Chapter 1

The Sentence as a Case Study in Cognitive Science

Much of the work of modern cognitive science has assumed that cognition is at least in part computational. That is, "to think" means "to manipulate symbols in a particular manner." In cognition, the symbols are mental representations that possess meaning or "stand for" mental objects. Regardless of the domain, the formal manner of manipulating symbols constitutes the syntax of the domain. Syntax is not a unique formal property of linguistic computation, but it is part of vision, motor behavior, and every other activity with a computational basis.

This canon of modern cognitive science recently has been under attack. One of the major problems of the syntactic conception of cognition has centered around how the formal operations take on appropriate meanings. That is, how does it come about that the formal manipulations of symbols produce computational effects that have meaning in our world? What we will try to show is that, in the case of language, the sentence provides the minimal domain into which elementary meanings can be placed and combined. Thus, when more elaborate structures are derived from sentences, the more elaborate structures will also have meaning.

A second line of attack on the assumption that cognition involves symbol processing comes from impressive demonstrations that models of intelligent behavior need not encode symbols at all. The nonsymbolic connectionist approach has attained remarkable success in predicting language behavior with networks of simple processing units that are associated with one another in various ways. Importantly, none of these simple processing units needs to represent anything. Nevertheless, connectionist models are able to behave in ways that mimic the behavior of human beings as they carry out such complex tasks as understanding language.

In this book we present a model that addresses both of these attacks on the symbol-processing metaphor for cognitive science. This chapter makes the point that mental processes in general and linguistic processes in particular come in two flavors —habits and computations. We argue that the sentence level is a natural level of linguistic representation, and that, despite many valid arguments for an associative aspect of sentence comprehension, sentence comprehension also is fundamentally

computational. Our integration of computational and associative approaches, furthermore, can resolve difficult problems that each approach faces when independently considered as a complete model of comprehension.

1.1 The Sentence Is a Natural Level of Linguistic Representation

One argument for the sentence level flows from consideration of a widespread phenomenon in cognitive science, inductive learning. Inductive learning is the acquisition of general knowledge based on experience with specific examples. Since it is based on experience, this type of learning should be influenced by how frequently specific patterns occur. We will see in this section that inductive learning of language requires considering the sentence as a natural level of representation, defined by a grammar.

Cognitive science has been founded on two alternate truths:

- Most of the time what we do is what we do most of the time.
- Sometimes we do something new.

Both statements are intuitively correct. Yet each alone has dominated the cognitive sciences for sustained periods of scientific history. For example, during the first half of the twentieth century, the first statement was enshrined within associationistic behaviorism as the only relevant fact. In this view, complex behavior was concocted out of associations between individual mental entities. The behaviorist constraint further restricted the associated entities themselves to being observable in actual stimuli and behaviors.

The basic paradigm was *stimulus-response* (S-R) theory, in which patterns of behavior are built up out of the environmental reinforcement of connections between particular stimulus configurations and response sets. Language was recognized as an extremely complex behavior because it involves stringing words together into long sequences. S-R theory was elaborated to explain this in terms of long S-R chains in which each successive word served as both the reinforcement of the previous word and the stimulus for the next one.

The S-R paradigm for language never got very far for several reasons. First, it was never implemented in a way complex enough to begin to approach the intricacies of actual speech. Second, by the late 1950s it was clear that associative behaviorism was not adequate to explain many different kinds of facts, ranging from animal behavior in nature and the lab to human language. For example, Karl Lashley (1951) noted that spoken language comprehension cannot be explained in terms of simple direct associations, because of the frequent presence of relations at a distance. When the following sentence is spoken

(1) Rapid righting with his uninjured hand saved from loss the contents of the capsized canoe.

the interpretation of the spoken form of *righting* depends on material presented much later in the sentence. Examples like these provided evidence against a simple chaining associative model of language behavior, in which the interpretation of a word is a conditioned response to the previous word.

The final reason was a logical failure of the S-R paradigm as a scientific theory. The problem here is that there is no independent way to define what counts as the relevant "stimulus," the relevant "response," or the relevant "reinforcement" all at the same time (Chomsky 1959). Even simple examples of conditioning, such as training a rat to press a lever to a particular tone, face this logical problem. When the rat appears to learn to press the bar correctly, there is still no evidence that it has conceived of the situation in the same way as the experimenter. Is the stimulus the tone? Is it the tone in a particular cage? Is it the tone a particular amount of time after some other event? Is it something else that the tone affects, such as vibrating the sawdust in a particular way? Of course, each of these possibilities can be studied, and the field of learning was becoming littered with evidence that any or all of the alternatives could be true (see Saltz 1971 for a review).

The problem is even more manifest when considering a complex behavior such as language. When children who have not yet mastered English hear the sentence

(2) The sky is blue today.

how do they know what to relate it to? What conceptual probabilities are reinforced? Which parts of the utterance are reinforced? Even if one restricts the domain of inquiry to sentence-internal pattern learning, the problem remains. *The* can be viewed as a stimulus for *sky*, but what is the stimulus for *is* or *blue*, and most specifically, *today*? Chomsky (1959) argued that a theory requires an independent definition of the natural objects under study, before one can investigate the effects of frequency and reinforcement on learning those objects. In the case of language, he suggested that the natural object is the sentence, and its definition is provided by a grammar. His arguments were generally taken as persuasive, and the S-R attempts to deal with language faded. From our standpoint, the important idea is that even inductive learning models of complex behaviors require structures that can define relevant patterns over which learning can be reinforced. The sentence is a level of organization at which such patterns can be defined.

Associationism is an ever-renewable resource. Connectionist models recently have resuscitated the power of habit-based theories and have rehabilitated the reputation of inductive learning. These models recapture the intuition that most behavior is made up of accumulated habits, themselves based on frequency. While they are descendents of associationist behaviorism, many connectionist models have broken with the behaviorist stipulation that only observable entities can be associated.

Connectionist models consist of simple processing units analogous to neurons. These processing units are interconnected in various ways, and the activity of any

particular unit depends on the input it receives from other units. Applied to language, units can be triggered by more than one word, by "memory" of prior words, or by internal units that have no direct correspondence to overt stimuli. As the system gains experience with language, the weight, or value, that is assigned to any particular input and internal connections may change depending on feedback about whether it has responded correctly. That is, a connectionist system can encode environmentally appropriate modifications to its behavior, meeting at least a rough definition of *learning*.

Connectionist models explicate comprehension as a matter of satisfying various constraints formed through experience with language. For example, the spoken form of *righting* has been used a certain number of times as "righting" and a certain number of times as "writing." The immediate interpretation of *righting* in Lashley's example above will depend on the frequency of use of the alternative meanings of the word. Contextual constraints that depend on the frequency of use of a word in particular contexts, such as how frequently the sequence *rapid writing* has been experienced, apply as well to influence the immediate interpretation of a word. Syntactic constraints, such as how frequently an adjective like *rapid* precedes a nominalized verb (like *righting*—a verb based on the noun *right*), also influence the immediate interpretation of a word. Thus, comprehension involves the application of many kinds of habits simultaneously to determine the most likely interpretation. This procedure is often called *constraint satisfaction* (e.g., Rumelhart 1989).

Most connectionist theorists maintain that "rules" are mimicked by the network of processing units as a by-product of the constraint satisfaction process. This is expressed in terms of *pattern completion*. Incomplete or ambiguous information is "filled in" by triggering the closest available pattern that has been learned by the system. Since patterns can involve abstract units that connect to many actual parts of an utterance, the patterns can be quite complex and can approach representations of sentence-level dependencies.

A brief consideration of a connectionist treatment of object recognition illustrates pattern completion. If you see this book at a distance and obliquely, its retinal projection might be a crooked rhomboid with some correspondingly crooked markings. You will immediately see it as a rectangular "book" with normal writing on it. In connectionist terms what happens is that isolatable features of the oblique book activate selected features that are strongly connected to an actual book, four corners, a certain thickness, recognizable letters (e.g., *o* or *l*). All those features are best integrated from experience as part of a book, which is why that is what you perceive. This homely example sets us up for the formal requirement: *In order for pattern completion to work, the environment has to have real objects that can be experienced with variable frequencies*. The visual system cannot build up frequency-based activation patterns to objects that do not exist independently. Since physical objects do

exist independently, the models can work swimmingly well, and may indeed capture important features that are neurologically relevant.

But what about language? Where are the "objects" of language over which learning can occur? Utterances do not have a constant independent existence, and they certainly do not wear their internal structure on their surface. Thus, Chomsky's suggestion that simple S-R theory required a prior theory of the sentence is exactly relevant to connectionist models, for exactly the same reasons. The sentence level defines the fundamental object of language perception and provides the mechanism for modifying weights in the processing system.

We noted at the beginning of this section that it may be surprising to find that inductive models, including current sophisticated ones, require an independent theory of linguistic structures. It is surprising (and disappointing) only to those who wish to eradicate symbolic structures as relevant to mental models of ongoing behavior. We find it heartening that both systems of symbolic creativity and of habits converge on the same double-edged truth: we mostly behave out of habit, except when we do something novel.

1.2 The Integration of Habits and Symbols

This book is devoted to meeting the challenge of how to integrate the symbolic computational basis for language with acquired habits. The more specific focus is on how sentence-level syntax might be organized together with frequency-based perceptual templates to be efficient and to predict a wide range of empirical phenomena. One can view this as an example of the current goal of creating "hybrid" systems, which have elements of symbolic and spreading activation models. We explore a version of *analysis by synthesis* as a theoretically attractive model with a surprising array of both trite and unexpected empirical support.

Our first task is to consider the classic history of psycholinguistics and current models of comprehension. Chapter 2 reviews the trials and tribulations of the concept of the sentence over the past century. We give special attention to the experimentally grounded revival of the sentence level as an independent representation during comprehension, mostly due to George Miller and his students. We recount the rise and fall of attempts to treat linguistic syntax as a direct model of behavior, and the emergence of the notion of a frequency-sensitive component of comprehension. Through all this, an essential psychophysical feature of sentences remains true —*words are especially behaviorally compelling when they are arranged in sentences.*

Chapter 3 presents some essential facts that psychologists need to know about syntax. We present a sketch of modern syntactic theory, with as little jargon and technical apparatus as possible. The essential features are that syntactic operations apply to abstract categories, they include movement, and they occur cyclically over

sentences. That is, sentences have computational *derivations* underlying them. This property motivates some form of sentence-level application of syntactic structure, rather than a simple left-right one (for reasons related to Lashley's observations). It is also difficult to attach statistical information to entire sentence derivations, since they involve a series of abstract computational steps, they are not susceptible to direct reinforcement, and the derivations are not susceptible to direct modeling in constraint-based systems.

Chapter 4 reviews many recent and contemporary approaches to comprehension, focusing on the influential structural model of Marcus and its many witting inheritors, and on the equally influential associative model of Osgood and its modern unwitting inheritors. The reader may find that we miss some of the virtues of particular models because our focus is specifically on the ways structural and habit-based knowledge of language are handled. For some theorists, this is either an oblique or an obnoxious question. Despite many differences, there are some consistent grains of agreement across sets of models. In particular, both statistical and structural constraints are evident in language comprehension.

Chapters 5 and 6 present an analysis-by-synthesis theory of sentence comprehension and some basic evidence for it. The analysis-by-synthesis model offers a way to accommodate the facts that comprehension is both inductively statistical and computationally derivational. In this model, statistically valid perceptual templates assign an initial hypothesized meaning, which is then checked by regeneration of a full syntactic structure. Accordingly, the model proposes that we "understand" every sentence twice, once when we project an initial meaning-form pair and then again when we assign a complete syntax to it. Hence, we refer to our specific analysis-by-synthesis model as *Late Assignment of Syntax Theory* (LAST). The model is completely consistent with current syntactic theories that include inflected lexical items and semantic functional projections in early stages of a derivation. It is also consistent with recently developed evidence that statistical properties of sentences play an immediate role in comprehension, captured in the frequency-based perceptual templates that assign the initial meaning.

LAST offers interesting twists on a number of classic and recently developed psycholinguistic phenomena. We contrast LAST with the nearly ubiquitous "syntax-first" models, which assume that syntax must be assigned before meaning can be analyzed. Perhaps the most salient fact in favor of LAST is that people understand sentences immediately, yet a number of syntactic features appear to have a behavioral role very late in processing, in some cases after a sentence is over. This fact is puzzling from the standpoint of any "syntax-first" theory. It is important to note that the initial comprehension is not purely semantic and syntax-free. Rather, it is based on "pseudosyntactic" structures that can be reliably assigned based on superficial cues. This has the consequence that in some cases, sentences are initially under-

stood with an incorrect syntax that felicitously converges on the correct semantics. For example, we argue that passives are initially understood as complex adjectives. That is, the following sentence

(3) *Passive*
Clinton was impeached by Congress.

is initially assigned a structure like that of either of the following:

(4) *Adjectival sentences*
a. Clinton was impeachable by Congress.
b. Clinton was insensitive to Congress.

This incorrect assignment leads to a correct semantic interpretation, which in turn is part of the basis for later regenerating the correct syntax that reflects the passive construction, as in:

(5) *Passive with trace*
Clinton was impeached [t] by Congress.

In chapter 6 we report a variety of experimental evidence suggesting that indeed the correct syntax in passive sentences is assigned very late in comprehension.

Chapter 7 explicates how the model treats garden-path constructions—perhaps the single most pervasive object of study in today's psycholinguistics. A garden-path sentence is one in which the initially assigned structure turns out to be wrong. A frequent example from the last three decades of study in psycholinguistics is the *reduced relative* construction,

(6) *Ambiguous reduced relative*
The horse raced past the barn fell

which is much more complex perceptually than its corresponding unreduced relative:

(7) *Unreduced relative*
The horse that was raced past the barn fell.

or a corresponding unambiguous reduced relative:

(8) *Unambiguous reduced relative*
The horse ridden past the barn fell.

LAST explains the strength of the illusory complexity of the ambiguous reduced relative as a function of the application of a pervasive perceptual template that assigns simple declarative "agent-action-patient" patterns to sequences. In this case, the first salient organization of the sentence is like that in

(9) The horse raced past the barn. Fell.

which is hard to avoid even though the result is an ungrammatical sentence.

Indeed any property of the initial sequence that increases its salience as a simple sentence also increases the garden-path effect. This includes information about the animacy of the first noun, the conceptual fit of the first noun as an agent of the verb, and the kind of roles required by the verb, as well as other types of information. The role structure of verbs turns out to be a critical controller of how the garden-path effects appear and how they interact with context. We review much of the current experimental literature along with some new studies showing that most of the processing difficulties with reduced relatives occur only with verbs that are potentially intransitive, such as *raced*, and less so with verbs that must have an object, such as *frightened*.

(10) *Reduced relative with potentially intransitive verb*
The horse raced in the barn fell.

(11) *Reduced relative with transitive verb*
The horse frightened in the barn fell.

Chapter 8 focuses on applications of LAST. We return to the question of why the sentence level is a basic unit of analysis in comprehension. The NVN pattern and its variations, such as NV for intransitive verbs and NVNN for "double-object" verbs, is a powerful template just because the sentence is the fundamental unit in our analysis-by-synthesis model. The sentence is the object of pattern-recognition processes in comprehension, and therefore serves as the conduit for modifying associative connections. Thus, the existence of canonical sentence patterns solves the problem of isolating a relevant analytic level for inductive learning. This chapter reviews recent experimental evidence for sentence-level templates.

The next two chapters broaden the application of the model and integrate it with other systems of language behavior. Chapter 9 extends the model to multiclause and discourse-level structures. It is useful to think of comprehension as simultaneously building up meanings and structural representations at different levels of representation at the same time. This gives a special status to the ends of clauses, which is the point at which word-, sentence-, and discourse-level structures can be integrated together. A variety of behavioral studies show that clause boundaries indeed involve rapid swings in attention from being internally to externally oriented—that is, oriented toward mental activities or the world. It is in this context that we discuss the issue of the "modularity" of sentence-comprehension processes and corresponding experimental evidence. We will suggest that LAST renders the issue of modularity a nonissue, since the comprehension system can be seen as both modular and nonmodular at different points during comprehension.

Chapter 10 sketches theories of acquisition and the neurology of language. In each case, the goal is to explore the implications of and for LAST. We do not propose to present complete or even correct models of acquisition or the representation and

processing of language in the brain. Rather, our goal is to see if these behaviors give evidence for the kinds of distinct processes that we postulate in the analysis-by-synthesis model. We think they do.

A natural model of acquisition has an analysis-by-synthesis form, in which children continually create structural representations for the sentences they can understand, which in turn are extended by statistically valid generalizations to understand more kinds of sentences. A model of this kind emphasizes the dual role of innate (or easily available) structural descriptions and statistically validated generalizations. It also offers a potential explanation of how the analysis-by-synthesis model of comprehension is naturally acquired.

The most stable neurological property of language representation is that it has a unique relation to the left hemisphere. Examination of some data from aphasics and some developmental data suggest that what may actually be lateralized is the *pseudosyntax*, the set of initial operations in the formation of an immediate initial structure and meaning. Knowledge of actual syntax might be represented more diffusely. This could explain why certain aphasics can make syntactic grammaticality judgments about sentences that they cannot understand.

The neurological experimental data we focus on primarily are evoked brain potentials, which can be collected during comprehension and language behaviors of other kinds. A common contemporary method is to introduce an anomaly of some kind into a sentence and study how long it takes to have a measurable effect, and what kind of effect it has. This allows contrast between quite local features, such as inflections, and global syntactic properties. Intriguingly, the evidence suggests that anomalies in features involved in pseudosyntax have immediate effects, while derivational syntactic properties are detected much later. The distinction between the kinds of syntactic features and the timing of their computation is exactly predicted by the analysis-by-synthesis model.

We hope that this book serves several purposes. First, we review a large segment of classic and current psycholinguistic research and theory. We also outline how current syntactic theory can fit well with behavioral theories in general. Most generally, we offer and adduce evidence for a model that integrates structural and habit-based knowledge. We hope that this inspires others to develop corresponding models in other domains of cognitive science.