Chapter Ten

Language in, on, and about Time

Elizabeth Bates, Jeffrey Elman, and Ping Li

Language is a system that lives in time—perhaps more so than any other aspect of mind and behavior. Literate individuals can write down their thoughts and leave them lying about for a while in some atemporal paper space. But these written forms are frozen and lifeless until someone picks them up and turns them back into a temporal stream. Spoken language can be used to talk about the past, and it lives in a rapidly changing present. But it always faces forward, rushing toward a future that is (more often than not) charted only a few words or syllables in advance. These forward-facing, temporally constrained properties have some important implications for the way language is acquired by children, implications that we will review in this chapter from three points of view: language *in* time, language *on* time, and language *about* time.

In the first section (*in* time), we deal with the central role of prediction (i.e., guessing what comes next) in language learning. In particular, we will review some recent simulations of language learning in recurrent nets, a kind of neural network architecture that also lives in time, using information about the present and the past to guess what word (or other linguistic element) is going to come next. Results suggest that a future-

Partial support for the writing of this chapter was provided by the Human Frontier Science Program Organization.

oriented system of this kind has enormous potential for capturing realistic aspects of language learning in children.

In the second section (*on* time), we view language as a complex skill that must be carried out under serious temporal constraints. In particular, we compare and contrast the effect of temporal constraints on comprehension (i.e., strategies for optimizing information before it fades away) and on production (strategies for the rapid access of words and sentence frames, and the problems of choice and commitment in sentence planning). This framework helps us to understand aspects of language development that go beyond the "acquisition" of words and grammatical rules, with an emphasis on developmental changes in the efficiency of real-time language use.

In the third and final section (*about* time), we look at how children learn to talk about time. This involves a brief review of cross-linguistic evidence on the development of time words (e.g., "tomorrow," "now"), verb tense (e.g., past, present, future), and verb aspect (e.g., complete vs. incomplete events; continuous, punctate and/or iterative events). Emphasis will be placed on cognitive and conceptual prerequisites for the development of "time talk."

LANGUAGE IN TIME

Time is clearly important to human behavior, if only because many activities are necessarily expressed as temporal sequences. We understand causality because causes precede effects; we learn that the coherent motion over time of points on the retinal array is a good indicator of objecthood; and it is difficult even to think about phenomena such as language, goal-directed behavior, or planning without some way of representing time. So it is particularly curious that most theories of human behavior completely neglect time.

To be sure, there are notable exceptions. Some theorists have seen that the processing of serially ordered events and the abstract representation of those events pose serious challenges. For example, an important issue in models of motor activity is whether the action plan is a literal specification of the output sequence or represents serial order in a more abstract manner (e.g., Fowler, 1977, 1980; Jordan & Rosenbaum, 1988; Kelso, Saltzman, & Tuller, 1986; Lashley, 1951; MacNeilage, 1970; Saltzman & Kelso, 1987). Linguistic theoreticians, on the other hand, have tended to be less concerned with the representation of time (typically assuming, for instance, that all the information in a sentence is somehow made available simultaneously in a syntactic tree); but the research in natural language processing suggests that the problem is not trivially solved (e.g., Frazier & Fodor, 1978; Marcus, 1980). Thus one of the most elementary facts about human activity—that it has temporal extent—is sometimes ignored and is often problematic.

Connectionist Models

In the past decade there has been renewed interest in computational models of behavior that have "brainlike" properties. The guiding principle underlying this approach (sometimes called "connectionism," or "parallel distributed processing," or "neural network models") is that many important aspects of human behavior may be explained by understanding the constraints and properties of the hardware (neural tissue) that supports it. Although it is undoubtedly true that we know far less about brain function than remains to be discovered, there are nonetheless a number of ways the brain is clearly different from the standard digital computer. On one hand, memory access time is dramatically slower in the brain (tens of milliseconds, compared with nanoseconds for the computer); on the other hand, digital computers execute instructions serially, one at a time, whereas processing in the brain is distributed over millions of units that may be active simultaneously. Programs in the computer consist of lists of instructions; "programs" in the brain are probably stored as modifications in synaptic strengths.

Connectionist models attempt to capture these basic properties by using a large number of simple processing units (each with roughly the computational power of a neuron), in which processing elements may display a graded range of outputs (but with nonlinearities in their output function), with massive interconnections between units. An important feature of this approach is the existence of relatively simple learning algorithms that allow networks of this sort to learn behaviors based on example rather than explicit instructions or rules.

Although one of the early attractions of such systems was their ability to carry out processing in parallel (in a way similar to what is believed true of humans), more recent developments have opened up a new way of dealing with serial processing. A variety of approaches have been studied (e.g., Elman, 1990; Jordan, 1986; Stornetta, Hogg, & Huberman, 1987; Tank and Hopfield, 1987; Waibel, Hanazawa, Hinton, Shikano, & Lang, 1987; Watrous & Shastri, 1987; Williams & Zipser, 1988). The most intriguing are those that treat the processor as a dynamic system (that is, a system in which the current state depends to some degree on its prior state). These systems are deceptively simple in appearance but have yielded surprising results.

One of the key insights that has emerged from simulations involving these models is that an enormous amount of information can be gleaned merely by attending to serially ordered input and attempting to predict what will come next. In some sense, of course, this is hardly surprising.

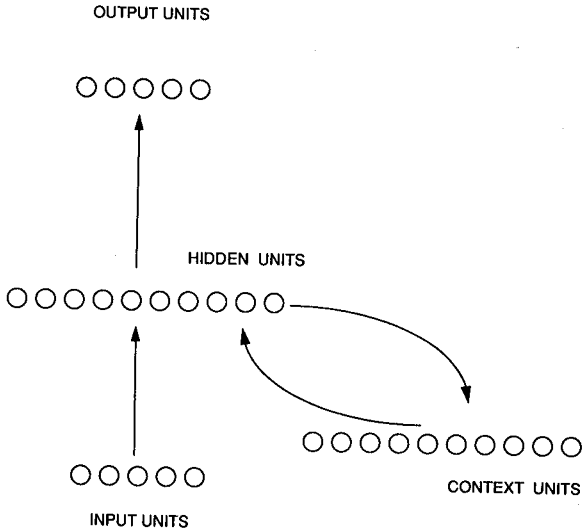

Figure 10.1 Simple recurrent network. Hidden-unit activations are copied into the context units on cycle t, then fed back to the hidden units on cycle $t + 1$.

The order of events in the world (for example, of words in utterances) is rarely either random or lacking structure, but it reflects a rich and complex interaction of constraints. It makes sense that a system that can predict which events will follow other events must be sensitive (if only implicitly) to those constraints. What is less clear is how much prior knowledge must be possessed in order to learn these constraints. The simulations we describe below suggest that rather less prior knowledge is required than one might think.

Learning about Lexical Categories

For instance, consider the problem of lexical semantics: How do children learn the meanings of words? Clearly there are multiple sources of information that allow children to learn word meanings. But what about the primitive semantic and grammatical categories themselves: animate, human, breakable, noun, transitive verb, and so on? Are these innately specified, or can they be induced from the input alone?

Elman (1990) attempted to answer this question by presenting a neural network with a simple problem. The network (shown in Figure 10.1) consisted of three layers. The nodes in the input and output layers were used to represent words (each word was represented by turning on a different node). The intermediate ("hidden") layer was used to allow the network to form intermediate or internal representations of the words it received as input. A context layer was used to recycle hidden-layer activation patterns. This recurrent, or feedback, layer provided the network with internal memory. The network was given a sequence of words as input, one at a time, and asked to produce on its output layer a prediction of which word would follow. The words formed simple sentences and were concatenated so there were approximately 27,000 words in all (a sample fragment is shown below:

input: dog chase woman woman smash plate   boy eat sandwich . . .
output: chase woman woman smash plate boy   eat sandwich girl . . .

The network itself was initially configured with random weights along the connections between layers, so that the predictions at the outset were random (typically all output units would receive some small random activation). As the training proceeded, the network prediction after each new input was compared with the next word that was actually received and the discrepancy between predicted and actual word was used to adjust the weights in the network (using the back-propagation of error-learning algorithm; Rumelhart, Hinton, & Williams, 1986). This procedure allows the network weights to gradually converge on a set of values such that the correct output can be produced in response to any reasonable input.

In fact, at the conclusion of training, the network's performance was quite good. Interestingly, the network was not able to predict the exact word that would follow in any given context. Short of memorizing the sentences, this would not have been possible. A sentence beginning "The woman . . ." might continue in a variety of ways. What the network learned instead was to predict all of the words that might occur in a given context, in proportion to their likelihood of occurrence.

At this point it is worth asking how the network solved the problem. Recall that each word was represented by a vector of 30 ones and zeros in which a single bit—arbitrarily chosen—was turned on. As a result, there is no built-in information about a word's grammatical category or about its meaning. (More precisely, the words are represented by basis vectors, which are orthogonal to each other and lie at the corners of a 30-dimensional hypercube). Indeed, the very notions of grammatical category and semantic feature are not part of the representations made available to the network.

However, grammatical categories and semantic features clearly play a role in the order of words in these sentences. Although it is true that many words might occur in any given context, not just any word was acceptable. Only words that "fit" by virtue of the grammar or semantics of that sentence could be predicted. So we might expect that to solve the prediction problem the network might have to learn something about grammatical and semantic categories. If so, where would this information be situated?

The most likely location would be in the hidden layer. As the network processes a sentence, each new word activates the nodes in the hidden layer. The hidden-layer activations then cause the output nodes to be activated (in order to predict the next word). Although the encoding for words is set in advance, the activation patterns in the hidden units are learned over time by the network, and the network is free to develop internal representations of words that reflect their properties in a way that allows the network to do the prediction task. Thus it is instructive to look at these internal representations.

This can be done by allowing the trained network to process additional sentences. As each word is presented, we save the hidden unit activation pattern it evokes (in the process of producing a prediction). Metaphorically, we might imagine that we are recording event-related potentials (ERPs) from the network as it processes the stimuli, much as we might collect scalp potentials from human subjects as they listen to sentences. Eventually we end up with hidden unit activation patterns corresponding to each of the possible input words. These patterns are vectors in 150-dimensional space (because there are 150 hidden units). It is not feasible to visualize such high-dimensional patterns directly, but we can see how they are distributed in space by looking at the similarity relations between them. A hierarchical clustering tree represents this information by joining vectors that are close in space low in the tree; vectors or groups of vectors that are more distant are joined higher up.

In Figure 10.2 we see a hierarchical cluster diagram of a trained network's hidden unit activations for the words used in this simulation. Remarkably, it appears that the network has discovered a great deal about both the grammatical and the semantic characteristics of the words. The hidden unit activation space is divided into two major regions: verbs (shown on the top of the tree) lie within one region, and nouns (shown on the bottom) within the other. Transitive, intransitive, and optionally transitive verbs are grouped separately within the verb space. And nouns are partitioned into regions corresponding to animate and inanimate, with further divisions marking human, nonhuman, breakable, large versus small animal, and so forth.

These groupings are not part of the input in any overt sense, but they do reflect the temporal structure of the sentences. The noun/verb distinction underlies the basic word order; the argument structure of different verbs and the semantic features of different nouns account for other ways word order is constrained. So while the task of the network was not to extract these distinctions explicitly, it is clear that the

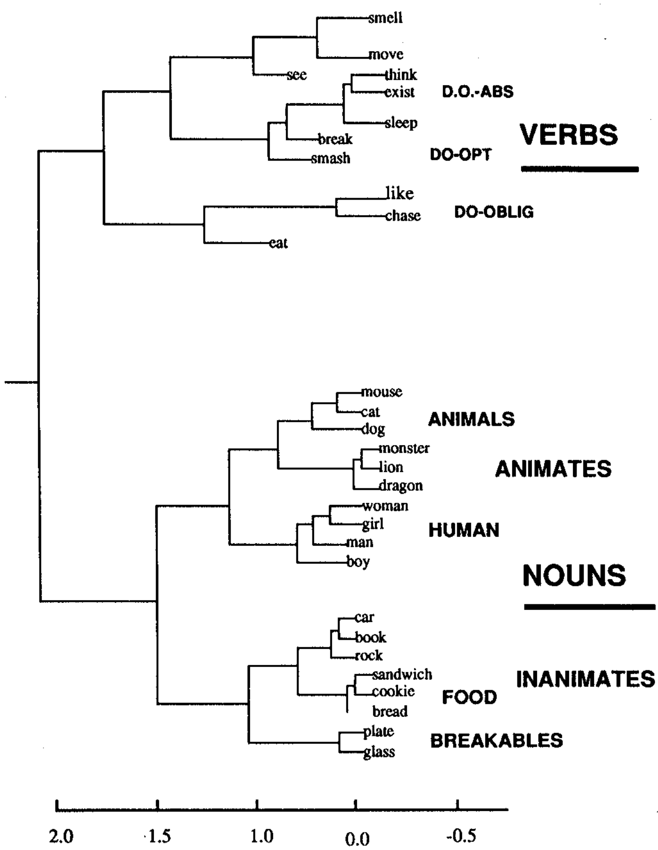

Figure 10.2 Hierarchical cluster of hidden unit activation patterns from network described in simulation 1. The input patterns are orthogonal vectors, but the hidden unit activations encode a similarity structure that reflects the semantic and grammatical characteristics of the lexicon.

(apparently simple) task of prediction provides a powerful motivation for inducing much more structure than we might have supposed at the outset.

Rules as Trajectories through State Space

In the last simulation we saw that the network was able to make useful inferences about the lexical category structure of words, based on distributional facts. The representation of the lexicon in this system looks very different from the standard lexicon we are familiar with in psycholinguistics. Rather than being a table of dictionarylike entries, the lexicon here is a region in state space, embodied in the activation patterns on the

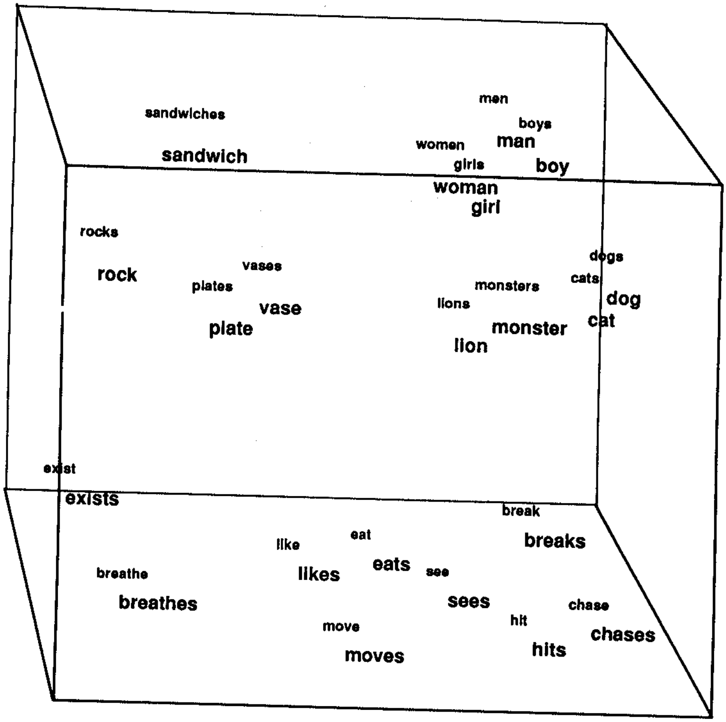

Figure 10.3 Schematic representation of hidden-unit activation patterns as vectors in an *N*-dimensional state space. Lexical items are points in space; different regions correspond to grammatical categories or semantic features.

hidden unit layer. The network learns to organize this space into regions that represent grammatical and semantic information. Thus the lexicon is simply the region of space into which the network moves when it "hears" a word. Figure 10.3 depicts how this might look in three dimensions. What is missing from this picture is a sense of how grammar itself is represented in such a scheme. That is, the lexicon is an important part of language, but we also need a way to represent the constraints on how words may be combined to form interpretable sentences. In traditional schemes, of course, this is accomplished through rules. We might wonder how such constraints could be represented in a system with dynamical properties, such as a recurrent network.

To explore this question, let us consider a somewhat more complex language than the one used in the first simulation, where sentences consisted of only two or three words. Obviously, natural language typically involves utterances that are not only longer, but more complex. For example, consider a sentence such as "The cat the dogs chase climbs up the fence." In some sense, there are really two sentences here: "The cat climbs up the fence," and "The dogs chase the cat." The sentences are combined here using relative-clause formation (as opposed to simply conjoining them with "and"). As a result, there is a part-whole relation between the subordinated sentence ("The dogs chase the cat") and the noun it modifies ("the cat"). There are not only conceptual consequences to this relation, but grammatical consequences as well (e.g., the first noun, "cat," is in the singular and agrees in number with the last verb, "climbs," whereas the second noun, "dogs," is in the plural and agrees with the first verb). An important question, thus, is how a network such as the one used in the first simulation might represent the structure of such complex sentences. The problem is that the significant structure of the sentence departs from the simple linear order of the words.

This problem was addressed in Elman (1991). A simple recurrent network, similar to the one shown in Figure 10.1, was trained on a prediction task involving complex sentences. The following sample illustrates the sort of sentences used:

boys who chase dogs see girls.
girl who boys who feed cats walk.
cats chase dogs.
mary feeds john.
dogs see boys who cats who mary feeds chase.

The prediction task was complicated by the many long-distance dependencies (e.g., between main clause subject nouns and their verbs, often separated by several embedded clauses; or verbs whose direct objects preceded, rather than followed, the verbs). The network was able to learn the grammar only when it began the task with limited resources and a narrow attentional window and then "matured" slowly over time (Elman, 1993). At the conclusion of this training regime, the network not only mastered the training data but was able to generalize its performance to novel sentences that contained new lexical items and structures it had not seen before. At least from its performance, it appears that the network learned the complex grammar underlying the data. How was the grammar encoded?

One can answer this question in much the same way as for the first simulation, by looking at the hidden unit patterns as the network processes the sentences. This time, however, it is instructive to study these patterns as they change over time (and so our network ERPs are really more analogous to the time-varying traces recorded with human subjects than the static traces recorded in the first simulation). Figure 10.4 shows the trajectory of the network's "mental state" as it processes two sentences: "boy who boys chase chases boy" (bottom trace) and "boys who

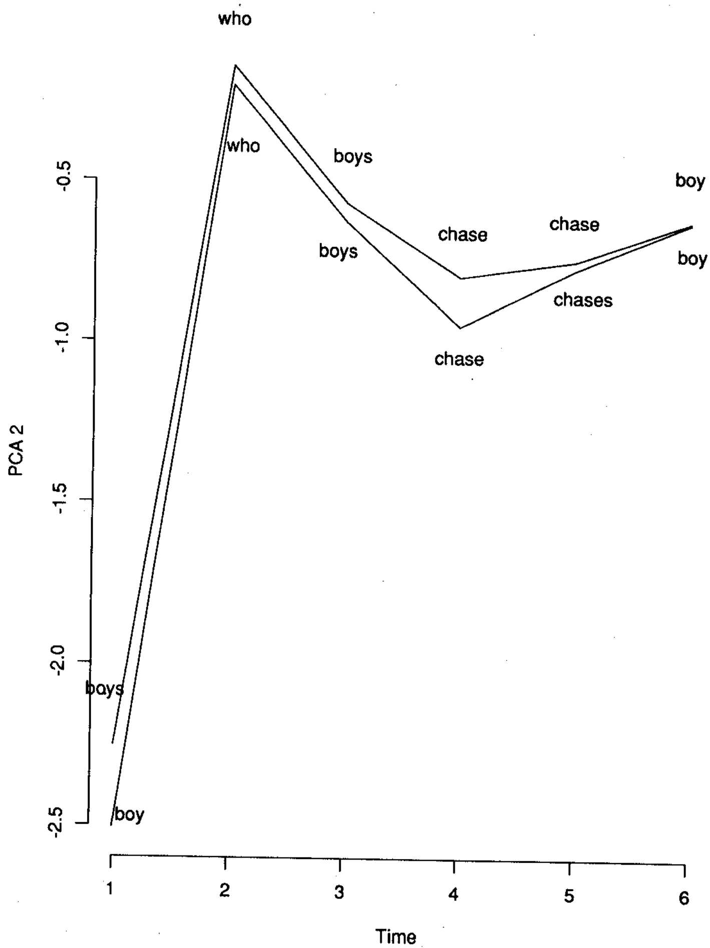

Figure 10.4 Trajectory through (hidden unit) state space as a network processes the sentences "boy who boys chase chases boy" and "boys who boys chase chase boy." The horizontal axis represents time; the vertical axis represents position in state space. The difference in the number of the initial noun (singular vs. plural) is encoded by a displacement in state space.

boys chase chase boy" (top trace). The two sentences are identical except for the embedded relative clauses, which have a singular subject in one case and a plural subject in the other. What Figure 10.4 shows is how the network marks the number of the main clause subject. This is encoded as an upward displacement (for plurals) in the plane of state space.

Figure 10.5 shows state space trajectories for the three sentences "boy chases boy," "boy sees boy," and "boy walks." The verbs in these

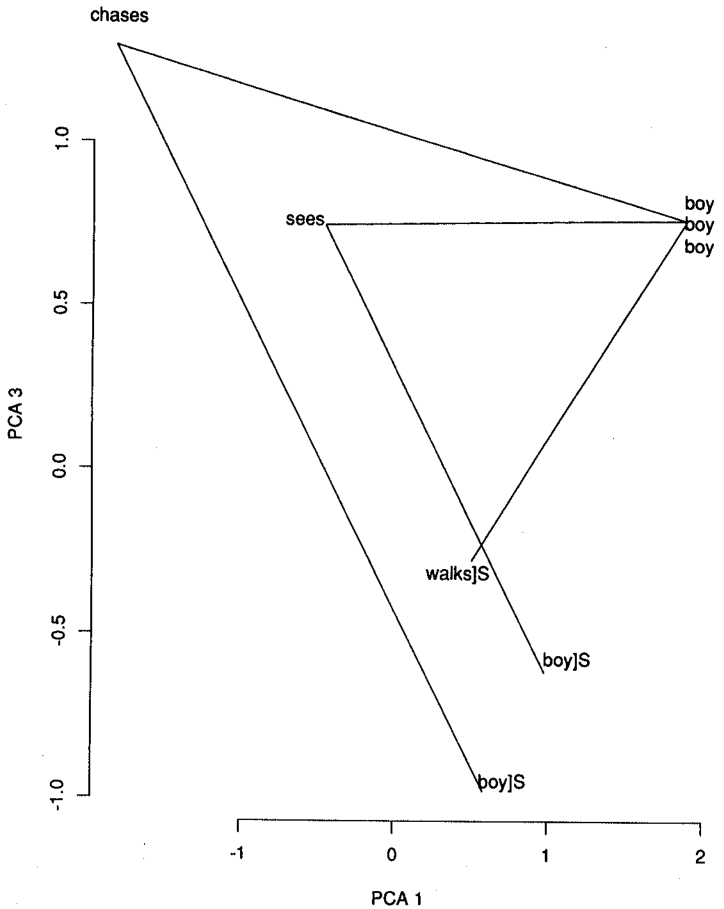

Figure 10.5 Trajectory through state space as a network processes the sentences "boy chases boy," "boy sees boy," and "boy walks." The horizontal and vertical axes represent different dimensions in the state space (see Elman, 1991, for details). The three verbs differ in their verb argument structure: in this grammar, "chases" is obligatorily transitive, "sees" is optionally transitive, and "walks" is obligatorily intransitive. The network represents verb argument expectations along an axis running from the upper left to lower right in state space.

sentences differ with regard to their verb argument structure. In this grammar, "chases" is obligatorily transitive and so always requires a direct object; "sees" is optionally transitive (in this example it is followed by a direct object, but it need not be); and "walks" is intransitive. The network learns to encode this difference in argument structure along an

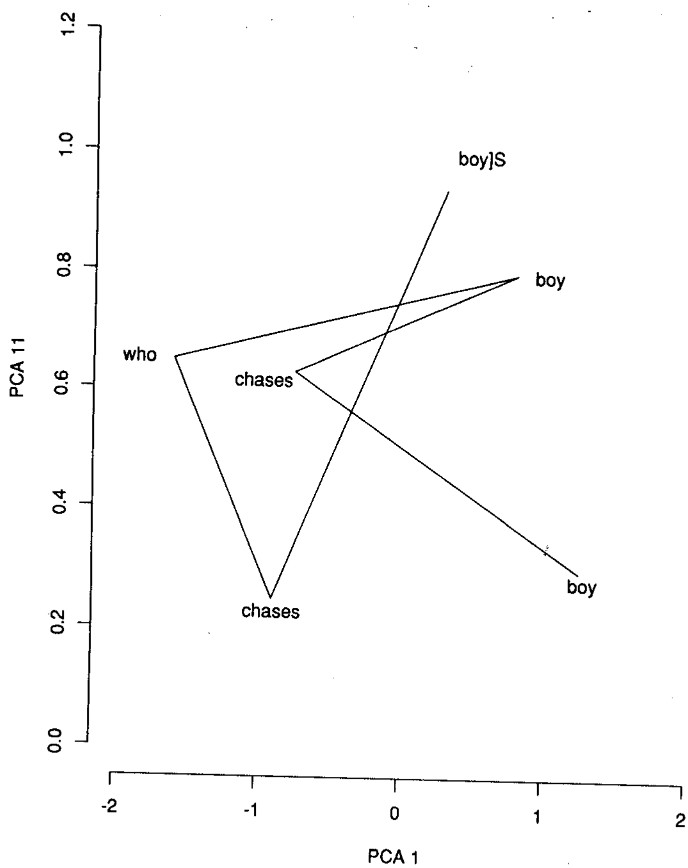

Figure 10.6 Trajectory through state space as a network processes the complex sentence "boy chases boy who chases boy." The horizontal and vertical axes represent different dimensions in the state space (see Elman, 1991, for details). The canonical pattern for a simple sentence is represented in embedded causes as a displacement (to the lower left) in state space.

axis that runs from upper left (for transitive verbs) to lower right regions of state space (for intransitive verbs).

Finally, in Figure 10.6 we see how the network represents levels of embedding. The first part of the trajectory (corresponding to the main clause, "boy chases boy") turns out to be the canonical pattern for all simple transitive sentences (the pattern looks different from that in Figure 10.5 simply because we have changed our viewing perspective and are looking at a different plane in the high-dimensional state space of the network's "brain"). In this figure we see that as the network processes additional relative clauses it replicates the basic canonical pattern, displacing it to the left and downward in state space. This spiraling pattern is typical for embedded clauses and is the way the network represents the hierarchical organization of complex sentences.

Let us now try to summarize what we have learned from these simulations. First, it is clear that there is a great deal of information to be gleaned from the linguistic input. The network starts off with minimal a priori information but is nonetheless able to discover the lexical category structure and grammatical structures implicit in the input. We do not imagine that children are tabulae rasae in the extreme sense that these networks are; yet these results suggest caution about exactly how much prior innate knowledge must be attributed to account for language acquisition.

Second, although the prediction task used in these simulations is elementary and children obviously do more than simply predict, it is also apparent that anticipating how the world will change over time is a powerful driving force for induction and leads to results that are not obviously related to the task itself. We find it plausible to believe that, particularly at early stages of development, children (and other organisms) might use prediction as a way to impose basic order on the world and to bootstrap further learning.

Finally, these networks give us new ways of thinking about the lexicon and about rules. The view that words are simply sensory stimuli that (through convention) move our brains into particular regions of mental space is appealing and is more consistent with the way we might view other sensory inputs. Sentences are not phrase structure trees, but rather trajectories through this mental space. The trajectories are constrained, of course. The dynamics of the system permit certain sequences to be processed; these are grammatically well-formed sentences. Ungrammatical sentences are sequences of words that move against the dynamics. Thus, rules are simply the sets of attractor and repeller regions contained within the system.

In short, we obtain a very different view of linguistic representations and language learning if both are viewed within the framework on temporal dynamics. Now we can turn to another approach to language and time: learning to use language under the severe temporal constraints imposed by the limits of information processing.

LANGUAGE ON TIME

There is a widespread tendency to think of linguistic knowledge as a relatively static set of meanings and rules, like the dictionaries and grammar books that gather dust on library shelves. Indeed, it is common to talk about the dynamic process of language learning as "language acquisition," a metaphor that brings to mind visions of consumers standing in checkout lines to pay for groceries, or squirrels storing nuts for the winter. We would like to argue instead that the state of "knowing a language" is better viewed as an extremely complex skill, one that must be executed rapidly and efficiently, under a number of very serious constraints on the way information can be received, stored, maintained, retrieved, transformed, and output in real time. From this point of view language development involves "learning how" as well as "learning that." If we take this straightforward and yet uncommon perspective on the nature of language learning, then a number of otherwise mysterious facts fall into place.

Under the developmental account that is implied by the term "language acquisition," words or rules have to make an abrupt transition from *outside* the child to *inside*. If this is correct, then we should also observe a large number of abrupt and discontinuous changes in behavior, reflecting the presence of linguistic knowledge that was absent just a short time before. In fact, such abrupt behavioral changes are rare (although they are occasionally observed). New words may appear quite suddenly (Carey, 1982), but their meaning and range of use are shaped gradually across months or years. And as Roger Brown (1973) has observed, grammatical regularities tend to come gradually, "like ivy growing between the bricks" (for extended discussions of this point, with corroborating evidence, see Maratsos, 1983; Marchman & Bates, 1994; Marcus et al., in press; Pizzuto & Caselli, 1992).

The usual move that linguists make to deal with such discrepancies between theory and data is the distinction between competence (linguistic knowledge) and performance (language use). On this argument, linguistic knowledge is indeed a discrete and discontinuous matter, but our view of it is often clouded by psychological factors that are independent of language proper (e.g., fading memory, changes of mind, momentary lapses of attention). This move does serve its primary purpose: It salvages abstract theories of linguistic knowledge from the vicissitudes of imperfect behavioral data. But it leaves the developmental psycholinguist with a number of difficult problems. In particular, to determine when a rule is "really there," the researcher has to develop complex criteria to separate the behavioral wheat from the chaff.

Within the field of child language, these criteria are usually treated together under the rubric of "productivity": a rule is productive when its usage can be demonstrated over some specified number or range of situations or contexts. Productivity criteria are supposed to help us eliminate situations in which the child imitates an application of a rule without any understanding, including repeated uses in well-practiced but poorly analyzed routines like "Whatchadoin?" (for "What are you doing") or

"Don't!" (uttered by a child who does not understand or use the separate elements "Do" and "not"). However, the distinction between rote usage and "true" application of a rule is not easy to make. This is particularly true for morphological markings that appear gradually—"Now you see it, now you don't"—moving from zero to 100% application across a period that may last as long as ten years. Typically, researchers compromise by establishing an arbitrary border, like Brown's criterion of "90% use in obligatory contexts" or Bloom's criterion of "use with at least five different word types in one session" (Bloom, 1970; Brown, 1973; for a comparison of these and several other criteria for productivity, see Bates, Bretherton, & Snyder, 1988). The problem with these criteria is that they are so arbitrary, assigning a sharp border to a gradual phenomenon in order to preserve a theory that does not fit the data very well. Other researchers are willing to give credit for productivity only when the child uses the rule in a novel and creative manner, with nonsense words provided by the experimenter (e.g. "Here is a wug . . . here is another wug. . . . Now I have two wugs"), or in creative errors that could not possibly have occurred in the child's linguistic input (e.g., "The cereal spilled on the floor, so I broomed it"). The problem here is that such errors or novel uses are rare in observations of spontaneous speech, so that investigators who use such criteria are likely to underestimate the child's control over the grammar.

The contrast between linguistic knowledge and linguistic use emerged quite clearly in a recent study on the "acquisition" of the English passive. Passive sentences occur roughly 1% of the time in informal English speech, yet they have been studied extensively in the psycholinguistic literature—perhaps more extensively than they deserve. The popularity of passives in psycholinguistic research is due in part to their complexity, requiring coordination of several aspects of grammar and meaning. For example, the two sentences "John hit the ball" and "The ball was hit by John" differ in word order ("John" in preverbal position vs. "the ball" in preverbal position), grammatical morphology ("hit" vs. "was hit by"), and meaning (the active form focuses on John as topic, while the passive form focuses on the ball as topic). Marchman, Bates, Burkhardt, and Good (1991) conducted a study of production of the passive in more than 170 children and adults ranging in age from 3 to 21 years. Subjects were shown a series of short black-and-white cartoons in which one character acted on another (e.g., a horse biting a goat). At the end of each cartoon, the subject was asked to describe what happened from one of two points of view; for example, "Tell me about the horse" versus "Tell me about the goat." The latter situation (focus on the patient or receiver of the action) can be viewed as the ideal context or "ecological niche" for a passive form (e.g., "The goat was bitten by the horse"), although it would of course be technically correct for the speaker to describe the situation in some other way. In this study mastery of the passive form was measured in two ways: (1) Was the speaker able to produce at least one well-formed passive sentence in response to a discourse probe that focused on the receiver of the action? (2) How often did the speaker use the passive form in its ecological niche (i.e., in response to items in which the experimenter focused on the patient)? These two measures gave very different views of the process by which passives are "acquired." All of the subjects aged 9–21 produced at least one well-formed passive; furthermore, production of at least one passive was observed in 67% of the 3-year-old children. From this point of view, we should conclude that passives are acquired (on average) before the age of 3. However, 3-year-old children produced the passive only 11% of the time on patient-focus items, compared with an average of 85% in adults. Between the ages of 3 and 21 years there was a monotonic increase in rate of passive production. From this point of view, when should we conclude that the passive is "acquired"? Based on traditional competence theories, it is acquired before the age of 3. And yet there is a very large and continuous change in the ability to produce passive forms that does not level off until at least 21 years of age. Three-year-olds "have" the requisite knowledge of passive forms. They also "have" the requisite knowledge of what passives are for, because they clearly know that an active form like "The horse bit the goat" is really not a good answer to "Tell me about the goat." Indeed, they find all kinds of direct or indirect ways to obey the discourse probe (e.g., "The goat just stood there and the horse bit him!"). So what is changing from 3 to 21 years of age? Marchman et al. (1991) conclude that there are slow and continuous developmental changes in the *accessibility* of linguistic forms. The requisite knowledge is there very early (in some rudimentary form), but speakers must learn to retrieve that knowledge efficiently and rapidly for use in conversation. Because passives are so rare (occurring less than 1% of the time in the child's input), children have relatively little experience with this complex form. They *have* it, but they cannot *do* it easily in real time.

There are many aspects of linguistic development that display these characteristics. For example, in the domain of phonology we find a phenomenon called "coarticulation," which refers to alterations in the pronunciation of one phoneme in preparation for a phoneme that is coming up shortly. Thus the sound *b* in the word "bee" is pronounced differently from the sound *b* in the word "boot." Skilled speakers make these adjustments rapidly and unconsciously. Furthermore, skilled listeners exploit these coarticulatory cues in order to predict the sounds they are going to hear next (see above). There is relatively little research on the development of coarticulation in children, but there is some reason to believe it is yet another very gradual process, in both comprehension and production, and constitutes one of the reasons little children sound the way they do.

In the domain of grammar, many more examples can be found. Bates and MacWhinney (1989) have reviewed a number of relatively late changes in sentence comprehension, reorganizations in the way preexisting knowledge is used to interpret sentences in real time. For example, Italian children "know" the rules that govern subject-verb agreement in their language by 2½ to 3 years of age (Pizzuto and Caselli, 1992) but they do not "use" that knowledge extensively to decide "who did what to whom" in sentence interpretation before 6–7 years of age (Devescovi, D'Amico, Smith, Mimica, & Bates, 1993). Given a sentence like "The dog are kicking the cows" (which would be quite grammatical in Italian), Italian adults respond by saying that "The cows did it" (i.e., they assign the actor role to the noun that agrees with the verb in number). But Italian children below age 6 are more likely to respond by choosing the dog as the actor (using word order instead of subject-verb agreement to make their decision). Devescovi et al. suggest that use of agreement cues comes in relatively late, presumably because they require the child to hold several elements in memory before a decision can be made (comparing agreement marking on the first noun, the verb, and the second noun). Thus even though very young Italians "know" the principles that govern agreement in their language (using them productively in free speech), they do not "trust" or "like" to use those rules in real time because the costs of crowding up memory are too high.

There are many other ways languages differ in how they are used in real time. From the perspective of speech production, they differ in the point at which certain *commitments* have to be made in real time. Consider, for example, the production of articles, modifiers, and nouns. Imagine an English speaker extemporizing happily about some topic, hoping to hold the floor for precious milliseconds while he figures out the word he wants to use to describe a particular object. This speaker can stall for time by saying something like the following: "The uhmmmm the little . . . knob there on the left . . ." By using the word "the," the English speaker has committed himself to nothing more specific than production of an English noun. By the time he arrives at the word "little," things have gotten a bit more precise—but not much. Consider how different things are for a speaker of Italian: "Il uhmmmm il piccolo . . . la piccola maniglia alla sinistra . . ." This speaker may have stalled for time, but at the expense of making a grammatical mistake or (at best) a false start that has to be retracted before he can go on. The reason is quite simple: The article in Italian is marked for number and gender, which means that production of the article commits the speaker to a particular number and gender. In the example above, the speaker did not know what word he would eventually use to describe "the little knob." He started out with the masculine article "il," compounded his problem by using the masculine adjective "piccolo" (for "little"), and then had to switch to the feminine form for each of these words before he could go on to produce the right target noun. The problem is even worse for a native speaker of German, because the article must be marked not only for number and gender, but for case (i.e., a marking on the article that indicates whether the noun that follows will be nominative, genitive, dative, accusative, etc.). In essence, German speakers have to have most of the sentence planned (grammatical forms as well as word selection) before they can start out.

On the other hand, these "disadvantages" in the timing of word and sentence planning end up as advantages from the listener's point of view. Assuming that the speaker does not make a mistake and does not have to retract the opening lines, German or Italian listeners can start to zero in on the identity of the noun that is coming several precious tenths of a second before English counterparts. Indeed, Kilborn (1987) has shown that German listeners have a slight but measurable advantage in word recognition experiments, owing perhaps to the exploitation of these early cues to noun identity. Like everything else in the evolution of natural languages, there appears to be a trade-off: lose something here, gain something there. All languages represent an imperfect, constantly changing, but reasonably workable set of solutions to the problem of communicating lots of information in a short time, under very serious constraints of perception, memory, and production. Each language represents a somewhat different configuration of skills, maximizing the timing and utility of information during comprehension while holding the costs of production and the likelihood of error down to some acceptable minimum. From this perspective the process of language learning has more in common with learning to walk, to dance, or to type than we might expect if we restricted ourselves entirely to the "acquisition of competence" view that has dominated the developmental literature to date. Language takes place *in time*, and the child has to learn to comprehend and produce it *on time*. This brings us to the final issue in our review: learning how to use language to talk *about time*.

LANGUAGE ABOUT TIME

The expression of time constitutes an indispensable part of language use. We describe situations as being in the past, present, or future and talk about events as ongoing or completed. Languages differ in the resources they offer for expressing time concepts, for example, whether they use grammatical devices such as tense and aspect markers or lexical items such as temporal adverbials. But they can all express basic notions concerning time.

Two of the most important devices for expressing time concepts in the world's languages are *tense* and *aspect*. Tense is typically used to locate the *event time* (the time of the event being talked about) with respect to the *speech time* (the time when the speaker utters the sentence). The event time can also be specified with reference to some time other than the speech time—the *reference time*. When the event time is before the speech time, we use the *past* tense; when speech time is before the event time, we use the *future* tense; and when the two overlap, we use the *present* tense. Note that in all these cases the speech time and the reference time overlap. When they do not, we use tense markings such as the *pluperfect* (event time before reference time and reference time before speech time) or the *future perfect* (speech time before event time and event time before reference time).

In contrast to tense, which is concerned with the relation between situations at different time points, aspect is the way speakers view the internal temporal constituency of a single situation, for example, as beginning, continuation, or completion. Traditionally, aspect is divided into two basic perspectives: *perfective* versus *imperfective* (cf. Comrie, 1976). Perfective aspect presents a situation in its entirety as a single whole (often with emphasis on the completion of the action), while imperfective aspect presents a situation with emphasis on its internal structure (focusing on the event as ongoing). As a grammatical category, aspect is expressed by inflectional morphology (e.g., "is doing" vs. "has done" in English), derivational morphology (e.g., in Russian), and other grammatical devices such as particles (e.g., in Chinese).

Clearly, children grow up speaking language in time, and they necessarily learn to use the language about time. In what follows, we will review two important theories originating from crosslinguistic studies of children's acquisition of time talk and present a general view on the relation between cognitive prerequisites and the development of temporal expressions. A first important theory (or "working hypothesis") about the acquisition of temporal expressions is Richard Weist's four-stage model. Weist (1986, 1989) proposes that children proceed through a sequence of four systems of time talk in the framework of speech time, event time, and reference time as discussed above.

The first system is the *speech time system* (about 1;0 to 1;6),[1] under

1. For consistency, all ages are given in the format "years;months" (e.g., 1;6 is 1 year, 6 months).

which both the event time and the reference time are frozen at the speech time. At this stage children can talk only about here and now, that is, visible objects in the immediate perceptual field. This system is characterized as having no tense, aspect, modality, and so on (cf. Brown, 1973). Weist presents crosslinguistic evidence from studies of a number of languages to support this system, including English, Polish, Hebrew, Japanese, and Mandarin Chinese. For example, Chinese has aspect but not tense markings; however, according to Erbaugh (1982), child Mandarin is not marked even for aspect at this initial phase.

The second system is the *event time system* (about 1;6 to 2;6), under which the event time has been separated from the speech time. This system is characterized by children's ability to talk about event time relative to speech time. At this stage children get out of the here and now constraints of the speech time system, and now can express both an ongoing event at the time of speech and an event at a past time, that is, before the speech time. According to Weist, they can also express their intentions and desires concerning potential subsequent events at this stage. For example, in Mandarin Chinese, children often use the modal verbs *hui* 'can' and *yao* 'want' to express their intentions (cf. Erbaugh, 1982), which prepares for the occurrence of future marking.

The third system is the *restricted reference time system* (about 2;6 to 3;0), under which the event time is now "anchored" within a framework of the reference time. The emergence of the concept of reference time is characterized by the occurrence of temporal adverbs and temporal adverbial clauses in children's utterances. At this stage children begin to use adverbs denoting reference time such as "today," "tomorrow," and "yesterday" and adverbial clauses that indicate the framework within which an event takes place, such as "when Mama was little . . ." In this case children have learned to use temporal adverbs and adverbial clauses to set up the reference time. However, it appears that the use of reference time is still restricted at this stage, because the event children talk about takes place only within the reference time; that is, children can relate only two points in time at this stage.

Only at the last stage, when the *free reference time system* about 3;0 to 4;6) is established, can they set up a reference time before or after the speech time and relate the event time to it at a third point in time. At this stage all three time concepts can be manipulated independently, and children can freely relate events to one another in time. This system is characterized by the emergence of words like *before* and *after* that allow the establishment of temporal sequencing at different time points in children's time talk. Weist also points out that although there is evidence the children can coordinate three temporal locations as early as 3 years of age, most research indicates that it is not until about 4;0 to 4;6 that children can freely express the concepts of speech time, event time, and reference time.

It has long been recognized that there is a close relation between children's expression of modality and their acquisition of the future tense. Weist (1989) points out that though researchers accept that children can express modality relatively early, they are nevertheless reluctant to credit them with the ability to express future. He suggests it is possible that all references to the future have some modal value, that is, some sense of possibility or necessity (intention, desire, obligation, permission, etc.), and therefore the expression of future is also relatively early in child language. As early as 1;7 years of age, Polish children can say things like "Mommy will take it out" as a request, and "will turn (it) over" as a statement. These kinds of requests or statements, whatever their modal value might be, surely express the deictic relation of the event time's being subsequent to the speech time.

The four-stage development of the temporal system is a very interesting phenomenon in children's expression of time. What is more interesting is that these stages correspond nicely to the cognitive and conceptual development of time. Weist suggests that children's linguistic coding of events and situations allows us to infer their conceptual development, and that crosslinguistic evidence is crucial in this respect because different languages encode time concepts in a variety of ways. When the linguistic devices used to encode a time concept are the same across languages (e.g., temporal adverbs), the emergence of the time concept into the evolving temporal systems is stable across those languages. In contrast, when the linguistic devices are different (e.g., tense and aspect markers), the emergence of the time concept will be variable. Weist suggests that each time the child makes a transition from one temporal system to another, a corresponding conceptual development is established. For example, there is evidence that children's conceptual development is no longer restricted to the here and now when they proceed to the event time system, because at this stage they have achieved the capacity for displacement—that is, they now can talk about objects outside the immediate perceptual field. Conceptually, they are now able to retrieve prior experience from memory, knowing that these experiences happened before the present time. When children have developed the reference time system at the last stages of temporal development, they can separate time points and express them either in a serial order or in reversible orders. Initially they may be restricted to expressing things only in the natural order in which things happen, such as "After John had his breakfast, he went to work," and only later can they say "Before John went to work he had his breakfast."

A second important line of theorizing about the development of children's language about time is Slobin's "Basic Child Grammar" (Slobin, 1985). Although this model is more general and not specifically designed for temporal development like Weist's four-stage model, it bears important relevance to the issue of the expression of time. The "Basic Child Grammar" assumes that children come to the language acquisition task with a prestructured "semantic space" containing a universal, uniform set of semantic notions that are at first neutral with respect to language-specific categories. According to Slobin, these semantic notions are "privileged" to be mapped onto grammatical forms of individual languages in the process of language acquisition. That is, before children's experience with specific properties of the grammar, these notions strongly attract grammatical forms of the input language in the form-meaning mapping processes. Slobin proposes two universal "temporal perspectives" in the semantic space: *process* vs. *result*. These temporal perspectives function early to define a semantic contrast in children's learning of tense and aspect systems. Slobin emphasizes that *result* is a particularly salient perspective to children and provides an early mapping point for speech segments associated with content words referring to actions. Specifically, whenever a language has an acoustically salient past tense or perfective marker on the verb, its first use by the child seems to be to comment on an immediately completed event that results in a visible change of state, such as situations denoted by the verbs "drop," "fall," "break," and "spill."

Slobin's hypothesis has been supported by crosslinguistic studies in a number of languages, including English, French, Italian, Turkish, Greek, and Chinese. In English, Bloom, Lifter, and Hafitz (1980) found that the distribution of children's tense and aspect markers was correlated with different semantic categories. In particular, between ages 1;10 and 2;4, "-ing" occurred in children's utterances almost exclusively with action verbs that named durative, noncompletive events, for example, "play", "ride," and "write," whereas "-ed" and irregular past-tense forms occurred overwhelmingly with verbs that named nondurative, completive events, that is, events "with a relatively clear result," such as "find," "fall," and "break."

Bronckart and Sinclair (1973) found that before age 6, the French children in their study used the *passé composé* (past perfective) forms significantly more often than the *présent* (present) for "perfective events" that had a clear end result and, conversely, used the *présent* significantly more often than the *passé composé* for "imperfective events" that did not lead to any observable result. These authors conclude that the distinction between process and result is the predominant and perhaps the only aspectual feature in the language of French children below 6 years old.

In Italian, Antinucci and Miller (1976) showed that their subjects first restricted their use of the past tense (*passato prossimo*) to "change of state" verbs that specified actions with a clear result, such as "fall," "find," and "break." The children did not combine activity and stative verbs with the past tense but rather used the *imperfetto* (imperfect). They also found that children made the inflectional endings of past participles agree with the number and gender of the *object* of the verb. In the adult language the participle agrees only with the subject—with one interesting exception. When there is a clitic (unstressed) object pronoun before the verb, the participle must agree with the object pronoun. Thus, the English sentence "The horse has pushed the cow" should be translated as "Il cavallo ha spinto la mucca," where the only agreement marking (indicated with underlining) occurs between subject, *cavallo* and the verb auxiliary *ha*. However, things change with the addition of an emphatic object pronoun before the verb, as in "The horse (it) has kicked the cow" (where the pronoun "it" refers to the cow). This emphatic form requires an additional agreement between participle and object, as in "Il cavallo la ha spinta la mucca," which is usually reduced further by rules of vowel elision to "Il cavallo l'ha spinta la mucca."

This is a very subtle rule in Italian, one we might not expect children to notice until they are fairly far along in development. However, it looks as though Italian children find this pattern very early and overgeneralize it to all past-participle forms, producing impossible sentences like "Il cavallo ha spinta la mucca." This "invention" or overextension of a syntactic agreement rule suggests that they tend to focus intently on the result of the event.

In Turkish, Aksu (1978) reported similar findings. Turkish children before age 4;6 tend to use the past-tense form *-di* with "change of state" verbs to mark punctual, resultative events, but use the progressive form *-iyor* with activity verbs that do not indicate any result.

In Modern Greek, Stephany (1981) found that perfective aspect occurred more frequently with resultative verbs than with nonresultative verbs in young Greek children's utterances, whereas the reverse was true for imperfective aspect. Furthermore, past tense occurred only with perfective aspect, and present tense only with imperfective aspect.

Finally, in Mandarin Chinese, Li (1990, 1991) conducted three experiments to test the process/result distinction in children's speech between ages 3 and 6. The comprehension experiment revealed that children understood the perfective marker *-le* better with resultative verbs and the imperfective marker *zai* better with process verbs. The production experiment also showed that there was a strong association between perfective aspect and resultative verbs and between imperfective aspect and process verbs in children's productive speech. The imitation experiment indicated that children repeated sentences with the perfective marker *-le* significantly better than those with the imperfective marker *zai* when the verb was resultative. Taken together, the results from cross-linguistic studies of tense and aspect systems are consistent with Slobin's hypothesis about the "temporal perspectives" in the time talk of early child language, with special emphasis on the child's interest in "getting results."

The temporal perspectives in Slobin's "Basic Child Grammar" assign a significant role to the child's cognitive predisposition, since Slobin hypothesizes that children are sensitive to the semantic distinction between *result* and *process* before their experience with specific properties of the target grammar. However, the extent to which the meaning-form associations in children's time talk conform to the distribution patterns in the adult language, together with evidence from studies of early child-mother interactions, suggests another equally plausible explanation without invoking prestructured semantic categories as Slobin does. There is now good evidence that children are influenced at a very early age by the input they are exposed to (see Snow, 1977, for discussion). In the domain of children's expression of time, we also have strong evidence that children's speech reflects adult language patterns and that mother-child interactions at an early stage promote certain temporal perspectives (see Li, 1990, for a summary). Brown (1973) noticed that children's use of past tense begins with a small set of verbs that refer to instantaneous events, such as "fell," "dropped," "crashed," and "broke," and he reasoned that these verbs may have always or almost always occurred in the past tense in mothers' speech at early stages of child language. Stephany (1981) specifically explored the correlation between children's use of tense and aspect markers and mothers' use of these markers. Her analysis indicated that the distribution of tense and aspect markers with different types of verbs in young Greek children's speech conforms surprisingly well to the patterns mothers use in speaking directly to children. In addition, Stephany also compared the mothers' speech to children and mothers' speech to adults and found that mothers modify their speech to children with respect to the use of temporal forms and the frequency of these forms. More important, in an investigation of the acquisition of Brazilian Portuguese, de Lemos (1981) showed how mothers model temporal markers in special contexts designed to direct the child's attention to certain situational properties, in particular, to the result of an event. De Lemos's (1981) detailed analyses of the patterns inherent in child-mother interactions suggest that the child's sensitivity to the contrast between process and result is strongly promoted and encouraged by the mother. Each time the child plays a game and the game leads to some observable results, the mother emphasizes the completed result by using a perfective marker with relevant verbs, and at one stage the child could also imitate the mother's speech. This kind of routine is very characteristic of child-mother interactions between ages 1;6 and 2;2, according to de Lemos. Results seem to be a very salient property of situations, especially in child play (e.g., the child builds a toy tower and knocks it down), and the world's languages place special emphasis on it (see Li, 1990, on the properties of Chinese, and Nedjalkov, 1988, for a discussion of resultative constructions of the world's languages). It is therefore not surprising that mothers intentionally direct children's attention to it at a very early stage in their acquisition of time talk.

Bowerman (1985, 1989) argues that children are sensitive to the characteristics of input patterns from early on and that their speech reflects their efforts in the analysis of the distributions of form-meaning mappings. With crosslinguistic evidence of the described kind, it is reasonable to propose that input patterns and child-mother interactions play an important role in the early stages of children's expression of time, and that young children attend to the temporal perspectives from early on as a result of the child's analysis of the learning environment, not because of prestructured semantic categories.

In this section we discussed some features inherent in children's learning of the language *about time*, that is, the development of time talk. Two theoretical hypotheses about children's expression of time were reviewed. First, Weist (1986, 1989) hypothesizes that in learning to talk about time, children proceed through four stages of the temporal system: the speech time, the event time, the restricted reference time, and the free reference time. The progress from one stage to another reflects corresponding cognitive and conceptual development. Second, Slobin (1985) proposes two temporal perspectives, process and result, to which children are sensitive from a very early stage in their acquisition of temporal expressions. Slobin hypothesizes that these perspectives are prestructured and "privileged" to be mapped to corresponding grammatical forms. However, crosslinguistic evidence suggests that the patterns observed with children's acquisition of time talk can be better accounted for by reference to the properties of the input and to child-mother interactions than by invoking predisposed semantic structures.

CONCLUSION

We have now come full circle. In the first section we stressed the extent to which a simple neural network that lives *in time* can succeed in learning very complex aspects of language, simply by trying to "guess what comes next." In the second section we pointed out that much of grammatical learning can usefully be viewed as a form of skill acquisition, as children learn to produce and understand speech as efficiently as possible, that is, *on time*. In the third section we reviewed current theories of the process by which children learn to talk *about time*. To a great extent, these theories have presupposed that learning to talk about time reflects strong innate constraints (whether they are linguistic or cognitive in nature). However, we end by concluding that these biases can (to some extent) be induced from exposure to a well-structured environment. In other words, a system that lives *in time* can learn to talk *about time*, by attending closely to the way language is used in the world around it. In other words, much of the exquisite organization that we see in language learning can be explained by temporal constraints.

REFERENCES

Aksu, A. (1978). *Aspect and modality in the child's acquisition of the Turkish past tense.* Ph.D. dissertation, University of California, Berkeley.

Antinucci, F., & Miller, R. (1976). How children talk about what happened. *Journal of Child Language,* 3, 167–189.

Bates, E., Bretherton, I., & Snyder, L. (1988). *From first words to grammar: Individual differences and dissociable mechanisms.* New York: Cambridge University Press.

Bates, E., & MacWhinney, B. (1989). Functionalism and the competition model. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing.* New York: Cambridge University Press.

Bloom, L. (1970). *Language development: Form and function in emerging grammars.* Cambridge: MIT Press.

Bloom, L., Lifter, K., & Hafitz, J. (1980). Semantics of verbs and the development of verb inflection in child language. *Language,* 56, 386–412.

Bowerman, M. (1985). What shapes children's grammars? In D. Slobin (Ed.), *The crosslinguistic study of language acquisition: Vol. 2. Theoretical issues.* Hillsdale, NJ: Erlbaum.

Bowerman, M. (1989). Learning a semantic system: What role do cognitive predispositions play? In M. Rice & R. L. Schiefelbusch (Eds.), *The teachability of language.* Baltimore: Brookes.

Bronckart, J., & Sinclair, H. (1973). Time, tense and aspect. *Cognition,* 2, 107–130.

Brown, R. (1973). *A first language.* Cambridge: Harvard University Press.

Carey, S. (1982). Semantic development: The state of the art. In E. Wanner & L. Gleitman (Eds.), *Language acquisition: The state of the art.* Cambridge: Cambridge University Press.

Comrie, B. (1976). *Aspect: An introduction to the study of verbal aspect and related problems.* Cambridge: Cambridge University Press.

de Lemos, C. (1981). Interactional processes in the child's construction of language. In W. Deutsch (Ed.), *The child's construction of language.* New York: Academic Press.

Devescovi, A., D'Amico, S., Smith, S., Mimica, I., & Bates, E. 1993). *The development of sentence comprehension in Italian and Serbo-Croation: Local versus distributed cues.* Manuscript, University of Rome "La Sapienza," submitted for review.

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

Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning, 7,* 195–225.

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

Erbaugh, M. (1982). *Coming to order: Natural selection and the origin of syntax in the Mandarin-speaking child.* Ph.D. dissertation, University of California, Berkeley.

Fowler, C. (1977). *Timing control in speech production.* Bloomington: Indiana University Linguistics Club.

Fowler, C. (1980). Coarticulation and theories of extrinsic timing control. *Journal of Phonetics, 8,* 113–133.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition, 6,* 291–325.

Jordan, M. I. (1986). *Serial order: A parallel distributed processing approach.* Institute for Cognitive Science Report 8604. University of California, San Diego.

Jordan, M. I., & Rosenbaum, D. A. (1988). *Action.* Technical Report 88-26. Department of Computer Science, University of Massachusetts at Amherst.

Kelso, J. A. S., Saltzman, E., & Tuller, B. (1986). The dynamical theory of speech production: Data and theory. *Journal of Phonetics, 14,* 29–60.

Kilborn, K. (1987). *Sentence processing in a second language: Seeking a performance definition of fluency.* Ph.D. dissertation, University of California, San Diego.

Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior.* New York: Wiley.

Li, P. (1990). *Aspect and aktionsart in child Mandarin.* Ph.D. dissertation, Leiden University.

Li, P. (1991). *Zai* and *ba* constructions in child Mandarin. *Center for Research in Language Newsletter* 5(5) (University of California, San Diego).

MacNeilage, P. F. (1970). Motor control of serial ordering of speech. *Psychological Review, 77,* 182–196.

Maratsos, M. (1983). Some current issues in the study of the acquisition of grammar. In J. Flavell & E. Markman (Eds.), *Handbook of child psychology* (Vol. 3). New York: Wiley.

Marchman, V., & Bates, E. (1994). Continuity in lexical and morphological development: A test of the critical-mass hypothesis. *Journal of Child Language, 21:2*, 339–366.

Marchman, V., Bates, E., Burkhardt, A., & Good, A. (1991). Functional constraints on the acquisition of the passive: Toward a model of the competence to perform. *First Language, 11*, 65–92.

Marcus, G. F., Pinker, S., Ullman, M., Hollander, M., Rosen, T. J., & Xu, F. (in press). Overregularization in language acquisition. *Monographs of the Society for Research in Child Development.*

Marcus, M. (1980). *A theory of syntactic recognition for natural language.* Cambridge: MIT Press.

Nedjalkov, V. (Ed.). (1988). *Typology of resultative constructions.* (Translation ed. B. Comrie). Amsterdam: John Benjamins.

Pizzuto, E., & Caselli, M. (1992). Acquisition of Italian morphology: Implications for models of language development. *Journal of Child Language. 19*, 3, 491–557.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning internal representations by error propagation. In D. E. Rumelhart and J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1). Cambridge: MIT Press.

Saltzman, E., & Kelso, J. A. S. (1987). Skilled action: A task dynamic approach. *Psychological Review, 94*, 84–106.

Slobin, D. (1985). Crosslinguistic evidence for the language-making capacity. In D. Slobin (Ed.), *The crosslinguistic study of language acquisition: Vol. 2. Theoretical issues.* Hillsdale, NJ: Erlbaum.

Snow, C. (1977). The development of conversation between mothers and babies. *Journal of Child Language, 4*, 1–22.

Stephany, U. (1981). Verbal grammar in Modern Greek early child language. In P. S. Dale & D. Ingram (Eds.), *Child language: An international perspective.* Baltimore: University Park Press.

Stornetta, W. S., Hogg, T., & Huberman, B. A. (1987). *A dynamical approach to temporal pattern processing.* Paper presented at the IEEE Conference on Neural Information Processing Systems, Denver.

Tank, D. W., & Hopfield, J. J. (1987). *Neural computation by concentrating information in time.* Paper presented at the IEEE International Conference on Neural Networks, San Diego.

Waibel, A., Hanazawa, T., Hinton, G., Shikano, K., & Lang, K. (1987). *Phoneme recognition using time-delay neural networks.* ATR Technical Report TR-I-00006; Japan: ATR Interpreting Telephony Research Laboratories.

Watrous, R. L., & Shastri, L. (1987). *Learning phonetic features using connectionist networks: An experiment in speech recognition.* Paper presented at the IEEE International Conference on Neural Networks, San Diego.

Weist, R. (1986). Tense and aspect: Temporal systems in child language. In P. Fletcher & M. Garman (Eds.), *Language acquisition.* Cambridge: Cambridge University Press.

Weist, R. (1989). Time concepts in language and thought: Filling the Piagetian void from two to five years. In I. Levin & D. Zakay (Eds.), *Time and human cognition: A life-span perspective.* Amsterdam: Elsevier.

Williams, R. J., & Zipser, D. (1988). *A learning algorithm for continually running fully recurrent neural networks.* Institute for Cognitive Science Report 8805. University of California, San Diego.