that pattern differently (for example, pour and fill). The traditional approach is to try to state these regularities as rules and then determine how the child could infer them on the basis of the limited information available. Framed in this way, the task is so complex that it appears unsolvable without innate grammatical knowledge (38).

From the probabilistic constraints perspective, these phenomena represent a classic constraint satisfaction problem. There are several sources of systematic though probabilistic information governing verbs and the structures in which they occur. These include facts about the semantics of verbs, such as how much they overlap; correlations between verbs and both the syntactic structures they license and the kinds of events they describe; and item-specific idiosyncrasies that are the result of processes of language change or historical accident. Connectionist networks are well suited to capturing systems with this character. Importantly, a network configured as a device that learns to perform a task such as mapping from sound to meaning will act as a discovery procedure, determining which kinds of information are relevant. Evidence that such models can encode precisely the right combinations of probabilistic constraints is provided by Allen (42), who implemented a network that learns about verbs and their argument structures from naturalistic input.

These phenomena are the focus of ongoing research, and the issues are by no means settled. However, even at this early stage several implications are coming into focus. First, it is clear that standard poverty-of-the-stimulus arguments have less bearing on solving this kind of problem than was assumed before the development of models such as Allen's. The degeneracy of the input (the fact that it includes ungrammatical utterances) has little impact because the model is not performing grammar identification. Input variability is not crucial because the model's performance on any given verb does not solely depend on experience with it; the model benefits from exposure to other verbs that pattern similarly and differently. Finally, such models provide a mechanism by which the child could avoid overgeneralizations such as sentence 7 without negative evidence. Much of the debate about the nature of the child's experience has focused on whether such evidence is available in useful form (43). From our perspective, however, what is more important is the vast amount of information provided by the statistical struc-

ture of the input and the power of constraint-satisfaction networks to distill what is relevant. The network has to find a set of weights that produce correct output for all verbs. The examples that the network is trained on provide positive evidence resulting in changes to the weights that favor attested constructions. Because a common set of weights is used to encode all verbs, these changes simultaneously provide evidence against other constructions to which it has not been exposed. The effect is very much as Chomsky conjectured in his discussion of indirect negative evidence. Direct negative evidence, to the limited extent it is available, is not crucial and merely represents another probabilistic constraint.

A second implication concerns the relevance of poverty-of-the-stimulus arguments to other aspects of language. Verbs and their argument structures are important, but they are language-specific rather than universal properties of languages and so must be learned from experience. Other initial successes of the network approach have been in areas such as vocabulary acquisition, which also involves language-specific knowledge (36, 44). Although the network approach is useful in such domains, it is the properties of universal grammar that are thought to be unlearnable. It is therefore important to observe that standard poverty-of-the-stimulus arguments apply to aspects of language that must be learned, as well as to putative properties of universal grammar (39). If the network approach applies to phenomena such as verb learning, it must be determined if it also applies to other aspects of language that present similar problems.

Conclusions

I have outlined an emerging view of language that is generating considerable interest. This approach rejects the older view's assumption that the starting point for investigations of language is an idealized competence grammar. The goal is instead to explain the nature of language in terms of facts about how language is acquired, used, and represented in the brain. As I have stressed, the approach is new and there are as yet few solid results in hand. Vast areas of language have yet to be addressed at all. Moreover, the claim that humans are born with innate knowledge of grammar does not rest solely on issues concerning acquisition; other phenomena such as universal aspects of language structure, creolization, and dissociations between language and other aspects of cognition are thought to converge on the same conclusion (8). As with the poverty-of-the-stimulus argument, it will be

necessary to reexamine these claims in light of the alternative theoretical framework before drawing definitive conclusions. Perhaps the major achievement to this point is the articulation of a framework in which these important phenomena can be considered anew. Clearly, what is innate about language cannot be determined without exploring the role of experience to its limits.

For many linguists, these developments represent an unwelcome regression to an empiricist view that modern linguistic theory was thought to have definitively refuted years ago (21). This approach does not deny that children are born with capacities that make language learning possible; rather, it questions whether these capacities include knowledge of linguistic universals per se. When the generative paradigm was created, the concept of innate grammatical knowledge provided a needed alternative to tabula rasa empiricism. The approach assumed that the nature of the biological endowment relevant to language could be deduced from studies of competence grammar. Research in developmental neurobiology and in cognitive neuroscience has since begun to yield more direct and specific evidence about how brains are structured and develop (45). Innate capacities may take the form of biases or sensitivities toward particular types of information inherent in environmental events such as language, rather than a priori knowledge of grammar itself (46). Brain organization therefore constrains how language is learned, but the principles that govern the acquisition, representation, and use of language are not specific to this type of knowledge.

REFERENCES AND NOTES

1. N. Chomsky, *Aspects of the Theory of Syntax* (MIT Press, Cambridge, MA, 1965); *Knowledge of Language* (Praeger, New York, 1986).
2. _____, *The Minimalist Program* (MIT Press, Cambridge, MA, 1995).
3. A. Prince and P. Smolensky, *Optimality Theory* (MIT Press, Cambridge, MA, 1997).
4. N. Chomsky, *Knowledge of Language* (Praeger, New York, 1986). To these questions we would now add, how is language represented in the brain and how did it evolve in the species?
5. D. Lightfoot, *The Language Lottery* (MIT Press, Cambridge, MA, 1982).
6. N. Chomsky, *Syntactic Structures* (Mouton, The Hague, 1957), p. 17.
7. E. M. Gold, *Inf. Control* **10**, 447 (1967).
8. S. Crain, *Behav. Brain Sci.* **14**, 597 (1991).
9. S. Pinker, *The Language Instinct* (Morrow, New York, 1994).
10. T. Roeper and E. Williams, Eds., *Parameter Setting* (Reidel, Dordrecht, Netherlands, 1987).
11. D. E. Rumelhart and J. L. McClelland, Eds., *Parallel Distributed Processing* (MIT Press, Cambridge, MA, 1986), vol. 1; C. M. Bishop, *Neural Networks for Pattern Recognition* (Clarendon Press, Oxford, 1995).
12. D. Zipser and R. A. Andersen, *Nature* **331**, 679 (1988); H. Barlow, in *The Cognitive Neurosciences,* M. Gazzaniga, Ed. (MIT Press, Cambridge, MA, 1995), pp. 1011–1014.
13. J. L. Elman, *Cognit. Sci.* **14**, 179 (1990); A. Cleeremans. D. Servan-Schreiber, J. L. McClelland, *Neural Comput.* **1**, 372 (1989). In this pioneering work, recurrent networks were trained on simplified tasks such as predicting the next word in a sentence. Although training does not include feedback about grammaticality, the information the networks encode can be used to support the secondary task of making grammaticality judgments. A network that picks up on the structure inherent in a set of patterns will respond differently to patterns that violate this structure. Grammaticality judgment is then a signal detection problem in which the perceiver must establish criteria that reliably differentiate signals produced by well-formed versus ill-formed utterances.
14. D. E. Rumelhart and J. L. McClelland, in *Parallel Distributed Processing,* J. L. McClelland and D. E. Rumelhart, Eds. (MIT Press, Cambridge, MA, 1986), vol. 2, pp. 216–271.
15. G. Hinton, *Artif. Intell.* **40**, 185 (1989); J. Hopfield, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 8429 (1987).
16. N. Chomsky, *Language* **35**, 26 (1959).
17. D. E. Rumelhart and P. M. Todd, in *Attention and Performance XIV,* D. E. Meyer and S. Kornblum, Eds. (MIT Press, Cambridge, MA, 1993), pp. 3–30.
18. The nature of representations underlying the surface forms of sentences has been a focus of linguistic theory for many years. The internal representations developed by connectionist networks can be seen as a type of underlying structure, although not identical to those proposed within linguistic theory [M. S. Seidenberg, *Cognition* **50**, 385 (1994)].
19. G. Bower and E. R. Hilgard, *Theories of Learning* (Prentice-Hall, Englewood Cliffs, NJ, 1981).
20. J. Berko, *Word* **14**, 150 (1958).
21. S. Pinker, *Science* **253**, 530 (1991).
22. M. S. Seidenberg and J. L. McClelland, *Psychol. Rev.* **96**, 523 (1989); D. Plaut, J. McClelland, M. Seidenberg, K. Patterson, *ibid.* **103**, 56 (1995). Such networks use the same weights to encode "rule-governed" forms and exceptions and therefore do not draw a categorical distinction between the two.
23. J. L. McClelland, in *Attention and Performance XII,* M. Coltheart, Ed. (Erlbaum, Hillsdale, NJ, 1987), pp. 3–36; A. Mackworth, in *Encyclopedia of Artificial Intelligence* (Wiley, New York, ed. 2, 1992), pp. 285–293.
24. Interestingly, in a footnote to the discussion of "colorless green ideas," Chomsky (6) noted, "Given the grammar of a language, one can study the use of the language statistically in various ways; and the development of probabilistic models for the use of language (as distinct from the syntactic structure of language) can be quite rewarding" (p. 17). There is a similar footnote in his manuscript *The Logical Structure of Linguistic Theory* (ms., 1955; Plenum, New York, 1975).
25. The Linguistic Data Consortium, at http://www.ldc.upenn.edu; and the CHILDES Project, at http://psyscope.psy.cmu.edu/childes.
26. F. Jellinek, in *Readings in Speech Recognition,* A. Weibel and K. F. Lee, Eds. (Morgan Kaufmann, New York, 1990), pp. 91–98; R. P. Lippmann, *Neural Comput.* **1**, 1 (1989).
27. J. C. Trueswell and M. K. Tanenhaus, *Lang. Cognit. Processes* **6**, 339 (1991); M. C. MacDonald, *ibid.* **9**, 157 (1994); L. Frazier and K. Rayner, *J. Mem. Lang.* **26**, 505 (1987).
28. M. C. MacDonald, N. J. Pearlmutter, M. S. Seidenberg, *Psychol. Rev.* **101**, 676 (1994); M. K. Tanenhaus and J. C. Trueswell, in *Speech, Language and Communication,* J. L. Miller and P. Eimas, Eds. (Academic Press, San Diego, CA, 1995), pp. 217–262.
29. These facts about language derive from both linguistic and nonlinguistic sources. For example, a particular construction may be high in frequency because it is relatively simple, because of the kinds of events it describes, and because these events happen to be common ones.
30. Similar principles have also been applied to language production [G. Dell, C. Juliano, A. Govindjee, *Cognit. Sci.* **17**, 149 (1993)].
31. P. Jusczyk, *The Discovery of Spoken Language* (MIT Press, Cambridge, MA, 1997); J. L. Morgan and K. Demuth, Eds., *Signal to Syntax* (Erlbaum, Hillsdale, NJ, 1996).
32. J. R. Saffran, R. N. Aslin, E. L. Newport, *Science* **274**, 1926 (1996).
33. C. Moon, R. Panneton-Cooper, W. Fifer, *Infant Behav. Dev.* **16**, 495 (1993).
34. Y. Munakata, J. L. McClelland, M. J. Johnson, R. S. Siegler, *Psychol. Rev.,* in press.
35. L. R. Gleitman, H. Gleitman, B. Landau, E. Wannter, in *The Cambridge Linguistic Survey,* E. Newmayer, Ed. (Cambridge Univ. Press, Cambridge, 1988), vol. 3; S. Pinker, *Language Learnability and Language Development* (MIT Press, Cambridge, MA, 1984).
36. M. Christiansen, J. Allen, M. S. Seidenberg, *Lang. Cognit. Processes,* in press; M. Kelly and S. Martin, *Lingua* **92**, 105 (1994); T. Mintz, E. Newport, T. Bever, in *Proceedings of the 25th Annual Meeting of the Northeastern Linguistic Society,* Amherst, MA, 1995, J. Beckman, Ed. (Graduate Linguistic Students Association, University of Massachusetts, Amherst, MA, 1995).
37. L. Gleitman, *Lang. Acquisit.* **1**, 1 (1990).
38. S. Pinker, *Learnability and Cognition* (MIT Press, Cambridge, MA, 1989).
39. C. L. Baker, *Ling. Inq.* **10**, 533 (1979).
40. B. Levin, *English Verb Classes and Alternations* (Univ. of Chicago Press, Chicago, IL, 1993).
41. N. Chomsky, *Lectures on Government and Binding* (Foris, Dordrecht, Netherlands, 1981), p. 7.
42. J. Allen, thesis, University of Southern California, Los Angeles (1997). My discussion of these phenomena draws heavily on this work, which I gratefully acknowledge.
43. J. Bohannon and L. Stanowicz, *Dev. Psychol.* **24**, 684 (1988); P. Gordon, *ibid.* **26**, 215 (1990).
44. R. Aslin, J. Woodward, N. LaMendola, T. Bever, in *Signal to Syntax,* J. L. Morgan and K. Demuth, Eds. (Erlbaum, Hillsdale, NJ, 1996), pp. 117–134.
45. P. Rakic, in *The Cognitive Neurosciences,* M. Gazzaniga, Ed. (MIT Press, Cambridge, MA, 1995), pp. 127–146; L. C. Katz and C. J. Shatz, *Science* **274**, 1133 (1996).
46. J. Elman *et al., Rethinking Innateness* (MIT Press, Cambridge, MA, 1996); E. L. Newport, *Cognit. Sci.* **14**, 11 (1990).
47. I thank J. Allen and M. MacDonald for discussions that have contributed in essential ways to my understanding of these issues. I also thank J. Saffran and M. Christiansen for helpful discussions. Research supported by National Institute of Mental Health grants PO1 47566 and KO2 01188.