Chapter 10
Implications

We have presented a model of language comprehension that incorporates and integrates both habit-based and rule-governed aspects of language. This model has implications for the broader field of cognitive science. In this final chapter, we first summarize the main points of our model, the historical background for it, and the evidence for it, roughly in the order of the preceding chapters. Then we review some topics on the implications of our model for cognitive science as a whole.

10.1 Summary

There are two foundational approaches to the study of sentence comprehension, the habit-based approach of behaviorists such as Skinner and Osgood and the rule-governed approach that emerged from the work of Miller and Chomsky (chapter 2). The greatest advantage of the habit-based approach was its characterization of the associative nature of many aspects of language behavior, and its goal of explaining the acquisition of language behavior through inductive mechanisms. In the end, however, the habit-based approach of the 1950s and 1960s failed largely because it did not have sufficiently sophisticated algorithms for modifying associative strength, it was limited by behaviorist restrictions, and it had no means of identifying the appropriate unit of reinforcement.

During the same period, experimental research on language perception gave special prominence to the sentence level of representation and the view that the natural unit of comprehension is larger than the word but smaller than the discourse. The natural unit of comprehension was thought to be the sentence, which was defined by a set of generative rules. We observed that a critical and enduring fact is that words seem to be physically clearer when they are arranged in sentences. This basic psychophysical phenomenon heightened the importance of understanding the psychological mechanisms for understanding and manipulating sentences.

The then-nascent transformational grammar offered a generative model for sentences, which sparked two decades of research on the "psychological reality" of

grammar. The outcome of this research was that the relation between grammar and comprehension was seen as indirect. Consequently, psycholinguistic models became more detached from linguistic theory.

Psycholinguistic research has been further complicated by rapid changes in grammatical theory. However, most syntactic theories since the 1950s share some central ideas about language (chapter 3). First, sentences have a hierarchical structure, in which larger units consist of two or more smaller units. Second, sentences have derivations, in which rules are applied to abstract structures that modify the linguistic representation. Third, derivations are cyclic, operating over sentence-like units. These theories converge on the view that a representation of meaning is based on the content of a sentence. The evolution of syntactic theory reflects greater insight into how best to describe the relation between the meaning of a sentence and its form. In the generative framework, meaning was based on an initial inner form, the "deep structure" of sentences, which was deformed by transformations under a surface representation. This was then supplemented by sensitivity to aspects of surface form, with concomitant constraints on the derivational relation between the deep and surface structures. More recently, it has become clear that there are many formal and empirical advantages to building up a sentence syntactic derivation, phrase by phrase, with concomitant intermediate stages in access to the meaning representation.

Current approaches to sentence comprehension come in two basic flavors, rule-based and habit-based (chapter 4). Modern variants of Marcus's rule-based model adopted the view that syntactic structure is essential and logically prior to the construction of sentence meaning. Accordingly, these models assume that the apprehension of syntactic structure comes first in the understanding of sentences. There are also modern variants of habit-based models, which adopt the view that sentence comprehension is primarily a process of applying associative knowledge that is represented as connections between simple processing units. Connectionist models have made tremendous progress in solving the methodological half of Osgood's dilemma: the development of sophisticated mathematical procedures for distributing feedback to associative connections, so that the model converges on language-like behavior. However, connectionist models have not been able to solve the other half of the dilemma that faces all inductive associative models: What is the appropriate linguistic unit that determines how feedback is distributed?

This set the stage for the development of a model of sentence comprehension that integrates the two types of linguistic behavior, an analysis-by-synthesis model that we call *Late Assignment of Syntax Theory* (LAST, chapter 5). LAST uses associative, habit-based knowledge, or pseudosyntax, to form an initial hypothesis about the relation between meaning and form. This initial representation of meaning and form is the input to a synthetic component. The synthetic component consists of abstract grammatical knowledge that generates a full syntactic representation and its cor-

responding meaning. The comprehension system compares the output of this full syntactic derivation to a stored representation of the input sequence. If the two representations match, the corresponding meaning is stored and the syntactic details become less available.

LAST has two critical implications. First, it is completely compatible with current syntactic theory. The output of pseudosyntax consists of overt lexical items, basic phrase grouping, and an associated meaning that specifies basic grammatical values and semantic functions. This pseudosyntactic representation corresponds well to the input to a derivation in current syntactic theory, the "numeration."

A second critical feature of LAST is that we understand sentences twice. The first time, an initial meaning-form pair is based on associative connections and templates in the pseudosyntax. The second time we understand a sentence it is based on the meaning-form pair derived by the grammar. Successful comparison of the derived form and the stored input confirms the final interpretation. This process explains a number of phenomena including the most important classical finding that words are clearer when they appear in sentences. In the dual-representation model, two surface forms are created and compared with the input, thus focusing extra processing and attention to the physical input.

There is considerable classical evidence for LAST (chapter 6). Current theories can explain different parts of this evidence, but LAST explains all of it without adding ad hoc assumptions. The initial representation of the relation between form and meaning approximates the final representation. It can be syntactically incorrect, as in the initial representation of passive sentences as adjectival, but "good enough" to arrive at a likely meaning-form, which then sparks a complete derivation (cf. Ferreira and Stacey 2000). The evidence shows that there is rapid access to low-level phrases, which are established in part by using function words and morphemes to identify the boundaries of phrases. In addition to function words and morphemes, pseudosyntax consists of sentence-level templates like NVN = agent-action-patient and its fuller form based on learned argument requirements for verbs, NV (N) (N) = agent-action-patient-recipient. The presence of explicit markers for *wh*-movement enables the pseudosyntax to identify the location of *wh*-traces. The evidence also shows that subcategorization frames and knowledge about word meanings quickly influence sentence processing. The synthetic component uses the initial representation of meaning and pseudosyntax to generate a detailed syntactic representation. The synthetic process is similar in certain ways to the processes of recalling or producing sentences. Some aspects of the syntactic representation, such as detailed phrase structure and the source location of moved noun phrases, become available only late in the process of sentence comprehension, after the grammar has generated a complete syntactic representation and the corresponding meaning. The results of studies that examined introspection about sentences were found to depend on whether the

task uses prospective introspection, which precedes full comprehension, or retrospective introspection, which follows full comprehension.

Sentence-level templates in pseudosyntax have tremendous explanatory power (chapter 7). The sentence template NVN = agent-action-agent and its variants NV (N) (N) = agent-action-(patient)-(recipient) can explain virtually all the experimental evidence on sentences with a reduced relative clause as in *the horse raced past the barn fell*, much of which has appeared contradictory. The reliance of pseudosyntax on sentence-level templates also explains a vast amount of the literature on full subject and object relative clauses and garden-path sentences other than reduced relatives. In each case, the comprehension difficulty occurs because of the initial application of the NVN sentence template and its variants.

The architecture of LAST entails that representations of the relation between meaning and form appear twice during comprehension, once in the initial representation of meaning based on pseudosyntax, and once in the complete representation of meaning based on grammar. This architecture has implications for the issues of syntactic modularity and discourse processing (chapter 8). We noted that LAST can be seen as both nonmodular, in the formation of the initial hypothesis about meaning and form, and modular, in the generation of a complete syntactic representation, which uses grammatical knowledge exclusively. The formation of two representations of meaning and form also has implications for the processing of multiclause sentences and discourses, since, in principle, representations of both meaning and form may be available for integration during ongoing sentence processing. There is corresponding evidence that ongoing comprehension involves fluctuations of attention to different kinds of representations depending on the requirements of the comprehension system as described by LAST. As far as we know, much of this evidence remains unexplained by other theories, except by ad hoc assumptions.

LAST also has implications for the acquisition and neurological representation of language (chapter 9). Much evidence from acquisition indicates that children develop early sensitivity to pseudosyntax, including sentence templates and grammatical morphemes. A review of studies of neurological processing during sentence comprehension reveals a surprising amount of evidence for LAST. In general, pseudosyntax and aspects of associative meaning produce very rapid brain responses, while more detailed syntactic representations produce relatively slow brain responses. Evidence from studies of agrammatic aphasia also can be interpreted naturally in terms of LAST.

Our discussion has focused on the need for a model of comprehension that integrates habit-based and rule-based knowledge, the organization of such a model, and the evidence for it. Along the way, we have referred to critical and complementary failures of habit-based and rule-based approaches, such as the grain problem for habit-based approaches and the learning problem for rule-based approaches. We turn

now to a summary of these issues and some broader implications of LAST for an integrated theory of language behavior and consciousness.

10.2 Cognitive Architecture, the Grain Problem, and Consciousness

Sentence comprehension may be one of the more complex cognitive capacities of humans. But its principles may have analogues in other cognitive behaviors. We now explore three implications of the analysis-by-synthesis model for cognitive science in general.

10.2.1 Habits and Rules Are Segregated in Modules

A constant theme of this book has been the construct validity of both habit-based and structural explanations of language behavior. In general, most of ongoing language behavior has a clear habitual component, while it also affords the possibility of constructing and understanding quite novel utterances. It has long been a puzzle how to integrate the two kinds of information in a single approach. Indeed, much of the antecedent history of cognitive science has involved radical swings back and forth between statistical and structural treatments of all behaviors, roughly three decades at a time for each approach since the turn of the twentieth century.

For example, the cognitive revolution of the 1960s was a triumph of rationalist structuralism over behaviorist associationism that had dominate the field since the 1920s. Abstract structures, such as the level of representation of sentences, were recognized by all as a dangerous challenge to the ''terminal meta-postulate.'' The *terminal meta-postulate* refers to the behaviorist claim that all theoretical terms and all corresponding inner psychological entities are grounded in more superficial entities, and ultimately in explicit behaviors and stimuli (see Bever, Fodor, and Garrett 1968). There is no obvious way in which inner syntactic forms are extracted from outer ones. Indeed, syntactic theory suggests that the relation is in the opposite direction. Rational structuralism sharpened the focus of research on the behavioral role of such inner forms, and gradually behaviorist strictures lapsed into the obscurity of irrelevance.

But arguments against behaviorism are not direct arguments against associationism. Associationism is merely the doctrine that all mental relations are associative rather than categorical, usually built up out of repeated experiences or co-occurrences. The compelling argument against the idea that absolutely everything is based on associative habits is the complexity and underdetermined nature of normal behaviors, such as the comprehension of language. The traditional answer by associationists has been that we underestimate the subtlety of behavior that a large complex associative network can compute. Even inner, abstract representations might be connected by a myriad of associations that is so complex as to mimic the appearance of hierarchies, part-whole relations, relations at a distance, and other

kinds of categorical facts. The thrust of connectionism, the major school of modern associationism, is to apply relatively complex networks of associatively connected simple processing units, and thereby explain apparent categorical and rule-like structures as the result of pattern completion and the exigencies of associative information compression. The idea is that with enough experience, associative networks produce behavior that looks structural in nature, but we allegedly know it really is not because we can manufacture a machine of associative networks in computer model of selected language behaviors.

That, of course, is just one of many human conceits. We manufacture electric generators, but we do not make the laws or the electrons that the laws govern. Rather, we construct machines that manipulate those phenomena to our own ends. Similarly, suppose that we wired up an associative network as complex as a human brain and placed it in a completely human environment and that it started to learn language. The process by which the associative network learns language would be as much a mystery as how a child learns language. We would be back to the 1960s' cognitive psychology joke. Someday the researchers in artificial intelligence will create a very human-like robot, and then we will have to practice psychophysics, theoretical psychology, and experimental "neuro"science on it to figure out how it works. We will have new sections of international psychological associations in "robopsychology" and the closely related field, "virtual cognitive neuroscience." In short, our creation of a language-learning machine would not mean that we have created language or the laws of language learning.

Similar modesty is indicated about mimicking the achievements of specific connectionist models. The ability of a model to converge on 90 percent accuracy in computing a specific syntactic structure after thousands of training trials or to differentiate lexical classes based on repeated exposure to local contexts is an important achievement. Such an achievement stands as an existence proof that statistically reliable information is available in the language stimulus world and can support the induction of recognition patterns. But to claim that such achievements show we can account for actual categorical syntactic structure is not warranted. It would be like claiming that the amazing height of a medieval church spire shows that humans can reach the heavens. The church steeple is a prototype of humanity's ambition, but the ambition itself refers to categorical concepts.

So, after the cognitive shouting and bombast of the 1960s and the more recent connectionist enthusiasms are absorbed, we are left with the same truth as before: Behavior is affected both by categorical and statistical processes. Some current models in cognitive science respond to this by proposing "hybrid models" that commingle elements of symbolic and associative processes. This can create conceptual puzzles if the two kinds of information are truly blended. What does it mean to state that a "rule" applies a particular percentage of time, especially if the rule is a stage in

a complex derivation? What does it mean to state that an "associative regularity" defines a structural category? In both cases, the attempt by one kind of information to modify the other can lead to theoretical incoherence.

The analysis-by-synthesis model we have sketched allows both kinds of information to play explanatory roles in behavior, and much of this book has been devoted to showing the empirical power of this model. But while it can be counted as a "hybrid" model, it clearly segregates the two kinds of information into separate computational domains. This maintains theoretical consistency and clarity, while allowing each kind of information to explain what it can explain in actual data.

10.2.2 The Grain Problem and Its Solution

The basic proposal of induction is that experience reinforces certain internal states, which then modify future behavior. This principle runs through the work of Skinner, Osgood, and current connectionist models. The problem has been repeatedly noted, starting with Chomsky's review of Skinner's book on language, that associative models generally do not have independent definitions of what counts as a stimulus in the world and a response pattern to it. That is, these models do not define what should be connected by each reinforcing experience independently of the intuition of an observer who already possesses the knowledge to be learned. In research on language behavior, this has become known as the "grain problem," the question of what level of linguistic analysis is confirmed by each environmentally successful use of an utterance.

We noted in chapter 5 that the analysis-by-synthesis model offers a very specific solution to the grain problem. There are two sets of constraints that must be met for the model to work. First, the output of the associative pseudosyntax must provide what is needed for the input to the syntactic derivation. In current grammatical terms, this includes a list of the (inflected) lexical items, phrase information, and information about functional relations, the latter most relevant to sentence/clause-level thematic roles. Accordingly, the levels for learning include mapping low-level information relevant for isolating words, higher-level information for grouping phrases, and semantic-pattern information relevant to sentence-level thematic roles. The last kind of information naturally emphasizes sentence/clause-level structures. This emphasis on sentence/clause-level structures is further confirmed by the fact that the sentence/clause is the natural unit of a complete grammatical derivation, produced by the syntactic component. The result is that the dominant focus of behavior is on sentence/clause-level information. Interpreted as a learning device, the analysis-by-synthesis model will reinforce connections between input information and the assignment of sentence-level structure.

This solves the grain problem for language comprehension. What we show here is that it is possible to extract an answer to the grain problem from the intrinsic struc-

ture of a behavioral model. But is this idea generalizable to other kinds of cognitive behaviors? It would be in the spirit of learning theorists to expect that there will be universal features, following on the general idea that there are common laws of learning. That is for others to argue in relation to specific models of other behaviors. Perhaps the solution will be different in every modality; perhaps it will have fundamental universal features that are similar to those we have developed.

10.2.3 Consciousness: The Universality of Analysis by Synthesis and Multiple Drafts

We close this book with a discussion of one of the broadest problems in cognitive science: consciousness. In its preoccupation with issues of structuralism and associationism since the 1960s, psycholinguistics has tended to be quite dry. Big psychological issues such as "beauty" or "happiness" are usually untouched by linguists and psycholinguists alike. A sensible interpretation of this situation is that the disciplines of linguistics and psycholinguistics are highly rarefied and technical. These disciplines illustrate an aspect of the structure of knowledge and a sketch of how that structure is deployed in certain behaviors. But even the most technical matter, such as the architecture of a model of sentence comprehension, may have extensions to broader questions.

Some of today's thinkers on consciousness divide it into technical (also known as "easy") problems of cognitive science, and traditional (also known as "hard") philosophical problems of epistemology. A leading example of the former is Dennett (1991), and a leading example of the latter is Chalmers (1996). We leave the philosophical problems (whatever they may be) to the philosophers. But we find the cognitive science approach to be compatible with our model, and we think that our model may have something to contribute to the problem of consciousness. In Dennett's model, what we think of as conscious representations of the world are actually a set of partially independently formed representations, some of which supplant the others in time, some of which merge. Each representation is a "draft" of reality, which blends in a stream of awareness and occasionally crystallizes into a relatively permanent form. Thus, "consciousness" does not occur in a set mental place, where we project reality to ourselves. Rather, what we mean and experience as consciousness is itself dynamic and self-reforming. Dennett draws on many different kinds of facts from cognitive science to support his view. We think our model has a small additional contribution to make to this enterprise.

We have sketched a comprehension model that chases its tail at least once. It starts with an associatively primed meaning, and then uses that to begin a syntactic derivation that rearrives at a detailed surface form and syntactically enriched meaning. This presents the comprehension process as a set of emerging representations of meaning and structure that converge. We have sketched the representations as

occurring in a series, but of course, they could be computed partially in parallel in many actual cases.

Thus, in our model we understand everything twice. This claim flies in the face of our phenomenological experience that we understand sentences once. Intuitively, we think we understand a sentence and are conscious of it and its meaning as a unified experience. Yet, the analyses and experiments we review in this book suggest that this apparently unified process is actually composed of converging operations of quite different kinds. The lack of consistency between our analytic and experimental results on the one hand and our phenomenological experience on the other is a problem that we eventually will have to face.

We can infer that our model of comprehension is an instance of Dennett's "multiple-drafts" model of consciousness. As in the multiple-drafts model, one representational draft has prominence over others, but all drafts are there as a part of the process. The compatibility of our model with the multiple-drafts view of consciousness allows us to enrich methods of studying how the drafts can work together and apart. It is a natural extension of our comprehension model to interpret the initial associative meaning-form hypothesis as one draft of a sentence and its meaning, and the syntactically derived representation of meaning-form as another draft. Thus, our model allows us to specify the nature and structure of each kind of draft, and to study how the formal attributes of each interact with conscious intuitions about them. A particular puzzle about the multiple-drafts theory of consciousness is the frequent intuition that one draft seems to have priority over the others—that is, one draft seems to be the "real" representation. In the case of sentences, the final computed sentence may appear to be the "real" one. What might this tell us about which kinds of drafts dominate others as part of reportable consciousness? There are various options:

- The "latest" draft seems to be the most real by co-opting those that precede it.
- The most "complete" draft seems to be the most real.
- The draft that depends most on internal computations seems to be the most real.

In a real machine, these explanations tend to co-occur. The latest draft is often the most "complete" by virtue of having had more internal computations involved in its production. But it is possible to test at least some of these alternatives by varying the computational complexity of the linguistic stimulus in specific ways, and studying the effects on its perceptual clarity. For example, we can make sentences that are relatively complex in terms of the associative pseudosyntax by violating canonical forms. Or we can make sentences more complex in terms of the formal grammar by requiring longer derivations. We can then examine the behavioral effects of different kinds of complexity. This may lead to some surprising predictions, in particular, that sentences that are more complex actually result in greater acoustic enhancement so

long as they are still comprehensible. A similar prediction may follow from certain ungrammatical sentences, namely, those that seem plausible at first.

Such predictions are highly unintuitive, which is why we like them. We have emphasized throughout this book the importance of being able to predict and explain phenomena that are prima facie unlikely or very puzzling. Any old theory can explain the obvious. We look ahead to predicting and explaining the nonobvious. This is the real meaning of the much-misunderstood saying: the exception proves the rule.