Chapter 4

Contemporary Models of Sentence Comprehension

The previous two chapters outline the logical arguments and "classic" evidence that knowledge and use of language involves sentence-level syntactic structures. In this chapter, we take up current theories of whether and how syntactic structures are deployed during comprehension.

4.1 The Problem of Comprehension

When we understand a sentence we are hardly conscious of its structure: rather, we usually grasp its "meaning" and move on to the next sentence, leaving behind the words, the phrasing, the entire syntactic architecture. For example, it is true of (1a) that it has the word order in (1b), the structures in (1c), and the meaning depicted in (1d). But we are ordinarily aware only of the literal word sequence and the meaning.

(1) a. The horse raced.
    b. the, horse, raced
    c. "the (= determiner) + horse (= noun)" = *subject*
       "raced" (= verb, past participle) = *predicate*
    d. [conceptual representation of horse racing]

This focuses us on the main theme of this book: How do syntactic structures support language comprehension? There are two lines of thought about this:

1. Meaning is assigned by reference to contextual and internal semantic constraints. Syntax, at most, is a perceptual or cognitive afterthought.
2. Syntax is assigned logically and, in fact, prior to meaning. Syntax is an unavoidably automatic perceptual prerequisite.

Similar conflicts have occurred in other perceptual domains, where critical experiments are possible. For example, there are two opposing views on the relationship between visually recognizing an object and perceiving all the details of its contours:

1. We first recognize the object and then "fill in" its shape.
2. We first perceive a shape and then recognize the object.

In the case of vision, one can come close to distinguishing these alternatives experimentally. At one extreme, it is possible to study the perception of "objects" (e.g., a face) with blurry or no contours. At the other extreme, one can study the perception of lines, spatial frequency gradients, or "contours" that do not correspond to real objects.

In the case of language, it is far more difficult to distinguish meaning- versus syntax-based comprehension theories. *Every* sentence has some meaning, even (2a), and meaningful sequences, such as (2b), can be interpreted, though they are relatively syntax-free.

(2) a. Twas brillig and the slithey toves did gyre and gymbal.
    b. Politician speak with forked tongue.

Thus, as a practical matter, we cannot contrast the perception of meaningless but well-formed sentences with meaningful ungrammatical sequences. We are consigned to indirect demonstrations and logical arguments. The most powerful logical argument would be a working comprehension model that either rests only on meaning, or only on syntax. Such a model would be an existence proof of the possibility of the corresponding claim about actual comprehension.

The problem is that we have a surfeit of alleged comprehension models, based on contradictory principles, and none of them actually works. Consider two extremes, a pure semantic model and a pure syntactic model as they would apply to (1a), which we repeat here:

(1) a. The horse raced.

In the semantic model, each lexical item has a unique meaning that it carries independently, as in (3). On purely semantic grounds, there is only one way they can be put together into a meaningful thought. Neither syntax nor sensitivity to surface order is required to arrive at the appropriate meaning.

(3) a. *The* = an aforementioned entity
    b. *horse* = domestic four-footed animal ...
    c. *raced* = run very fast, past tense

In the syntactic model, each word initially is categorized and placed within a phrase, and phrases are assigned functional labels. In most models, this process occurs from left to right, as each new word comes in (see figure 4.1).

With sentences of normal complexity, it immediately becomes clear that at least surface order must be attended to, even in a semantics-based model. In (4a), the individual words can be put together in only one sensible manner, but in (4b), the word order makes all the difference.

the	*horse*	*raced*
the = determiner		
	the = determiner *horse* = noun *the horse* = NP	
		the = determiner *horse* = noun *the horse* = NP *raced* = verb *the horse* = subject *raced* = predicate

Figure 4.1
A parse of a simple sentence based on lexical information.

(4) a. The horse raced past the barn.
b. The horse raced past the greyhound.

Furthermore, certain words have meaning only in regard to the structural functions they convey, which a semantics-based model must also respect. In (5a), *was* must be recognized as a "past passive marker" that inverts the normal effect of word order, in this case making the initial noun the patient rather than the agent of the action. In (5b) and (5c), *that* must be recognized as a modifying clause marker, marking the following verb as a modifying predicate rather than a main predicate. Such examples show that at best, a pure lexical semantics-based model cannot succeed without a significant syntactic armamentarium.

(5) a. The horse was raced past the greyhound.
b. The horse that raced past the barn fell.
c. The horse that was raced past the barn fell.

Conversely, it often appears that syntactic assignment relies crucially on conceptual knowledge of the world. In (6a), the adjunct prepositional phrase is attached to the sentence node and modifies the initial noun phrase; in (6b), the corresponding phrase is attached to the second noun phrase, which it modifies. Yet the difference in level of modification is entirely conveyed by our world knowledge.

(6) a. The jockey raced the horse *with a big smile*.
b. The jockey raced the horse *with a big mane*.

A similar point is made by the contrast between (7a) and (7b), both parallel in structure to (7c), in which *the horse* is unambiguously the patient of the first verb.

While (7b) may be a bit confusing, it is nowhere near as difficult to comprehend as (7a). The only difference is that it is very unlikely that a horse can lasso anything, and simultaneously very likely that it is the patient of that patient-requiring verb; thus, the comprehension process is lead into the correct interpretation by virtue of this conceptual knowledge. In (7a) the corresponding knowledge does not help, because, in fact, horses do race at least as often as they get raced—in fact more often, by definition.

(7) a. The horse raced near the barn fell.
b. The horse lassoed near the barn fell.
c. The horse ridden near the barn fell.

Finally, even very basic segmentation aspects of syntactic structure may be controlled by local semantic knowledge. In both (8a) and (8b), *the socks* must be interpreted as the agent of *fell*. But in (8a) it is more likely to be miscoded as the patient of *darn*, and thereby attached as the object of the preceding verb. This is much less strong following *wiggle*, with correspondingly less miscoding.

(8) a. When Mary was darning *the socks* fell off her lap.
b. When Mary was wiggling *the socks* fell off her lap.

The inescapable conclusion from these examples is that comprehension involves a combination of knowledge about the likelihood of conceptual combinations with knowledge of possible syntactic structures. It is not reasonable to argue that comprehension is based entirely on either conceptual or syntactic knowledge alone, but it is possible to specify an architecture in which one or the other kind of knowledge has logical and temporal priority of application. Thus, contemporary theories still fall into two classes: those that emphasize the priority and independence of rule-governed syntax, and those that emphasize the priority and independence of frequency-based associations.

4.2 Structural Models

The usual architecture in a structural theory first assigns the syntactic structures, and then semantic analyses. Semantic facts about the analyzed meanings may ultimately lead to choice of syntactically ambiguous structures or rejection of conceptually anomalous ones. But, in general, "syntax proposes, semantics disposes" (Crain and Steedman 1985).

By definition, the only information available to a structural parser is the syntactic category and related syntactic frame information conveyed by each word. The sparse nature of lexically carried syntactic information underlies the central problem of a left-to-right online syntax assignment scheme:

- At almost every point, there are multiple possible syntactic assignments.

Consider first (1a), which we repeat here:

(1) a. The horse raced.

In figure 4.1 we assumed that it had only one assignment that was unambiguous, as the parser moves through it. But consider the actual state of affairs as the process proceeds from left to right. Even at the word *horse*, the continuation could be quite different, one in which *the horse* is not a separate noun phrase, as in sentences (9a) and (9b).

(9) a. The horse races ended.
b. The horse and buggy raced.

That means that at the word *horse*, more than one structural option must be available. In a case like (9a), the next word—*races*—would seem to determine which option is correct: *horse* is part of a compound phrase, with *the horse races* being the actual noun phrase. But this, too, can be wrong, as in (10a) and (10b). In these cases, the original decision that *the horse* is a phrase, rescinded at the following word *races*, is ultimately correct.

(10) a. The horse races frighten whinnied.
b. The horse races usually frighten whinnied.

Accordingly, in almost all sentences there are multiple options at many points. This provides a dimension on which specific parsing architectures can differ. On the one hand, the parser can assign the "best" structure and be configured to allow multiple backtracking, or the parser can assign multiple structures in parallel and wait for the ultimately irrelevant ones to be excluded. We return to modern studies of this question later in the chapter. We start with a specific proposal by Mitchell Marcus (1980) that laid out most of the issues and many of the proposals that have dominated the structure-based approach for more than two decades. It is not our purpose to explicate the Marcus model with its original detail or nomenclature. Rather, we now review some salient properties of it, which are relevant to subsequent structural models.

Marcus argued that models that make incorrect assignments and then backtrack, or that hold alternate parses in parallel, simulate a "nondeterministic" model. He noted that these approaches set no limits on what a parsable structure could be, which means that they cannot aspire to be explanatory models, nor can they reveal how parsing constraints themselves might shape the form of language. He suggested that the parser should be "deterministic," that it should make only structural commitments that it never recants. Part of the motivation for this is to explore what such a parser must have in order to work. That is, the deterministic constraint actually weakens the power of the parser, and makes more revealing the particular properties it requires. A related motivation is to explore the extent to which a limited

The Active Node Stack

S16 (S DECL MAJOR S)/(SS-FINAL CPOOL)
[Rule for Declarative Sentence]
NP: (John)
AUX: (has)
VP: ↓

C: VP14 (VP)/(SS-VP INF-COMP CPOOL)
VERB: (scheduled)

The Buffer

1. NP41 (MODIBLE NS INDEF DET NP): (a meeting)
2. PP11 (PP):(for Wednesday)
3. WORD133(*.FINALPUNC PUNC):(.)

Yet unseen words: (none)

Figure 4.2
A snapshot of Marcus's parser (adapted from Marcus 1980).

parser can explain certain linguistic phenomena as a function of the parsing limits themselves.

Marcus outlined a model, *Parsifal*, with two major data structure components: an active-node stack of incompletely analyzed constituents, and a buffer of stored completed constituents. As he put it, the active-node stack is a set of high nodes in a phrase structure tree, awaiting nodes to dominate. The buffer is a set of low nodes, with analyzed material they dominate, awaiting dominating nodes to complete a connected tree. At each word, one or more nodes are placed in the active-node stack, and as they are filled, the completed nodes are moved to the buffer. For example, the beginning of sentence (1a) would access an S and a subordinate NP in the active-node stack at the initial word *the*. The word *horse* could then complete the NP, which could be sent to the buffer, allowing for new incomplete high nodes to be activated.

Figure 4.2 illustrates a snapshot of the model during a parse. There are two data structures. The *active node stack* has two memory cells, while the *buffer* has three memory cells.

One cell of the active-node stack contains structures that are attached to higher-level nodes but that need to have additional lower-level nodes (daughters) attached. The second cell in the active-node stack is a workspace for assembling constituents. This is called the *current active node*. The active-node stack operates as a pushdown stack in which the last item entered is the first retrieved. Thus, when a constituent is assembled but has not been attached to a higher node, it is dropped into the second data structure.

The second data structure is the buffer. The buffer consists of three cells that contain words or phrases that need to be attached to a higher node. The buffer corresponds to the look-ahead component of the parser. Each buffer cell can hold either a word or a complete multiword constituent.

So far, this model seems vulnerable to the difficulties enumerated above, basically that there are too many options at each point. The model is able to parse deterministically because it can examine cells in the active-node stack and the buffer. By examining the active node stack, the model is able to look down the tree to see where it is going; by examining the buffer it is able to look past already-completed nodes. This gives the model the power to consider more than one level in the hierarchy the grammar is building, and to skip over completed constituents. If the model could consider any number of active nodes and of completed constituents it would surely be deterministic, and would never make a mistake. In the limit, it could consider an entire sentence before definitively assigning any grammatical rule-based structures. However, to reduce Parsifal's undifferentiated power, Marcus limits the number of nodes that can be examined in the active stack to two, and in the buffer tothree.

This framework has a number of features relevant for current models.

I. The grammar provides a rich annotated surface structure.

Every structural parser must presuppose some kind of grammatical framework that defines the possible and particular grammatical structures that the parser must assign to each well-formed sentence. That is, there is a "covering grammar," which describes sentences and their full descriptions independently of the mechanisms of the parser. For this, Marcus chose an "annotated" phrase structure, roughly the surface structure of the then-emergent government-and-binding theory (Chomsky 1981; see sections 2.5, 3.4.1). The model includes the usual phrase hierarchies and grammatical relations such as "subject" and "object." For example, in sentence (11a), *the horse* and *raced past the barn* are phrases, with further internal structure for the latter.

(11) a. (The horse) = subject (raced (past (the barn))).

In addition, there are two notions of "trace," inaudible noun phrase elements that connect overt noun phrases with distant locations. For example, (11b) would be given an additional kind of structure, an NP-trace that marks the fact that there is an object relation assigned to *the horse* in its underlying structure, as well as the subject relation in the surface structure.

(11) b. (The horse)$_1$ = subject (was (raced (NP$_1$) = object (past (the greyhound)))).

"*Wh*-trace," another kind of trace, links questioned phrases to their grammatical relation in a clause:

(12) a. Who_1 did wh_1 race the horse?
b. Who_1 did the horse race wh_1?

II. Rules for constructing constituents occur in hierarchical packets, which result in giving simple declarative main clauses default status.

Another critical part of the parser is a set of pattern-action rules (sometimes called *productions*). One of these rules applies when the contents of the active node stack and the buffer match the conditions (or "structural description") of the rule. When a match occurs, the parser either attaches the contents of a buffer cell to the contents of the active node stack, or creates a new node in the active node stack.

To take a simplified example that ignores details, the presence of *the* triggers an application of a left–noun phrase bracket before it. The presence of a noun phrase bracket at the beginning of a sentence triggers an application of a sentence clause node and also the assignment of "subject" to that noun phrase. The presence of a verb following a noun phrase triggers the formation of a verb phrase, which then triggers an attachment to the S-node, and the default assignment of "simple declarative" sentence.

The fact that the conditions that trigger rules can trigger parallel rules in cascades, motivates grouping rules into "packets," in part defined by the triggering conditions they share. Packets themselves can be assigned rankings of priority or can be given default status. For example, the packet that assigns simple declarative sentence status at the potential main verb is the default—in other words, what happens if no other condition has triggered a more exceptional rule application.

Marcus notes that packets and rule priorities are not strictly necessary, nor are they a crucial component of the fundamental motivation or architecture of the model. Rather, he argues that they make the model work much more efficiently by capitalizing on statistical facts about the likelihood of particular constructions. The moral we draw from this is that even structural models may include statistical information to their advantage. In addition, we note that the prime statistically predominant construction from which others depart is the simple declarative transitive sentence.

III. Noun phrases are easy to isolate, based on surface cues.

This is important to the model, because many of the rules that assign structure at the clause level must presuppose that noun phrases have been isolated and already stored in the buffer. In effect, the parser can shift the buffer within a sentence, find noun phrase left edges, segregate and store the noun phrases, and then return to its original preshift serial location for higher-level structure assignments.

As a result, Marcus has two stages of syntactic processing: assembling constituents and attaching them. Phrases are first assembled, then attached to higher nodes. In a sentence like (13),

(13) Is the block sitting in the box red?

the word *is* initiates the creation of an S-node and is attached to the S-node in the active node stack as auxiliary in a yes-no question. When the noun phrase node is completed NP, it is attached as well to the S-node. The phrase *sitting in the box* is assembled as a verb phrase in the second cell of the active node stack but is not attached initially, because it may be the main verb phrase or a reduced relative clause. Hence, the verb phrase is dropped back into the first cell of the buffer. Once the word *red* is read, the parser recognizes that *red* can serve as the main predicate with *is*. At this point the verb phrase is attached to the noun phrase *the block* as a reduced relative clause.

Of course, being able to shift like this changes the intuitive clarity of the limit of considering three nodes in the buffer. It is still technically the case that the limit holds for any given constructive operation. The general importance of this feature is that it emphasizes the role of various closed-class morphemes and syntactic category patterns that define the left edges of noun phrases.

IV. NP-traces are treated as normal noun phrases stored in the buffer; wh-*traces are treated in a special* wh-*comp position.*

It is characteristic of language that noun phrases can have a role in more than one clause. In (14a), *horse* is the subject/object of *proposed* and the subject of *race*. In (14b), it is the subject of both *raced* and *fell*.

(14) a. The horse was proposed to race.
b. The horse that raced fell.
c. *The horse$_1$ was proposed Bill to have raced NP$_1$
d. (The horse)$_1$ = subject ((that)$_1$ (wh_1) = subject (raced (past the barn))) (fell)
e. (The horse)$_1$ = subject ((that)$_1$ (Bill) = subject (raced (WH_1) = object (past (the barn))) (fell)

NP-traces across clauses are limited to the overt subject position in the lower clause. Thus, we can have (14a) but not (14c).

This restriction does not hold for *wh*-trace, which can be either subject or object in its own clause. The grammar marks the difference between the traces as a function of how they are generated. *Wh*-trace is assumed to be in a special *wh*-node, known as *wh*-complement, of its own at the beginning of sentences and clauses. Whenever a clause-initial *wh* is encountered, it is attached to *wh*-complement and placed in the buffer. It then can join with the first available empty position in its own clause, or be passed down to the next clause and placed in the *wh*-node for that clause. In this sense, the parsing of *wh*-words is cyclic, moving from clause to clause, until a clause-internal position is available for it to fill.

In Parsifal, NP-trace is also placed in the buffer when the morphology indicates—for example, passive morphology ("was + pp") is a trigger to link a trace to the surface subject of the sentence and place the trace in the buffer. It is then available to be assigned to a constituent anywhere in the sentence, including its rightmost boundary. But it cannot literally move to a node of its own in a lower sentence. The result is that it can be assigned to fill the initial position of a lower sentence, since it is adjacent to it. But it can never be moved elsewhere in the lower sentence. This explains why NP-trace is always linked to the surface subject of a lower sentence, never any other surface position. The implication of this differentiation of kinds of traces in the parser is that *wh*-trace and NP-trace may be parsed by quite different mechanisms.

V. In certain cases, semantics determines constituency.

We come back to the problem of how semantic information might be embedded in the parser. To quote from Marcus (1980:52, including his numbering system):

> [Consider] the problem of prepositional phrase attachment. In sentence (.1) for example, the parser can be sure that "with" starts a prepositional phrase when this word enters the buffer, but it cannot possibly decide at that time whether the resulting PP should be attached to "the man" ... or to the clause itself, as in the most plausible reading of .2
>
> .1 I saw the man with the red hair
> .2 I saw the man with the telescope
>
> To solve this attachment problem deterministically requires access to some sort of semantic reasoning capability, but a necessary precursor to the application of this knowledge is the ability to first parse the prepositional phrase independent of its higher level role. This part of the problem can be solved simply by creating an unattached PP node when the preposition comes into the first buffer position, attaching its preposition and object, and dropping the resulting PP into the buffer where whatever [processes that] bring non-syntactic knowledge into play can examine the PP at their leisure.

The implication of this move is that adjunct phrases can be computed and held in abeyance until semantic information guides the particular attachment most consistent with their meaning.

VI. Cases in which the model fails correspond to human failures.

Even with its rich structure, the model makes systematic errors. For example, the priority of building verb phrases in simple declarative sentences can create misparsings that span such a distance that the model cannot recover from them. In (15a) the model gives priority to treating *has* as an auxiliary when it precedes a past participle, as opposed to treating it as the main verb and the past participle as an adjective. This priority leads to a garden-path misparse in sentences like (15b) that unambiguously require that *has* be a main verb.

(15) a. The store has assembled models.
b. The store has assembled models, as well as kits, in stock.

Marcus argued that such cases in fact correspond to cases in which humans also misparse the sentence, and have to access higher-level problem-solving strategies to recover from it. Thus, if the model is garden pathed just in those cases that humans are as well, this actually increases its plausibility as a psychologically correct model.

VII. There must be a means of recovery when the parser fails.

When the parser fails, it does not simply restart. Instead, repairs are carried out by "some higher level 'conscious' grammatical problem solving component" (Marcus 1980:204). In a garden-path sentence, the parser actually continues to the end of the sentence to produce two grammatical sentence fragments. The grammatical problem solver uses heuristics to operate on the grammatical fragments. For example, the parser initially produces for the garden-path sentence (16a)

(16) a. The cotton clothing is made of grows in Mississippi.

two grammatical sentence fragments, (16b) and (16c):

(16) b. [[The cotton clothing]$_{NP}$ [is made [of [???]$_{NP}$]$_{PP}$]$_{VP}$]$_{S}$
c. [[???]$_{NP}$ [grows in Mississippi]$_{VP}$]$_{S}$

The following heuristic applies:

> Given two consecutive sentence fragments, if the first fragment is a sentence that is complete through some part of the verb phrase and the second fragment is a clause that lacks a subject, then see if the bulk of the first fragment can be turned into a relative clause of some type (using another set of heuristics), thereby converting the entire fragment into a single NP which will become the subject of the second fragment. (Marcus 1980:205)

The use of this heuristic leads to the realization that *clothing is made of* is a relative clause; see (16d):

(16) d. [The cotton [clothing is made of t]$_{S}$]$_{NP}$

The heuristic therefore converts the two grammatical fragments, and the actual structure becomes available immediately; see (16e):

(16) e. [[The cotton [clothing is made of t]$_{S}$]$_{NP}$ [grows in Mississippi]$_{VP}$]$_{S}$

To summarize, in order to make parse assignment deterministic by application of grammatical rules, it is necessary to postulate that the parser can operate on several vertical nodes and can examine at least three horizontal constituents. These vertical and horizontal windows provide enough information in combination with a repertoire of syntactic possibilities to arrive at the uniquely correct syntactic assignment. In the course of making this model clear, Marcus defined a variety of dimensions along which all subsequent structural models can be examined. We now turn to a review of several models that have driven more recent and current research.

4.2.1 Minimal Attachment: Garden Paths and Construal

Frazier's Garden Path Theory (GPT) makes two basic assumptions about the process of combining words into phrases. First, processing resources are limited. This limitation means that the parser attempts to minimize complexity in order to preserve resources. Second, to generate candidate structures for a sentence, the parser uses only syntactic category information, such as whether a word is a noun, verb, or adjective. Thus, the GPT proposes that, as each word is received, the parser selects the structure that integrates the word into the previous structure with minimal structure change. Because of limitations on processing resources, only the first analysis that becomes available is maintained. The first available analysis is the "minimal attachment." Other analyses that are possible but not preferred are called "non-minimal attachment." In cases in which the first available analysis ultimately does not prove to be correct, there is a garden path, and the parser must perform relatively costly reanalysis. (See Frazier 1987a, 1987b for reviews.)

Basic evidence for initial computation of only the simpler, minimal attachment structure comes from the existence of garden-path sentences, such as (17):

(17) *Sentential complement*
John knew the answer was wrong.

In this sentence, readers frequently believe that the sentence is complete on reading *answer*. Reading *was*, however, makes this initial decision incorrect, forcing reanalysis. The GPT explains this phenomenon by noting that *knew* may have a noun phrase object (e.g., *the answer*), or it may have a sentential complement (e.g., *the answer was wrong*). Since the latter structure will require rules for expanding the verb phrase in the complement clause whereas the former does not, the parser adopts the simpler structure in which *the answer* is a noun phrase object.

Similar considerations apply in the more complex garden path in (18).

(18) *Reduced relative clause*
The horse raced past the barn fell.

This sentence is difficult because the active structure for *the horse raced past the barn* is initially adopted as soon as *raced* is received. At the point of receiving *raced*, the only rule needed is (19).

(19) VP $\rightarrow$ V (NP) (PP) (Adv)

When *fell* is processed, it becomes apparent that *raced past the barn* is actually a relative clause reduced from

(20) *Full relative clause*
The horse that was raced past the barn fell.

The reduced relative structure requires more elaborate phrase structure compared to the active structure, and so is not initially followed.

Much of the work on the GPT concerns its second assumption: *only* syntactic category information is used to develop the initial hypothesis about structure regardless of conceptual information. Rayner, Carlson, and Frazier (1983), for example, found evidence for a garden-path effect in sentences with reduced relative clauses similar to those in (21a) and (21b)

(21) *Reduced relative clause, plausible garden path*
a. The florist sent the bouquet of flowers was very flattered.

Reduced relative clause, implausible garden path
b. The performer sent the bouquet of flowers was very flattered.

compared to the control sentences:

Full relative clause
c. The performer who was sent the bouquet of flowers was very flattered.

Coordinate clause
d. The performer sent the bouquet of flowers and was very flattered as well.

Reading times were longer in the reduced relative sentences for the words *was very*, where it first becomes clear that the active structure is not viable. The fact that this increased reading time occurred equally for the reduced relatives with *florist* versus *performer* suggests that plausibility is not relevant for the initial syntactic decision. If plausibility were relevant, we might expect a smaller garden-path effect for the reduced relative with *performer*, since performers are more likely than florists to receive a bouquet of flowers.

Even though the GPT does not use plausibility information in the initial decision about structure, it may use this information to reject a structure that has been proposed on purely syntactic grounds. The effects of plausibility may appear rapidly, according to the GPT, but they do not appear initially.

The parser uses the *most recent filler strategy* to determine the structure of sentences with *wh*-movement. According to this strategy, a gap is coindexed with the most recent potential filler (Frazier, Clifton, and Randall 1983). The most recent filler strategy correctly predicts that (22a) is easier than (22b).

(22) *Late filler*
a. This is the girl$_1$ the teacher$_2$ wanted t_2 to talk to t_1

Early filler
b. This is the girl$_1$ the teacher wanted t_1 to talk.

Whenever the parser detects a gap, it searches backward from a gap to the first noun phrase, and assumes that this noun phrase fills the gap. The late filler sentence above is easier because the most recent filler strategy provides the correct structure. In making decisions about which noun phrases fill gaps, the parser uses only syntactic category information. Not only does it ignore plausibility constraints, it does not even use verb subcategorization constraints, even though they are arguably "syntactic." Frazier, Clifton, and Randall (1983) found that the late filler sentence (22c) is easier to understand than the corresponding early filler sentence (22d):

(22) *Late filler*

c. This is the girl$_1$ the teacher$_2$ decided t_2 to talk to t_1

Early filler

d. This is the girl$_1$ the teacher forced t_1 to talk.

They argue that decision times are longer for the second sentence because the parser initially posited the incorrect noun phrase (*teacher*) as the filler. If the parser used verb subcategorization information to assist in filling gaps, this mistake would not occur: one can "decide" to take action oneself, but one ordinarily "forces" someone else.

Frazier and Clifton (1996) have suggested that the parser makes initial decisions solely on the basis of syntactic information only for "primary phrases." Primary phrases include subject, predicate, sentential complements, and syntactic positions that occupy argument positions such as agent, goal, instrument, and theme. Initial syntactic attachment explains the processing differences in the following pairs of sentences, in which the minimal attachment sentence appears in (a) and the non-minimal attachment appears in (b):

(23) 1. *Main clause/relative clause*
   a. The horse raced past the barn and fell.
   b. The horse raced past the barn fell.
2. *Noun phrase as object/verb complement*
   a. John knew the answer very well.
   b. John knew the answer was wrong.
3. *Direct object of initial clause/subject of second clause*
   a. While Mary was mending the sock it fell off her lap.
   b. While Mary was mending the sock fell off her lap.
4. *Coordinate object/coordinate sentence*
   a. Jacob kissed Miriam and her sister.
   b. Jacob kissed Miriam and her sister laughed.
5. *Verb, direct object, indirect object/verb, object, modifier*
   a. Sandra [wrote [a letter] [to Mary]].
   b. Sandra wrote [[a letter] [to Mary]].
6. *Complement object/relative clause*
   a. John [told [the girl] [that Bill liked the story]].
   b. John [told [[the girl [that Bill liked]] [the story]]].

7. *Direct object|relative clause*
   a. Fred gave the man the dog.
   b. Fred gave the man the dog bit the package.
8. *Purpose|rationale*[1]
   a. Nixon bought [Trivial Pursuit] [to amuse us].
   b. Nixon bought [[Trivial Pursuit] [to amuse us]].

Nonprimary phrases are attached using the *construal principle*. These phrases are first "associated" with a thematic processing domain, then interpreted within that domain using both structural and nonstructural information. Nonprimary phrases are phrases that are elaborations of arguments, such as relative clauses, prepositional phrases, and adjunct phrases. Examples of sentences that use the construal principle are[2]

(24) 1. *Main clause|subordinate clause—prepositional phrase*
   I put the book that you were reading in the library (into my briefcase).
2. *High vs. low adverb attachment*
   We remembered that the assignment will be due yesterday/tomorrow.
3. *Relative clauses with a complex head*
   The reporter interviewed the daughter of the colonel that had had the accident.
4. *Secondary predication*
   John ate the broccoli yellow/raw/naked.
5. *High vs. low adjunct clause attachment*
   The doctor didn't leave because he was angry.

Figure 4.3 shows a flowchart of the GPT. Using words and syntactic rules, a constituent structure module determines the constituent structure. A thematic processing module operates in parallel to determine whether the constituent structure is consistent with the argument structure. The thematic processing module also indicates the most plausible host for associating adjunct phrases. The flowchart shows that plausibility enters into the processing system in two ways. One way is through the thematic processing module. This module uses all available information to check whether the constituent structure module has attached phrases in plausible argument positions. The second way plausibility information has an effect is through construal. The construal process establishes which of two or more ways of attaching an adjunct is most plausible.

To recapitulate, the recent versions of the GPT establishes some general concepts:

1. Following the simplest possible structure yields the correct interpretation in many cases, and in the long run may be cost-effective.
2. Purely structural theories cannot live without semantics. Semantics can have a role in initial hypotheses about sentence structure.

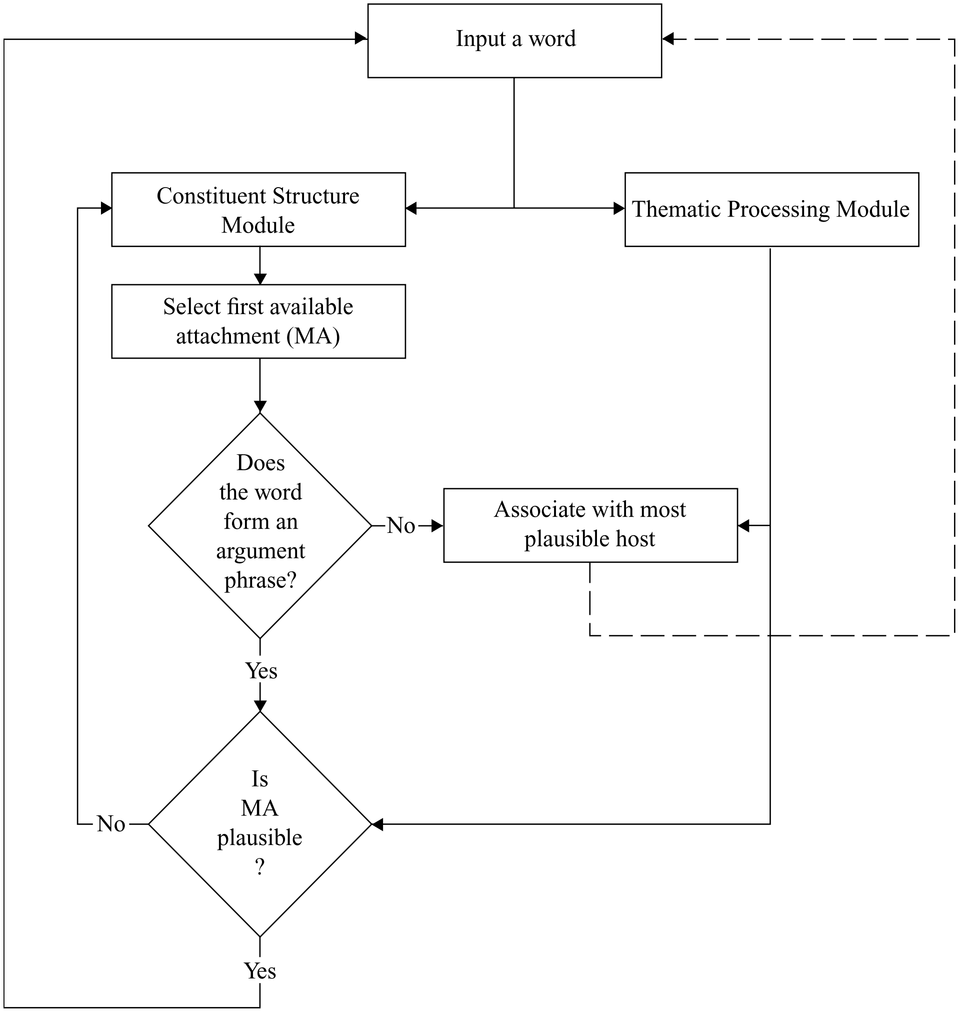

Figure 4.3
An information processing model of the Garden Path Theory.

4.2.2 Simplicity in Structure Building

Gorrell (1995) starts with two assumptions. First, the parser may work most efficiently by adopting the simplest allowable structure. The simplest structure generally provides the correct interpretation. It also costs less in terms of processing resources. Second, if the need arises, making a simpler structure more complex is easier than making a more complex structure simpler, if the simple structure is preserved as part of the complex one.

Gorrell expanded on these assumptions by proposing that the structural component of the parser can only build structure. It cannot destroy structure that has already been built. As each word is received, the parser builds the simplest possible dominance and precedence relations, a skeletal tree that indicates which syntactic category precedes another and which dominates another. The rationale for this is that the parser cannot change dominance and precedence relations.

Like Frazier's parser, Gorrell's builds a simple structure. Unlike Frazier's parser, however, Gorrell's uses plausibility information to guide initial decisions about dominance and precedence. This is allowed on the grounds that changing these relations is to be avoided. Since plausibility information can help the parser arrive at the correct structure, its use can reduce the need for subsequent reanalysis.

Figure 4.4 shows the architecture of Gorrell's parser. The structure builder utilizes words, including subcategorization information, semantic context, and principles of X-bar theory, to build a phrase structure tree. The structure interpreter uses this information together with principles such as those of government, binding, and case theory to interpret the tree for government, case, and so on. An example will clarify how these components work. Consider sentence (25):

(25) Bill knows Ian buys books.

When processing *knows*, the structure builder has access to the subcategorization information that *know* can have a NP-object or a sentential object. Since the NP-object is simpler, the parser builds a tree in which there is a VP dominating V and NP, with the V preceding the NP. This information is presented to the structure interpreter, which assigns a theta role and case to the NP that is forthcoming. When the structure builder receives *buys*, it builds a node for the complement phrase (CP) that is dominated by VP and preceded by V. Nodes are added to the tree structure so that the CP now dominates the NP for *Ian*. With the additional structure that arises from changing the NP *Ian* from an object of *knows* to subject of the embedded *buys*, there is no change in dominance or precedence relations. There has been a change in direct dominance, but the VP-node still dominates the NP for *Ian*. Consequently, there is no conscious garden path.

The situation is different when the parser must change precedence relations. When processing *raced* in a sentence like (26),

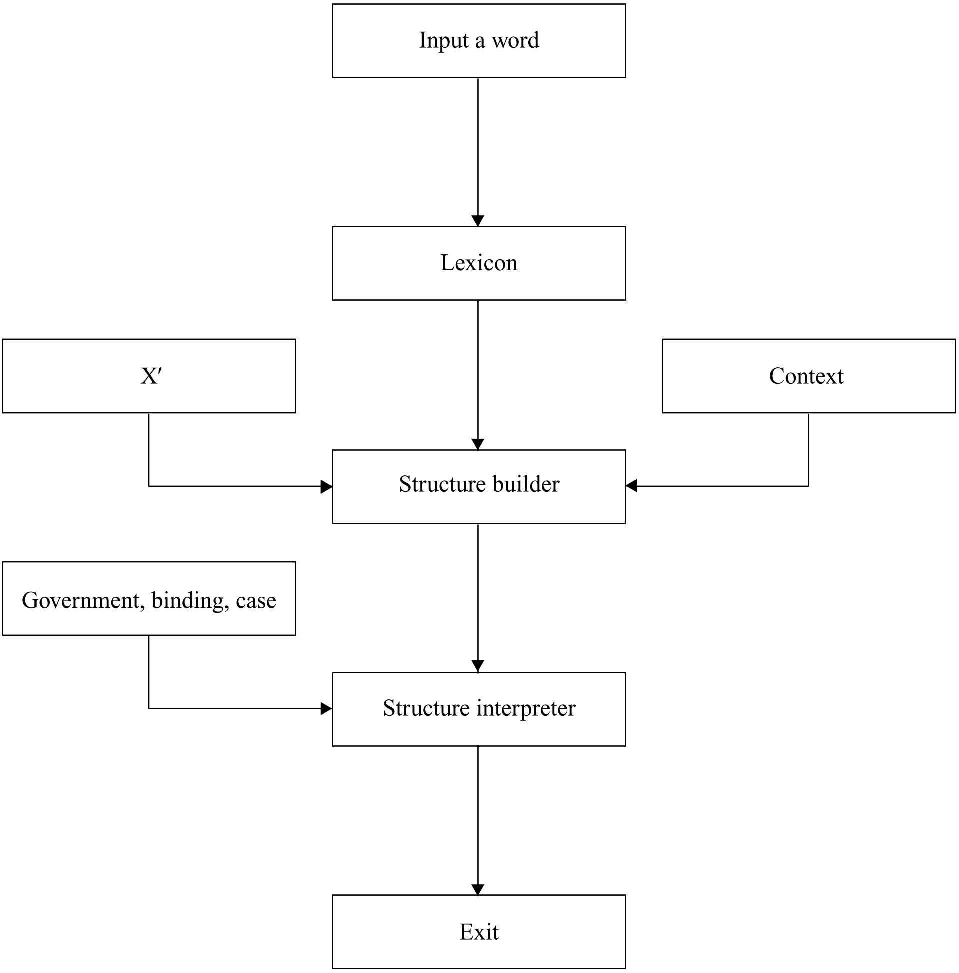

Figure 4.4
The architecture of Gorrell's parser.

(26) The horse raced past the barn fell.

the parser builds a VP-node that dominates a V-node for *raced*. In this structure, the NP precedes the VP. On processing *fell* the parser must retract this precedence relation. In the reduced relative structure, the NP-node for *the horse* dominates the node for the embedded relative clause that dominates the VP for *raced*. Since the structure builder cannot do this, conscious problem-solving mechanisms are employed, and there is a conscious garden path.

A garden path occurs as well in sentences like (27a).

(27) a. Ian told the man that he hired a story.

The parser initially constructs a complement phrase for *that he hired a*, as in (27b).

(27) b. Ian told the man that he hired a secretary.

In this initial structure, the NP-node for *the man* precedes the node for the complement phrase. Receiving *story*, however, causes reattachment of *that he hired* so that the complement phrase is not dominated by the NP-node for *the man*. Changing precedence relations is beyond the capability of the structure builder, and conscious reanalysis is required.

The principle of simplicity guides the processing of sentences with NP-movement. Examples are:

(28) a. This book_1 was read t_1 by Ian last night.
b. Ian wants PRO to leave to planet.

In the first sentence, passive morphology (*was ... by*) triggers postulation of a trace following the verb. In the second sentence, the infinitive *to leave* in triggers the postulation of PRO after the verb.

The parser handles sentences with *wh*-movement with the same principle of simplicity. In (29a) the *wh*-phrases is attached in a nonargument position.

(29) a. Which city did Ian visit?

Processing *visit* initiates construction of a NP-node, which is required for *visit*. Since the *wh*-phrase is not in an argument position, a trace is attached following *visit*.

Gorrell's parser handles more complex sentences such as (29b)

(29) b. What_1 did Ian say t_1 that Bill thought t_1 Fred would eat t_1.

in a similar fashion. Each verb leads to precomputation of an object NP, and the trace for *what* is attached. For example, *say* allows precomputation of a NP dominated by VP; if the sentence ends at *say*, the trace is attached to this NP, and the structure interpreter determines binding, case, government, and so on. When *that* is processed, the structure builder builds a complement phrase (CP); the attachment of the trace under NP remains but it is now dominated as well by CP. The trace at this point serves as an antecedent for additional traces after *thought* and *eat*. Thus, the structure analyzes the antecedent-trace relation as a series of local relations. Such sentences are not garden paths because there is no need to retract dominance or precedence relations.

Verbs with alternative subcategorization properties present special problems for processing *wh*-sentences. The following sentences from Fodor (1978) illustrate the problem:

(30) a. Which book$_1$ did Ian read t_1 to the children last night?
b. Which book$_1$ did Ian read to the children from t_1 last night?
c. Which child$_1$ did Ian walk t_1 to the office last night?
d. Which child$_1$ did Ian walk to the office with t_1 last night?

Both *read* and *walk* may be transitive or intransitive. However, *read* is more often used as transitive, as in the first sentence above, while *walk* in more often used as intransitive, as in the fourth sentence above. Intuitions suggest that sentence (30a) is indeed easier than sentence (30b), while (30d) is easier than (30a). Gorrell's parser predicts that sentence (30b) produces a garden path, if it is assumed that the parser has access to information about lexical preferences. The reason is that, in sentence (30b), the parser posits a trace after the V-node and preceding the PP-node. When *from* is received, the parser must change the precedence relations and posit the trace after the PP-node. Sentence (30c) is difficult because the parser does not posit a trace after the V-node, but at the end of the sentence it becomes clear that there must be a trace after the V-node so that the *wh*-phrase can occupy an argument position. According to Gorrell (1995:155), there must be "intervention of nonsyntactic factors to prevent the parser positing the primary relations [i.e., dominance and precedence] consistent with the structurally simpler reading." Gorrell refers to eye-movement results from Altmann, Garnham, and Dennis (1992) in processing sentences with reduced relative clauses such as (31).

(31) a. He told the woman he'd risked his life for ...

The simpler structure for *he'd risked his life for* is a complement clause, as in (31b), rather than a relative clause, as in (31c) and (31d).

(31) b. He told the woman he'd risked his life for her.
c. He told the woman he'd risked his life for that the fire was out.
d. He told the woman that he'd risked his life for that the fire was out.

A relative clause interpretation is appropriate when there are two women and the relative clause indicates which of the two women is being referred to. In the Altmann et al. study, sentences like these were embedded in a story biased toward either the complement interpretation or the relative clause interpretation. Altmann et al. found no increase in reading times for the disambiguating region in stories that supported the relative clause interpretation. Such a result indicates to Gorrell that the NP *the woman* is immediately related to the context. If there are two women mentioned in the context, the parser builds the more complex relative clause structure.

To sum up, some of the useful ideas in Gorrell's model are the following:

1. Dominance and precedence relations are harder to change than government and binding.
2. To avoid frequent and costly reanalysis, it is necessary to use conceptual information and lexical preferences.

4.2.3 Governing Categories

Theta roles are the primary semantic functions of phrases in sentences. Theta roles include functions like agent, theme, patient, goal, proposition, and so on. It is apparent that a primary goal of comprehension is to determine the theta roles as soon as possible. It also is plausible that comprehension will be difficult if the parser must change a commitment to a particular theta role that was established earlier.

This is the central concern of Pritchett's model of parsing. The parser follows what he calls *The Theta Attachment Principle*: fill theta roles as soon as possible and find a theta role for an unassigned phrase as soon as possible. Of course, the parser may make incorrect assignments on occasion; revisions in the assignments of words to roles are easy if the reassigned word remains in the same theta domain as before the revision. A theta domain consists of an element that assigns a theta role and all of the constituents to which it assigned the theta role. For example, a prepositional phrase, consisting of a preposition and a noun phrase, is a theta domain because the preposition assigns the role of object to the noun phrase. A verb phrase, consisting of a verb and its complements, is a theta domain as well, because the verb assigns theta roles of agent, theme, goal, experiencer, proposition, and so on. Different verbs have different requirements for theta roles, and some verbs are ambiguous, in having more than one set of requirements. (See Pritchett 1992 for a review.)

There are two kinds of revisions in parsing. Revisions that maintain theta domains are unconscious and can be performed without cost; revisions that change theta domains require conscious attention and high cost. Sentences (32a) and (32b) illustrate the two kinds of revisions and how they depend on the assignment of words to theta roles:

(32) a. Without her donations to the charity failed to appear.

In (32a) the first mistaken assignment concerns *her*. In its attempt to fill theta roles as soon as possible, the parser assigns *her* to the role of theme of *without*, just as it assigns *him* to the same role in

(32) b. Without him donations to the charity failed to appear.

However, when the parser receives *donations* there is no theta role for it to fulfill if *her* has the role of theme of *without*. To provide each word with a theta role as soon as possible, the parser revises its assignment so that *donations* becomes the theme of *without* and *her* modifies *donations*. Reassignment of *her* is not costly, however, since *her* still falls within the domain of *without*.

In (32b), *to the charity* is attached as a complement of *donations*, which retains its assignment as theme of *without* until *failed* appears. *Failed* requires the external theta role of subject, but the only way to fill this role is to change the assignment of *donations* from object of *without* to subject of *failed*. Since this change involves a change in governing category from *without* to *failed*, the revision is costly, and the comprehender is aware of having been led down a garden path. While (32a) illustrates both conscious and unconscious revisions, (32b) has only an unconscious revision.

(32) c. Without her donations to the charity Bob failed to appear.

Sentence (32c) has only the cost-free revision of reassigning *her* from object of *without* to modifier of *donations*. Once *donations* is assigned to object of *without* it does not need to be reanalyzed as subject of *failed*, since *Bob* is available to occupy that role.

Pritchett's parser successfully distinguishes several cases of conscious versus unconscious garden-path sentences—for example, (33a) and (33b).

(33) a. After Todd drank the water proved to be poisoned.
b. Susan knew her mother hated her.

Sentence (33a) is a conscious garden path because *water* changes from being assigned the role of theme by *drank* to being assigned the theme role by *proved*. Since *water* changes its theta domain, this is a conscious garden path. Sentence (33b) is not a conscious garden path. In this case, *mother* changes its role from theme of *knew* to experiencer of *hated*. However, since the clause *her mother hated her* remains as theme of *knew*, *mother* remains within the theta domain of *knew*.

The architecture of Pritchett's model is shown in figure 4.5. The central idea is that the parser attempts to maximally satisfy the theta criterion at all points during sentence processing. The strategy means that a word is assigned to any available theta role, and any available theta role is assigned a word as soon as possible.

If no word is available for a role, the parser examines the assigned tree to determine whether an assigned word may satisfy the unfilled role. If there is a word available for a role, a revision is made. If the revision does not involve a change in theta domain, the revision is cost-free and the next word is examined. If the revision

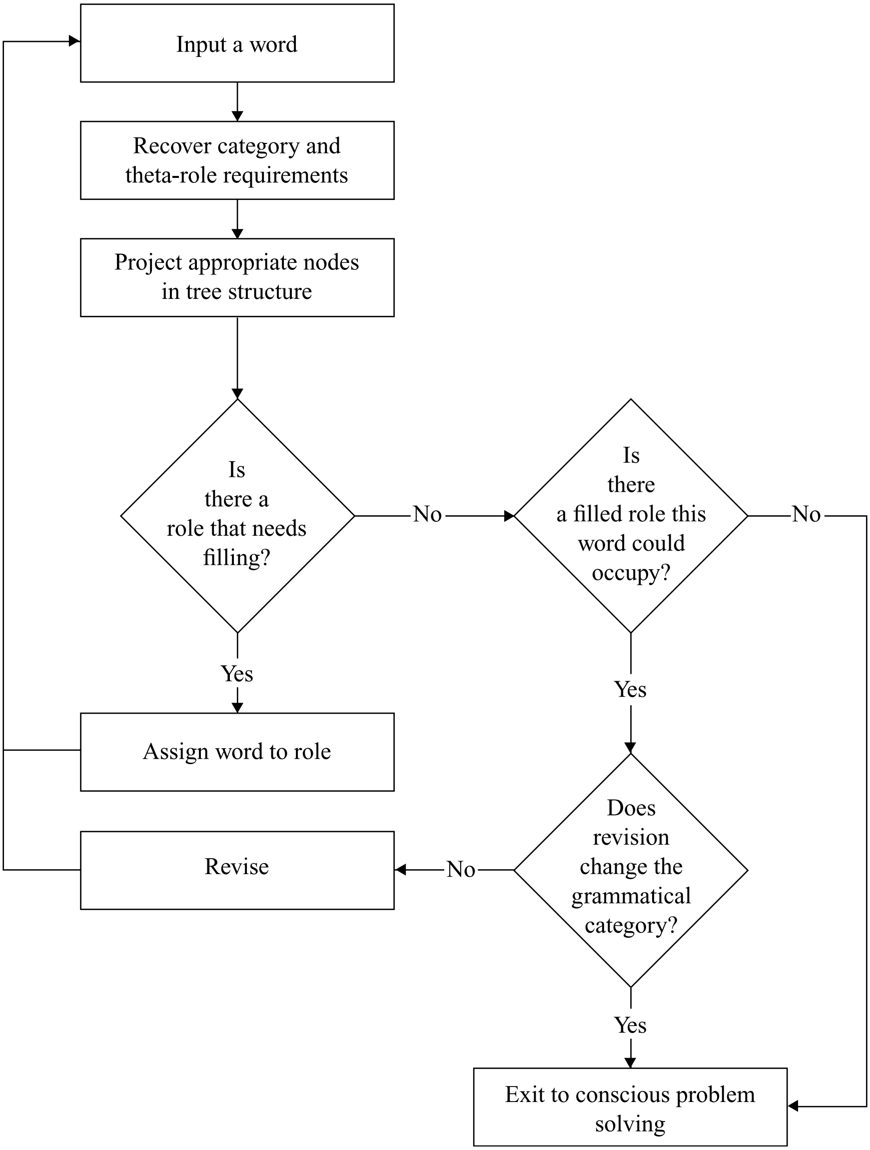

Figure 4.5
The architecture of Pritchett's parser.

does involve a change in theta domain, the revision the performed by conscious problem-solving mechanisms, which are conscious and costly.

The theta attachment principle states that the parser attempts to maximally fulfill theta-role requirements at all points. Consider the classic garden-path sentence (34):

(34) The horse raced past the barn fell.

When *raced* is processed, its maximal theta requirements are retrieved from the lexicon. Even though *raced* may simply take the theme role, it maximally requires an agent (as external subject) and a theme. The parser may adopt the active structure and assign *horse* as agent (external subject), leaving the theme role that is required for *raced* unfilled for the moment.

Alternatively, the parser may adopt the reduced relative structure by creating an empty category that has the theme role of *raced* in a relative clause, leaving the agent role that is required for *raced* unfilled. However, the parser rejects the reduced relative structure because this structure will leave *horse* without a theta role. Since the theta criterion can be satisfied more fully with the active structure, the parser assigns *horse* as the agent of *raced*. *Past the barn* is attached as the location role of *raced*. When *fell* is received, the parser looks for a NP to fill its theme role. The only possible NP to fill this role is *horse*, which already occupies the agent role for *raced*. There is substantial cost in making this revision, since *horse* must change to the new theta domain of *fell*.

Pritchett considers why some sentences that have the form of the classic garden-path sentence do not actually produce a noticeable garden path—for example, (35):

(35) The spaceship destroyed in the battle disintegrated.

The crucial difference between (35) and (34) is that *destroyed* is unambiguously transitive, while *raced* is optionally intransitive. The unambiguous cue to transitivity that *destroyed* provides prevents the violation of the theta reanalysis constraint (i.e., reanalysis that changes the theta domain is costly). When *destroyed* is processed, the active structure is adopted, the agent role is filled, and the theme role is left unfilled. When *in the battle* is received, the only way it can assign this phrase a theta role is if *destroyed* is posited as a reduced relative modifying *spaceship*. When *disintegrated* is received, *spaceship* fills its theme role. Consequently, there is no conscious garden path.

Why doesn't the reassignment of *spaceship* from subject of *destroyed* to unattached NP cost? Pritchett's answer is that it is the immediate detachment and reattachment of a phrase that is outside the capacity of the parser. When *in the battle* initiates reanalysis to the reduced relative structure, *spaceship* is detached from its role as subject of *destroyed*, but it is not reattached until *disintegrated* is received. Since the detachment and reattachment do not occur at the same time, this revision does not cost. Pritchett rejects the possibility that the parser bypasses the active analysis by filling obligatory internal arguments before optional external arguments. Thus, *spaceship* may be assigned as theme of *destroyed*, since *destroyed* requires a theme but not an agent. While this "no-misanalysis hypothesis" would fill an obligatory role early, it also would create a NP that needs a role (the head of the relative clause *spaceship*).

Some of the key ideas that emerge from Pritchett's parser are:

1. Sometimes reanalysis is costly and conscious; sometimes it is not.
2. Changing the assignment of a word to a new theta domain makes it costly and conscious.

4.2.4 Modules for Structure, Chains, Theta Roles

The structural models exhibit a range of views on how to deal with the fact that comprehension is fast, accurate, and largely effortless. One approach is to construct the parser so that it is serial and deterministic: only the correct structure is built. As we have seen, this approach requires lookahead and recoding of words into abstract categories, as in Marcus's Parsifal. A second approach is to relax the requirement of complete determinism without backtracking, but design the parser so that it can recover quickly when revision is occasionally necessary, as in Frazier's GPT (section 4.2.1). A third approach is to construct the parser so that only some aspects of structural analysis are deterministic and prohibited from participating in revision, as in Gorrell's model (section 4.2.2). Fourth, the parser may be designed so that there is no determinism: all aspects of structural analysis are subject to revision, but only certain operations of revision have a detectable cost, as in Pritchett's model (section 4.2.3).

Crocker (1996) extends this fourth approach. Crocker's parser mirrors the organization of government-and-binding theory. There are separate modules for building phrase structure, for assigning thematic roles, for constructing movement chains, and for evaluating the semantic/pragmatic function of a sentence. Each module carries out a specific function on a specific type of information. The processing system as a whole is incremental, so that as soon as a module performs its task, its results are passed on to the next module. If a later module detects an error in a previous module, the analysis is sent back for revision. Whether a revision is costly or not depends on the degree of commitment to the incorrect analysis. This in turn depends on how many modules have operated on and accepted the incorrect analysis.

Figure 4.6 shows how Crocker organizes the modules. Each syntactic processor applies the principles of corresponding components of government-and-binding theory. Based on lexical items, which include only the barest grammatical information, and X-bar theory, the phrase structure module determines sisterhood and constituency relations. The chain module establishes long distance dependencies that

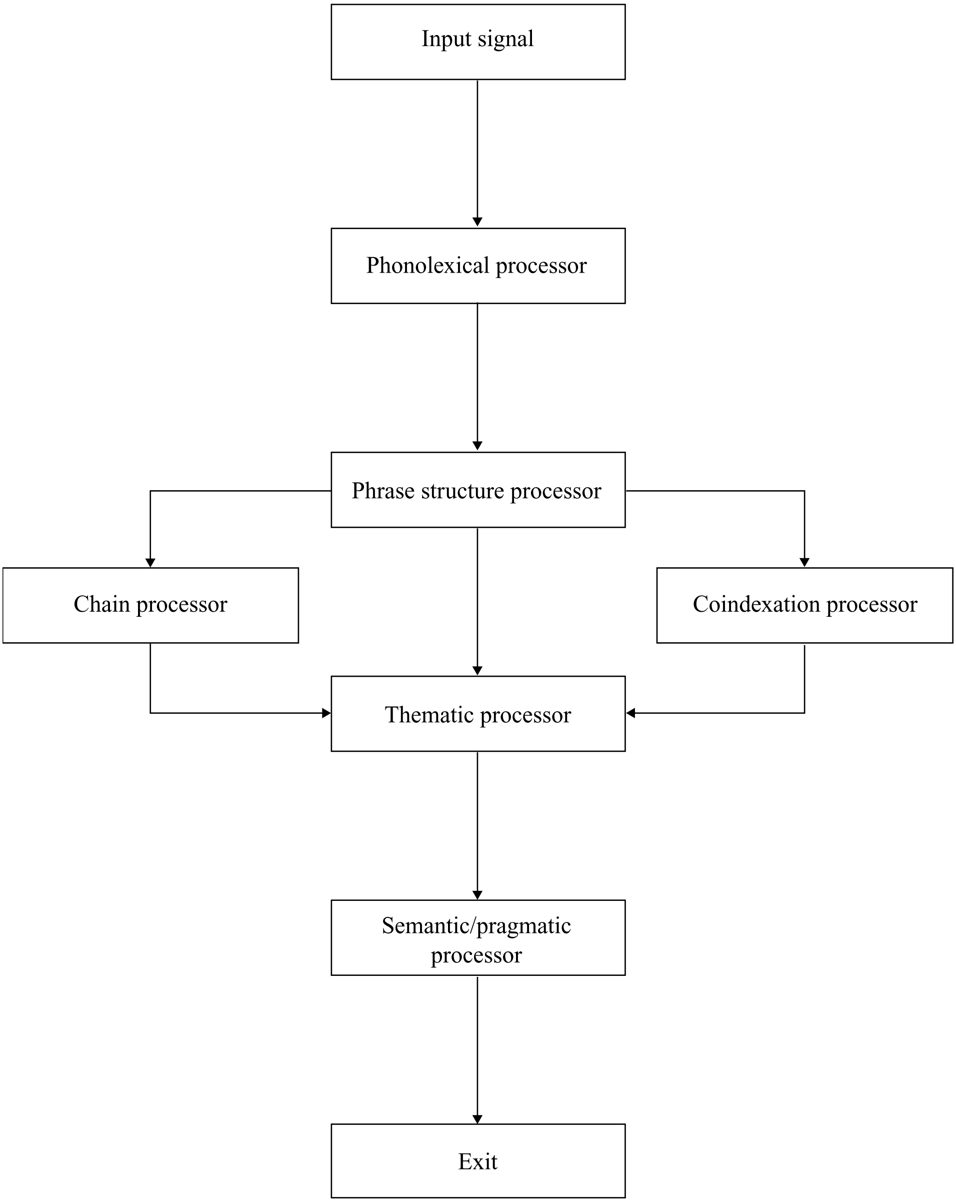

Figure 4.6
The organization of modules in Crocker's parser.

arise from movement, and the coindexation module uses binding theory to determine reference. To the information about phrase structure, chains, and coindexation, the thematic module adds subcategorization information and determines the argument structure. Based on the argument structure and knowledge about plausibility, the semantic/pragmatic processor provides a semantic representation for the general cognitive system. Thus, the various modules have access to only certain types of information.

The phrase structure module uses X-bar theory (see chapter 3) and basic lexical categories to project the phrase structure hierarchy dominating each word. Information about the basic lexical category of a word is simply N, V, and so on, and it does not include subcategorization information, such as that a particular verb is transitive or intransitive.[3] The phrase structure module uses strategies to resolve ambiguities.

A-attachment expresses the preference for attaching material into positions that potentially could receive a theta role (complements, external arguments, subject). A-attachment accounts for the preference to attach a prepositional phrase to a verb phrase to serve as an instrument, as in (36a)

(36) a. I saw the girl with binoculars.

rather than to a noun phrase to serve as a modifier, as in (36b)

b. I saw the girl with the flu.

A-attachment expresses a preference for attaching words into licensed positions rather than potential positions, accounting for the relative ease of the first sentence below:

(37) a. While Mary was mending the sock it fell off her lap.
b. While Mary was mending the sock fell off her lap.

And A-attachment accounts for the preference for active analysis rather than reduced relative analysis of (34), repeated here:

(34) The horse raced past the barn fell.

so that *the horse* is attached as subject (agent) of *raced* rather than receiving the patient role from *raced*.[4]

The phrase structure module also uses deep structure attachment: when it is necessary to attach to a position that does not take an argument, attach to a canonical base-generated position such as modifier, rather than to one that results from movement.[5]

The phrase structure module uses an active trace strategy: *posit a trace in any potentially vacated position*. Thus, the phrase structure module initially posits a trace at each t_1 in (38)

(38) Who_1 t_1 did you want t_1 Mother to make t_1 a cake for t_1?

since all of the following are grammatical:

(39) a. Who_1 t_1 left?
b. Who_1 did you kick t_1?
c. Who_1 did you want Mother to kick t_1?

Crocker's phrase structure module posits these traces without regard to subcategorization information, which is only available to the thematic module. According to Crocker, sentence (38) is not very difficult because the thematic module quickly rejects the chain and sends the sentence back to the phrase structure module before there is a strong commitment to it.[6]

As the phrase structure is determined, it is simultaneously passed on to the thematic module, the chain module, and the coindexation module. The thematic module uses information about the phrase structure positions of words and knowledge about theta-role requirements to assign theta roles to words. Importantly, this module does not have access to semantic features of words. Once the theta roles are determined, they are sent to the semantic/pragmatic processor. The semantic/pragmatic processor uses semantic features and the theta-role assignments to determine whether the thematic representation is plausible.

There is incremental processing. This means that as soon as a processor computes a representation, it is sent on to the next processor. The degree of difficulty in reanalysis is related to the level of commitment to an analysis—that is, to the number of processors that have accepted the analysis as well formed.

The following sentences illustrate degree of difficulty in reanalysis:

(40) a. After the child sneezed the doctor prescribed a course of injections.
b. I broke the window with my sister.
c. While Mary was knitting the sock fell off her lap.

In (40a), the phrase structure module attaches *doctor* as a complement of *sneeze*, but this is immediately rejected when the thematic processor tries to assign *doctor* as patient of *sneeze*. Thus, the phrase structure module needs to reanalyze. This reanalysis is neither costly nor conscious.

In (40b), the phrase structure module attaches the PP as an instrument, according to the strategy of argument attachment. Whenever possible, the phrase structure processor attaches words into potential argument positions. The thematic module assigns *sister* the role of instrument. However, the semantic/pragmatic system rejects *sister* as a plausible instrument, forcing reanalysis by both the phrase structure module and the thematic module. Reanalysis in this case is costly but unconscious because it occurs within the syntactic system.

In (40c), the phrase structure module attaches *sock* as complement of *knitting*. The thematic module assigns *sock* the role of patient. The semantic/pragmatic processor accepts these assignments as plausible. *Fell*, however, forces the phrase structure module to reanalyze *sock* as subject of *fell*. This requires the thematic module and the semantic/pragmatic processor to reanalyze as well. Since the reanalysis requires the work of three processors, it is costly and conscious.

To summarize, the key points in Crocker's parser are the following:

1. Some revisions can be performed quite easily, and others cannot.
2. Degree of commitment is related to difficulty of revision.

4.2.5 A Cost Metric for Online Computation

Some sentences are difficult not because they induce a garden path, but because their requirements for integrating words into a syntactic structure strain the limits of processing resources. This is the motivation for Gibson's Syntactic Prediction Locality Theory. The theory is unique among the contemporary structural theories we review here in that it makes clear predictions about the relative difficulty in comprehending sentences that are not garden-path sentences. (See Gibson 1998 for a review.)

Gibson developed a metric for predicting the processing difficulty of sentences. The central idea is that comprehension involves assigning words to thematic roles. Each word is stored in short-term memory until it is integrated into a syntactic structure that provides it with a thematic role. Similarly, each thematic role that is required for a predicate is stored in short-term memory until a word is found that can fulfill the role. Thus, there are two factors that increase processing difficulty. One is the number of thematic roles that have not been assigned a word. The second is the number of words that have not been assigned to a thematic role.

Gibson illustrates the cost metric with center-embedded sentences like (41a) and (41b):

(41) a. *Object relative clause*
The reporter who the senator attacked admitted the error.
b. *Subject relative clause*
The reporter who attacked the senator admitted the error.

In the object relative clause, *who* has been extracted from its position as the deep structure object of the relative clause verb *attacked*. In the subject relative clause *who* has been extracted from its position as the deep subject of *attacked*.

A sentence with an object relative clause typically is harder to understand than one with a subject relative clause (see chapter 7). Gibson's explanation of this difference is that there are increased processing costs momentarily associated with unfilled syntactic commitments in a sentence with an object relative clause.

Table 4.1 shows the word-by-word processing requirements for integrating words into a syntactic structure for sentences with different types of relative clauses. The table shows how each word is attached to previous words, and what new discourse referents have been introduced since the parser received the attachment site. The integration cost $I(n)$ depends on the number of new discourse referents n introduced since the attachment site. Comparing the two types of relative clause constructions in table 4.1, we see that there is substantial cost, $I(3)$, associated with the main verb *admitted* in both constructions. However, the object relative requires that the embedded verb *attacked* be integrated in two ways while maintaining three discourse referents in memory, for an integration cost of $I(1) + I(2)$. The processing load requirement for a subject relative clause on the last word of the relative clause (*senator*) is lower, with an integration cost of $I(0) + I(1)$.

Gibson's version of simplicity is motivated by the parser's need to reduce the demands on processing resources. It follows that whenever there is a choice between alternative structures, the parser will select the one that makes fewer commitments. In cases in which alternative structures require similar commitments, both structures are retained until a definitive choice can be made. Gibson's model is an example of "ranked parallelism," in which the parser maintains alternative structures simultaneously but ranks them in terms of preference. The structure that requires the least processing resources is ranked first.

Gibson allows the parser to use plausibility and frequency information to assist in breaking a tie between alternative structures. The justification for this feature of the parser comes from evidence that information about plausibility can eliminate the garden-path effect in reduced relative clauses. For example, Trueswell, Tanenhaus, and Garnsey (1994) found that word-by-word reading times were longer during *by the lawyer* in reduced relative clauses than in unreduced relative clauses (see chapter 6):

(42) a. *Reduced relative clause*
The defendant examined by the lawyer turned out to be unreliable.
b. *Unreduced relative clause*
The defendant that was examined by the lawyer turned out to be unreliable.

There was no difference in reading times for the corresponding phrase in the sentences like the following:

(42) c. *Reduced relative clause*
The evidence examined by the lawyer turned out to be unreliable.
d. *Unreduced relative clause*
The evidence that was examined by the lawyer turned out to be unreliable.

Table 4.1
Integration costs in object relative and subject relative sentences

Word	How it is integrated	New referents	Cost
A. Object Relative			
The	Not integrated	None	None
reporter	Attach to *the*	None	I(0)
who	Attach to *reporter*	None	I(0)
the	Attach to *who*	None	I(0)
senator	Attach to *the*	None	I(0)
attacked	1. Attach to *senator* by assigning *attack*'s agent role to *senator*	*attacked*	I(1)
	2. Attach empty category as object and coindex with *who*	*senator* *attacked*	I(2)
admitted	Attach as main verb to *reporter*	*reporter* *attacked* *admitted*	I(3)
the	Attach as start of NP	None	I(0)
error	1. Attach to *the*	None	I(0)
	2. Attach *the error* as object of *admitted*	*error*	I(1)
B. Subject Relative			
The	Not integrated	None	None
reporter	Attach to *the*	None	I(0)
who	Attach to *reporter*	None	I(0)
attacked	1. Attach a gap in subject position	None	I(0)
	2. Attach *attacked* to its subject	*attacked*	I(1)
the	Attach as start of NP	None	I(0)
senator	1. Attach to *the*	None	I(0)
	2. Attach *the senator* as object of *attacked*	*senator*	I(1)
admitted	Attach as main verb to *reporter*	*attacked* *senator* *admitted*	I(3)
the	Attach as start of NP	None	I(0)
error	1. Attach to *the*	None	I(1)
	2. Attach *the error* as object of *admitted*	*error*	

Evidence apparently is a poor agent for *examined*, making the reduced relative interpretation more likely than in the case of the sentence with *defendant*.

According to Gibson's theory, plausibility has this effect on processing reduced relatives only because there is a small difference in processing costs between the main verb and reduced relative interpretations of *examined*. When reading *examined* in *the defendant examined* . . . , the main-verb interpretation leads to a commitment for a forthcoming noun phrase for the direct object. The reduced relative interpretation leads to the prediction of a main verb (which never has a cost in Gibson's model) and an optional adverbial modifier, as in (43).

(43) The defendant examined yesterday was dismissed.

In both the main-verb and reduced relative interpretations, there is one syntactic commitment and no unassigned referents at the point of *examined*. Since the processing load for the two structures is similar, semantic information can shift the activation levels of the structures. For example, inanimate initial nouns lead to the reduced relative interpretation in *the evidence examined* . . . and animate nouns lead to the main-verb interpretation in *the defendant examined*. . . . Thus, the *by*-phrase produces an increase in reading time only for *the defendant examined by the lawyer* . . . , where the main-verb interpretation is shown to be incorrect.

To recapitulate, the grains of truth in Gibson's model are the following:

1. More than one interpretation can be available at any one time.
2. Frequency and conceptual information can influence the preferred interpretation.
3. Processing load increases with the number of syntactic commitments and unattached words.

4.2.6 Comparison of Structural Theories

We began this section with several ideas due to Marcus (1980). We then sketched some of the contemporary structural theories descended from Marcus, which introduced variations on different aspects of his work. Table 4.2 summarizes variations in Marcus in the contemporary theories. We can differentiate the theories with a set of questions:

What kinds of linguistic structures are assigned?

Like Marcus 1980, Gorrell's model and Pritchett's model assign a surface structure that is annotated to indicate coreference and the location of traces. Some models develop multiple representations of linguistic structure. Frazier's model develops separate representations of phrase structure and argument structure. Gibson's model develops separate representations of phrase structure, argument structure, and discourse structure. Crocker's model develops independent representations of phrase structure, movement chains, coreference, argument structure, and pragmatic meaning.

Table 4.2
Comparison of structural theories

Issue	Marcus	Frazier	Gorrell	Pritchett	Crocker	Gibson
What structures are assigned?	Annotated surface structure	Phrase structure, argument structure	Annotated surface structure	Annotated surface structure	Phrase, chain, coreference, argument	Phrase, argument, discourse
Are candidate structures formed serially?	Yes	Yes	Yes	Yes	Yes	No
Is parsing deterministic?	Yes	No	"Yes"	No	No	Yes
What operations have a processing cost?	None	Building branches	Changingdominance or precedence relations	Changing theta domain	Changing a strong commitment	Retaining unfilled roles and unassigned words
How does reanalysis occur?	General problem solving	Visibility	Adding nodes; general problem solving	Trimming	Recycling	Reinstating an old tree
Is nonstructural information used?	Yes	"No"	Yes	No	No	Yes
How is NP movement determined?	?	Minimal attachment	Cues to traces	Theta attachment	A-attachment	Theta attachment
How is *wh*-movement determined?	?	Most Recent Filler Strategy	Subcategories; frequency	Theta attachment	Active Filler Strategy	Active Filler Strategy

Are candidate structures formed in parallel or serially?

Are two or more candidate syntactic structures generated and one selected at some point, or is the model serial, with only one structure generated? Frazier, Gorrell, Pritchett, and Crocker all follow Marcus in proposing that the parser forms syntactic representations serially. In each of these models, an initial syntactic representation is formed and modified until it "crashes." Gibson's model is the one exception to serial formation of candidate structures. In this model, all possible structures are initially computed and the computationally simpler take priority. Structures that are similar in costs are maintained until there is information that distinguishes between them.

Is the parser deterministic?

Marcus (1980) defined strict determinism to mean that once a syntactic decision is made it cannot be revoked—that is, there are no revisions. By allowing a look-ahead buffer, Marcus was able to demonstrate determinism for sentences that cause no processing difficulty. Contemporary models of human sentence processing have dropped Marcus's notion of a look-ahead buffer, and, as a result, they have changed his idea of determinism.

On the issue of determinism, Gorrell and Gibson are perhaps closest to Marcus. Gorrell proposed that decisions about dominance and precedence relations are deterministic. Other syntactic decisions are not deterministic, and the parser easily modifies these decisions. Gibson proposed that several structures are computed and maintained simultaneously. Structures may be abandoned if they are costly, or reintroduced, but they are never modified. Frazier relinquished the idea of determinism and proposed that incorrect syntactic decisions are fairly common. To reduce processing load, the parser adopts the "minimal attachment." The cost of using this strategy is that the parser must frequently revise. Both Pritchett and Crocker abandoned determinism. Their parsers sometimes carry out cost-free revisions when no change in theta domain is required (Pritchett), or when there is no strong commitment (Crocker). Otherwise, the parser carries out costly revision.

What operations have a cost in processing resources?

Structural models differ in whether syntactic operations have a measurable cost in processing resources. Operations that require conscious attention are costly; those that occur "automatically" are cost-free. For Marcus (1980) the parser's syntactic operations are cost-free; however, a grammatical problem solver operates at substantial cost when the parser fails.

Contemporary structural models have introduced variations on this theme. Some models retain Marcus's idea that syntactic decisions are essentially cost-free but elaborate on the idea that revisions are costly. For example, Gorrell maintains that the parser builds structure at no cost; revisions of dominance and precedence relations can be conducted only by a grammatical problem solver and are costly. Pritchett and Crocker also suggest that basic structure building is cost-free, but in their case, the parser can carry out revisions that are costly if they involve changing a theta domain (Pritchett) or a decision with strong commitment (Crocker).

In other models, syntactic operations are costly. For Frazier, it is the cost of building structure and the need to reduce processing costs that leads to the minimal attachment strategy: rules that create more branching in the phrase structure tree are more expensive and therefore are avoided. Gibson suggests that maintaining syntactic commitments and unassigned words are both costly. A structure that is more costly than another is abandoned.

How does reanalysis occur?

Information that resolves a temporary ambiguity often occurs shortly after the ambiguity. Thus, Marcus's model was able to parse deterministically by using a three-item look-ahead buffer. For Marcus, a conscious grammatical problem solver carries out reanalysis. Contemporary models have distinguished between revisions that are easy and hardly noticeable, and those that are difficult and require conscious effort. There is general agreement on the basic facts to the explained, but there are differences in the explanation of the facts. The facts are that the first sentence in each pair of (44) is easier than the second.

(44) *Fodor and Inoue* (1998)

a. The boy noticed *the dog* limped badly.
b. While the boy scratched *the dog* yawned loudly.

Pritchett (1992)

c. Susan knew *her mother* hated her.
d. After Todd drank *the water* proved to be poisoned.

Sturt and Crocker (1998)

e. The wedding guests saw *the cake* was still being decorated.
f. While the wedding guests ate *the cake* was still being decorated.

Sturt and Crocker (1998)

g. Once the students had understood *the homework* was easy they quickly finished it.
h. Once the students had understood *the homework* was easy and they quickly finished it.

Lewis (1998)

i. Mary forgot *her husband* needed a ride yesterday.
j. Although Mary forgot *her husband* didn't seem very upset yesterday.

In each case, our intuitions tell us that we interpret the italicized material initially as being the object of the preceding verb, but the word following the underlined material shows that this interpretation is incorrect. The garden path is hardly noticeable in the first sentence of each pair, but it is quite striking in the second sentence. These intuitions suggest that it is easier to change the assignment from direct object to subject of a complement clause than to subject of a main clause. (Note, however, that the last pair from Sturt and Crocker 1998:395 is the only pair that controls for clause structure and verb subcategorization.) Here we consider how contemporary models explain these basic facts about ease of revision.

In Frazier's model, the parser carries out revisions when attached material conflicts with the grammar; revisions are focused at the location of the grammatical conflict (see also Fodor and Inoue 1998; Lewis 1998). The Visibility Hypothesis states that reanalysis that involves attachment to a recently postulated and uninterpreted node is relatively easy (Frazier and Clifton 1998). The Visibility Hypothesis can explain the facts, because easy reanalysis involves attachment to the recently postulated Verb node rather than to the earlier postulated Sentence node.

In Gorrell's model, reassigning the noun phrase to subject of a complement clause preserves dominance (the VP dominates the NP in both cases, although not directly for the complement structure). When the noun phrase is changed to subject of a verb in an independent clause it is no longer dominated by the VP. Since dominance must change in the second case, the revision is costly (see also Sturt and Crocker 1998). The parser can perform reanalysis that involves adding nodes, but only a general problem solver can perform reanalysis that involves destroying structure. Gorrell (1998) adds that *the horse raced past the barn fell* is difficult because there is no attachment site available for *fell* at the rightmost edge of the developing phrase structure tree once a main clause structure has been established for *the horse raced past the barn.*

In Pritchett's model, the parser carries out reanalysis by trimming the phrase structure tree. The only revision that is costly is one that requires changing the theta domain of the word. In *Susan knew <u>her mother</u> hated her*, the parser must change *her mother* from theme of *knew* to experiencer of *hated*, but in each case, *her mother* is within the theta domain of *knew*. In *After Todd drank <u>the water</u> proved to be poisonous*, the parser must change *the water* from theme of *drank* to theme of *proved*, which involves a change in theta domain. As a result, the second sentence is harder.

In Crocker's model, reanalysis occurs when a module detects an incongruity. The degree of difficulty in changing an analysis increases as the number of modules that have "accepted" the analysis increases. Crocker's model does not distinguish the sentences within pairs above. In each case, the assignment of the underlined material as direct object passes through the thematic module and the semantic/pragmatic module, and in each case, the next word signals a need for reanalysis by the phrase structure module. Reanalysis occurs by recycling through the modules.

Gibson's model makes no clear prediction about the basic facts of revision. If two alternative structures differ greatly in processing load, the more costly structure is no longer pursued. "Reanalysis" consists of reintroducing a previously abandoned structure. In the easier cases the parser must reintroduce the complement structure, and in the harder cases it must reintroduce the intransitive. If we assume that the parser drops the intransitive structure when it processes the underlined material, Gibson's model can explain the facts. However, Gibson's model provides no reason why the intransitive should be dropped while the complement clause interpretation is retained.

What is the role of nonstructural information such as conceptual representations and probabilistic information?

The structural models emphasize structural processing, and they tend to reduce the role of meaning and probabilistic information. Marcus, for example, allowed for meaning to influence the attachment of prepositional phrases, but did not specify how the relevant meaning was obtained.

Frazier departed from Marcus in stating that all initial structural decisions are made solely on the basis of syntactic category information. Plausibility and frequency of use may influence syntactic decisions in reanalysis, but these types of information do not initially influence syntactic decisions. Frazier and Clifton (1996) modified this view by acknowledging that the initial attachment decisions about nonprimary phrases involve meaning.

For Gorrell meaning has a prominent role in ensuring that dominance and precedence relations are established deterministically. The parser can use conceptual information to guide structure building. Information about frequency of use can be incorporated into the parser, but frequency of use of syntactic patterns cannot.

According to Pritchett and Crocker, the parser does not use information about meaning or lexical preferences to build structure.

For Gibson, neither frequency nor conceptual information is ordinarily accessed. These kinds of information are used only to resolve local ambiguities for which each of the readings is roughly the same complexity as the others.

How does the parser handle NP- and wh-*movement?*

While most theories provide similar treatment of *wh*-movement, the notion of NP-movement is characteristic of derivational grammatical theories. In the Marcus model, cues such as passive morphology, verbs with special properties, and *wh*-words cause the parser to attach a trace to a node and drop the node into the buffer. But *wh*-trace differs from NP-trace in that it has a special *wh*-node that moves to the first available position. NP-trace is coindexed with its referent clause by clause. Gorrell treats NP-movement such as passive and subject raising and *wh*-movement in a fashion similar to Marcus. In Frazier's model, NP-movement is handled by minimal attachment: the parser always chooses the simplest structure (at least for primary phrases). *Wh*-movement is processed with the most recent filler strategy. Pritchett, Crocker, and Gibson treat NP-movement in similar ways. Pritchett's theta-attachment principle means that the parser adopts the structure that assigns the most roles and leaves the fewest roles unassigned. Crocker considers NP-movement indirectly through the preferences that are exhibited in A-attachment and deep structure attachment. Crocker and Gibson treat *wh*-movement using a version of the active filler strategy.

4.3 Statistical Models

Associative models that extract regularities from a complex environment are by definition learning models. Our discussion of statistical models of comprehension necessarily blends theories of inductive learning and resulting representations. Statistical models of induction face two computational questions:

- Is the information available in the input for statistically based inferences about structure? Superficially, language structures appear to be wildly varied, making it difficult to conceive of how any inductive procedure with a small amount of data could extract structurally relevant information.
- What is the inductive learning mechanism that can extract the relevant information as it goes along—even if the information is technically "in" the language, what kind of mechanism can pick it up during some kind of actual behavior?

The mathematical tools of multiple regression, multidimensional scaling, and hierarchical analysis are powerful ways to examine if there are statistically distinct factors isolable from a heterogeneous database. For example, lexical categories can be strongly differentiated and categorized just from a distributional hierarchical analysis of the contexts in which items can occur. This seems mysterious, but it is not. Words from the same lexical category tend to occur in similar environments. For example, in normal sentences a word following *the* and preceding *has* is a noun, and a word between *shall* and *a* is a verb. Any scheme that analyzes the similarities of words in terms of their immediate environments will tend to arrive at syntactic-like categories. For example, Mintz (1997) applied a cluster analysis to the words in their immediate environments, in a large sample of "motherese"—utterances by caregivers to children. He classified each word in terms of the word preceding and following it. Then, he used a clustering analysis that grouped words together based on the similarity of their environments. Words that are very similar in their environments are grouped as very low level clusters; words that differ in their environments are linked only at a very high level in the hierarchy. A hierarchical clustering model applies stepwise criteria to establish an actual single hierarchy that best fits the data. The analysis yielded a hierarchical structure in which "nouns" are well differentiated from "verbs."

A simple analysis of this kind works because of the frequency of structurally defined frames, especially in simple sentences drawing on a restricted vocabulary. At first, it might seem that this "solves" the problem of category learning with no access to meaning or prior knowledge of syntactic categories. In that view, what a language-learning child does is apply the psychological equivalent of a cluster analysis to discover the syntactic categories, and classify words into them. But how to apply a tool-like cluster analysis is not obvious. There has to be a method that intrinsically flows from ongoing behavior. This need is reflected in several important limitations, which are common to all mere demonstrations that structurally relevant information is statistically available in the input.

1. *The Grain Problem*. How do we know what size grouping to pick? Indeed, if we pick just the right level, we obtain a good separation of nouns and verbs, but this depends on the judgment of the statistician, who already knows what the goal is. Mintz suggests that there may be an optimal group size that can be maintained by a real child, which automatically determines the right level in the hierarchy. But this choice is delicately sensitive to the particular input data, and we think there probably is no single criterion that works in general.

Mitchell et al. (1995) discuss the grain problem as it applies to higher-order analyses. To use prior experiences with language to facilitate sentence comprehension, the reader or listener must record and store relevant features of those experiences and must match those features with current linguistic material during comprehension. The *grain problem* refers to the question of what level of structural features is the appropriate level to store. Consider, for example, the following sentence:

(45) Someone praised the wife of the football star who was outside the house.

In deciding whether to attach the relative clause *who was outside the house* to *the wife* or to *of the football star*, the parser might refer to information at a variety of grain sizes. From relatively coarse grain to relatively fine grain, some possibilities are:

- Decisions may be based on the relative frequency of attachment locations in sequences of the form NP-(modifying constituent)-(modifying constituent).
- There may be separate records depending on the type of initial modifier, so that the records are based on frequency of attachment locations in NP-PP-(modifying constituent) sequences.
- There may be separate records for particular types of modifying constituents that follow the PP, so that in this case the relevant records contain information about the frequency of attachment locations in NP-PP-RC ("relative clause") sequences.
- There may be separate records for experiences with NP-PP-RC sequences depending on the preposition that introduces the PP, with different records for *of*, *with*, *by*, and so on.

- There may also be different records depending on whether the structure appears early versus late in the sentence.
- There may be different records depending on the animacy of the potential noun hosts.
- There may be different records depending on the specific nouns that are potential hosts.

2. *The Category Problem.* Even if the words are strongly segregated into groups, how does the mechanism know which groups are linguistically significant and which are "noise"? This problem has two aspects.

First, the grouping system has to know what a "noun" versus a "verb" is, in order to use the grouping to establish what words fall in each. But "noun" versus "verb" is in part a linguistic differentiation that depends on their function in sentences: "verbs" require argument positions to be filled; "nouns" fill those positions. Grouping words together based on immediate contexts does not provide that definition. Rather, the category learner either has to know enough to look for specific kinds of distributional evidence about the word groups, or has to access some kind of automatic mechanism that can extract that differentiation from the input strings. It is hard to see how an automatic mechanism could word without prior knowledge of enough syntactic structure to note the differences. And, of course, postulating that the learner knows what kind of categorical information to look for is simply an example of structural "nativism."

Second, the system has to know how many other categories it is looking for. And, these too, are defined in terms of their syntactic distribution.

3. *The Dialectical Mush Problem.* Consider a world in which the word groups are learned with the hierarchical cluster analysis. Each analysis has some error in it—that is, certain actual nouns may be classified as verbs and conversely, leading to a certain degree of error. Suppose that the learner then has children: the input to the next generation will be degraded, leading to an increased amount of error. The errors will increase with each generation. To put it differently, the system works as well as it does, just because the regularity of syntactic structure guarantees that words of the same category will sometimes appear in the same environment. Without accessing an actual syntax, in which lexical categories are defined in terms of their structural distribution, the system will endure entropy with each successive learner, and lose all structure after some number of generations.

Analytic procedures like hierarchical grouping algorithms are sensitive to lexically coded regularities, but do not correspond in an obvious way to associative relations between mental entities. The cluster analyses suggest that the ecology of language provides a lot of statistically reliable information relevant to linguistic structure, but little about how to discover it. Connectionist modeling offers a mathematical technique related to multiple regression analyses, but with a transparent utilization of associative relations, and a traditional learning rule that changes strength of association. Such models can be organized to become sensitive to aspects of syntactic structure that transcend the individual lexical item category.

4.3.1 Mediation Association Theory and Connectionism

In brief, mere existence of a statistically valid structure in the language environment will not guarantee discovery or application of the structure. As we reviewed in chapter 2, Osgood postulated a learning model with an intermediate level of structure, with r-s modules connected to numerous inputs and overt outputs. But Osgood did not offer a completely explicit mathematical formula as to how the variation of connection strengths between the mediating modules and the explicit responses percolates back to vary the strength of the connections from the initial stimuli to the mediating modules. In spirit, he was drawing on Hull's (1943) fractionation of responses and goal reinforcements to allow for such backward spread of selective reinforcement, but there was no clear calculus for doing this. Thus, while intriguing, S-R mediation theory did not catch on as a major tool for the study of language acquisition and behavior. It seemed right (given the precondition of utilizing associative theory at all), but imprecise and untestable. Furthermore, the strict adherence to behaviorist-reductionist principles to account for the formation of intervening variables made the entire scheme vulnerable to classic attacks.

Recent developments in computer-based "connectionism" have offered several different kinds of schemes for learning by adjusting the weights on connections. One learning scheme called *back-propagation* requires determining the difference in the observed and desired activity of output units (Bechtel and Abrahamsen 1991; Rumelhart, Hinton, and Williams 1986). Once the difference between observed and desired activity, or "error," is computed, it is used to adjust the weights on the connections. For example, if the activity of an output unit is too high, the weight on the connection is reduced; if it is too low, the weight is increased. The error further propagates backward to adjust the weights on connections between input units and intermediate units, according to how much each intermediate unit contributes to the error. This gives an automatic and powerful scheme for training networks with intermediate levels.

Consider how this scheme would apply to Osgood's word association network (see section 2.2.2). In a learning framework, the model's task is to match as output the word and associates of the word that was given as input. Of course, with only a few words of each, and a few intermediate nodes, there is no problem, so consider a world of hundreds of words as input and output, just as Osgood did. A learning trial consists of activating an input word and examining the output. If the output is the incorrect word, the first step is to reduce the weight on that word and increase the weight on the correct word. The activation links that lead to these responses or potential responses from the intermediate layer will be adjusted according to their degree of match to the correct word. Then for each intermediate node, the weights will be adjusted as well.

This simple scheme has changed the fortunes of associative mediation psychology. It is now possible to formulate and test specific models without having to tweak the strengths of connections to the intermediate layers of connection points. We have mediation associationism without behaviorist limits on intervening structures—a decidedly liberating advance.

4.3.2 Limits of Connectionism, Same or Different?

What can this tell us about the potential effect of statistical eccentricities on the learning and use of syntactic structures? A variety of learning models have been constructed that convey a structural analysis of many aspects of syntactic structure (e.g., Rumelhart and McClelland 1986; Elman 1990). Many of these are "toy" models, in that they work only on highly tailored input. In an example pertinent for the model we present in chapter 5, Juliano and Bever (1988) explored the extent to which phrase structure segmentation can be based on statistical properties of sentence boundaries. They set up a training model with an input array for examining three words in a row, a hidden layer of nodes, and an output node. The model had a recognition vocabulary of about 200 function words, and recognized the letter length of all words. Unrecognized words were classified in terms of the distance to the nearest recognized word to its left; see (46).

(46) *3-Word Input*
The horse that
horse that was
that was raced
was raced past
raced past the
past the barn.

The model examined three-word sequences, as shown above—for example, words 1–3, then 2–4, then 3–5, and so on. Punctuation and capitalization were opaque to the model but were used after each trial to give feedback on the location of actual sentence boundaries. At each new subsequence, the model was trained on sentence boundaries to predict that the space between the second and third words is in fact a sentence boundary; the higher the output value, the more likely there is a sentence boundary. In the preceding example, only the sixth triplet would have been reinforced as positive. After training on a text of several thousand sentences, the model did a good job of predicting sentence boundaries. This reflects the fact that sentences end and begin in characteristic patterns. But the boundary between two sentences is also the boundary between two phrases. This raised the possibility that the model would generalize well in predicting where the phrase boundaries are within sentences. In fact, this is true, as shown in (47)—the numbers correspond to the value of the output.

(47) a. Who (.036) is (.020) it (.498)?
b. We (.018) saw (.009) the (.073) children (.076) walking (.185) in (.079) the (.055) rain (.755).
c. What (.257) did (.009) you (.057) hit (.102) that (.079) time (.333)?
d. There's (.006) more (.131) in (.023) here (.501).
e. We'll (.005) save (.009) this (.108) one (.186) for (.085) later (.367).
f. What (.257) are (.009) you (.073) cooking (.101) down (.087) there (.427)?
g. Children (.071) rain (.315), walk (095) rain (.366), yes (.875)

These values can be converted to actual segmentation by two simple interpretive rules:

1. Every local maximum corresponds to a phrase boundary. For example, a value lower than the preceding value indicates that there is a phrase boundary after the preceding word.
2. Single words are joined to the right, unless sentence final.

These rules convert the initial output of the model into phrase structured sequences, which correspond well to where linguistic boundaries would occur. Some examples appear in (48). To illustrate, in (47a) the fact that the value of .020 after *is* is lower than the preceding value of .036 indicates that there is a phrase boundary after *what*, as shown in (48a).

(48) a. [[Who] is it?]
b. [[[We] saw the children] [walking] [in the rain].]
c. [[[What] did you] [hit [that time]].]
d. [[There's more] [in here].]
e. [[We'll save this one] [for later].]
f. [[[What] are you cooking] down there.]
g. [[Children rain,] [Walk rain, yes].]

This is not a toy model—that is, it generalizes quite well to normal English text, not just some chosen subset. Thus it would seem to "solve" the problem of how children might discover within-sentence phrase structure segmentation, by generalizing from the patterns of sentence boundaries. But it is important to examine closely how the model really works. First, it solves the lexical decoding problem by fiat. Words are simply operationally defined as letter strings between spaces. Second, it resolves the stimulus-identification problem by having a vocabulary of fixed words of interest—the closed class—and defines other words in terms of the most recently occurring closed class item. Third, the grain and reinforcement problem is solved for it by several constraints—(1) considering only three words at a time; (2) aiming at predicting the ends of utterances; (3) focusing the prediction on the second of the three words, rather than, for example, the third. Fourth, the model generalizes well to utterance-internal phrasing only if the utterance-final boundaries are "sharpened."

Thus, the model reveals that there is a great deal of structure in utterances that can be extracted by a properly configured associative system. No doubt other preconfigurations might end up with a similar result, but each will work because it has sufficiently pretuned mechanisms designed to pick up the actual phrase structures that structurally constrain word sequences.

Several current comprehension models have been developed that attempt to use statistically valid structures as their basis. As we will see, they all share the same kind of limits and virtues.

4.3.3 The Competition Model

MacWhinney (1987:251) formulated a scheme for comprehension that centers on the attachment of roles to predicates. The system is based on lexical knowledge, "as an organizer of auditory semantic ... and role-relational (i.e., syntactic) knowledge." Possible frames associated with lexical items "compete" with each other to determine the assigned information during comprehension. In general, the attempt is to base language structure on allegedly general cognitive principles. Indeed, conventional aspects of syntax are viewed as artifacts of linguistic theory:

> Phrase structures are epiphenomena, with the core of the grammar being composed of the arguments entered on particular predicates. By relating arguments to predicates, the listener builds up something that looks like a parse tree, [but because the parse is built up out of locally statistically valid attachments] there is no need for a separate encoding of phrase structure rules, since the correct patterns emerge from the operation of predicate-argument relations. (MacWhinney 1987:268)

The kind of "phrase structure" that emerges from this scheme is a form of directed graph, with predicates having connected links to their arguments.

Consider one of MacWhinney's examples (adapted from MacWhinney 1987):

(49) The dog ate the bone
((the → dog) ← ate → (the → bone))

The comprehension system arrives at this structure by linking a series of roles and cues that are associated with each word as it is encountered.

The essential concept is "competition" between structures, based on the relative validity of "cues" for particular analyses. The primary kinds of cues to role relations are word order and morphology. Children learn the validity of these cues as markers of particular verb roles, and automatically weight the competing cues in ongoing comprehension. Each verb specifies a set of possible roles it can assign to nouns, and then various cues combine to specify a particular role for the nouns in its argument domain.

For example, *the* is specified as having a head that follows it, so when *dog* appears next, it can be linked to the Head position defined by *the*. *Ate* has a characteristic description of its subject (animate and preceding it), and its object (following it); these roles and associated cues then facilitate the linking of the nouns into the appropriate predicate roles.

In addition to conventional roles (e.g., subject, object, indirect object), there are a number of additional, more structural sorts of "roles" such as "exohead" and "relhead." An exohead allows for structures that take more than one argument. For example, "prepositions take two arguments, [the one separate from the head] is the verb or noun which the whole prepositional phrase modifies ... the 'exohead'.... In a sentence such as 'they discussed the dogs on the beach', the exohead of 'on the beach' could be either 'the dogs' or 'discussed'" (MacWhinney, 1987:262). (Note that this corresponds to S/VP prepositional phrase attachment ambiguities discussed in sections 4.2, 4.2.1.)

The most complex grammatical role is that of the head of a relative clause ... "relhead." Relheads allow for linking of main and relative clauses. This kind of scheme can account for garden paths in terms of the relatively strong cues for an initially misleading analysis. Thus, in reduced relative constructions, the initial cues for a main-verb analysis are so strong that it is difficult to arrive at the correct analysis. In this case, the wrong set of cues and links "compete" so effectively that they create a misleading analysis.

The competition model is an ambitious project and is quite consistent with the concept that perceptual strategies assign structure, as discussed briefly in chapter 2. The notion of relying on surface and local cue validity to assign structure is exactly the same in both theories. Furthermore, the inclusion of both morphological markers and word order "cues" supports the role of patterns distributed across lexical items. The cue $N \ldots V$ is a pattern relation between lexical items in a sequence, not an intrinsic property of either lexical item.

The competition model has been elaborated and spelled out in somewhat greater detail than the strategies model. It has also sparked wide-ranging research on the acquisition of sensitivity to language-specific cues in different languages. Thus, it serves as a useful framework for proceeding with various kinds of empirical studies of the role of statistically valid cues in sentence comprehension. The model necessarily exhibits both the virtues and failings of a pure statistical model. First, it does not even address the grain problem—how to establish a natural level along which the evidence for cues is assembled. The authors often stipulate which kinds of cues are universal—for example, word order and morphology. But they do not offer a theory as to how a learning mechanism would automatically arrive at just these cues and not others (Chomsky 1959; and discussion in chapter 2).

Second, the model does not deal clearly with the question of whether syntax is prior to it, unnecessary, or mimicked by it. Over two decades of writing, the model's proponents consistently attack the notion of an autonomous syntax, yet the model itself looks very much like a variant of syntax, insofar as it works, can deal with recursive structures, and so on. The authors face a conundrum. Either the model simulates the essential features of a correct syntax or it does not. If it does, then it serves merely as a strongly equivalent structure, without guaranteeing the claim that it can be learned through independent cognitive mechanisms. If it does not, it is not adequate to the facts of a recursively rich linguistic system.

Consider the implications of the fact that language is recursive. For example, there is no principled limit on the number of relative clauses and other elaborations on an initial noun. Such elaborations intervene between the noun and the verb it agrees with.

(50) a. The evidence has proved my innocence.
b. The evidence in the box has proved my innocence.
c. The evidence in the box surrounded by armed guards has proved my innocence.
d. The evidence in the box surrounded by solemn armed guards in bright uniforms has proved my innocence.

None of the preceding sentences is particularly difficult to comprehend. Even if one limited such material to ten words, the number of different constructions would be enormous. Intuitively, the reason is that the first noun phrase is treated as the head of a phrase, which includes all of the intervening material up to the verb. But accessing the abstract notion of noun phrase (and computing the correct structure and import of the intervening material) is tantamount to utilizing syntactic structure. Technically, the competition model could propose that the only aspects of syntax involved in language learning and comprehension are just those that make reliable and salient cues available enough for the language to be usable. Thus, the model might presuppose a "syntax," but not one with a derivational history for sentences, empty categories, and so on. Of course, such a syntax would be held accountable for the same kind of other behavioral facts about language, such as native speakers' intuitions. It does not seem likely that the competition model has hit on just the right scheme, where other theories have failed. However, if it did, or does, it will be a significant achievement to show that apparent syntactic structures really are functionally determined by the needs of acquisition and processing by a general-purpose learner and perceptual system.

For the moment, however, we see the usual difficulty. If a syntax is presupposed, it is clear what the cues are valid for, and the model makes sense. If syntax is not presupposed, the model has no framework to specify potential cues or to determine the significance of their validity.

4.3.4 Constraint Satisfaction

With the fundamental tools of connectionism—simple processing units whose strength of association to each other can vary over time—it should be possible to model the effects of various syntactic and semantic properties in the initial interpretation of a sentence. This section reports several attempts to model the use of associative information at the earliest points of sentence comprehension. MacDonald, Pearlmutter, and Seidenberg (1994:682) propose a lexical model of processing that indeed presupposes all the syntactic information of the kind proposed in a current generative grammar: "Both lexical and syntactic ambiguity are governed by the same types of knowledge representations and processing mechanisms." They resurrect Chomsky's (1957, 1965) traditional distinction between competence and performance and argue that performance is completely determined by local lexical knowledge, while the actual representations that build up can be abstract and rich. The main goal of their proposal is to show that even so-called "syntactic" ambiguities are actually resolved on the basis of local lexical knowledge (following Ford, Bresnan, and Kaplan 1982). For example, in (51)

(51) The witness examined by the lawyer was lying.

the garden-path ambiguity is resolved as a function of the lexical categorization of *examined*, either as a simple past or as a past participle. The garden path itself is created by the initial illusion that *examined* is the simple past, supported by the relative frequency that *examined* has been experienced as a simple past, as well as the likelihood that *witness* is an agent because of its animacy, and other factors. Thus, the constraint-based lexicalist model should contrast with models that build syntactic structure by directly accessing sentence-level grammatical information. It also should contrast with those probabilistic models that use category patterns that extend across lexical items. It is not clear that the theory successfully contrasts with either.

Their method is first to demonstrate that "syntactic"-level ambiguities have behavioral properties similar to lexical ambiguities. They are both sensitive to context information and frequency information with similar parameters. At both levels, there is an intrinsic "base" probability of a particular meaning and structure. Context influences the processing most when the "base" probabilities are low and relatively equal. This eclectic array of information can include access to adjacent [prior] words.

MacDonald, Pearlmutter, and Seidenberg (1994) illustrate the formation of syntactic structure from lexical information with the sequence in (52)

(52) John cooked.

As shown in figure 4.7. MacDonald, Pearlmutter, and Seidenberg (1994) assume a distributed representation of lexical information, so that a lexical item is "activated" when the nodes that correspond to its properties are activated. Reading the word *John* activates the semantic features associated with *John*, including the fact that *John* is animate, human, male, and so on. It also activates the lexical category information associated with *John*, including the fact that *John* is a noun and by X-bar theory *John* appears in a noun phrase. The various thematic roles in which *John* can appear are activated, and since it is animate, the agent role is strongly activated. Other factors, such as how frequently *John* appears in different roles, discourse constraints, and so on, will influence the activation of the thematic role. When *cooked* is read, information about its meaning, lexical category, voice, argument structures, and X-bar structures is activated. In each case, the level of activation of alternative interpretations is related to their frequency of use. If we assume that *cooked* most frequently appears with an ⟨agent, theme⟩ argument structure, the corresponding X-bar structure, providing syntactic slots for an agent and a theme, will be activated most strongly. When the most strongly activated X-bar structures for *John* and *cooked* correspond, as in this case, they are linked, and *John* is assigned the agent role, with the theme role needing to be filled. This example illustrates how syntactic structure emerges from lexical information.

The model developed by MacDonald and colleagues explicitly presupposes all necessary syntactic information, thereby relegating the explanation of the syntactic representations to the problem of language acquisition. This is fair enough, given the goal of accounting for details of reading comprehension. If the connectionist approach is to be a general theory of language behavior, it should provide an associative-connectionist model of acquisition that will arrive at the kinds of syntactic knowledge the reading model presupposes to be available. It should also provide a general solution to the grain problem: Which kinds of syntactic and semantic information are reinforced on each comprehension event? Attempts to solve these problems are being made, but results are not yet available (Seidenberg 1997).

The last half decade has seen the development and implementation of a number of specific, working connectionist models using the ideas of constraint satisfaction in a number of important papers. Much of this work is due to Tanenhaus and his colleagues. The remainder of this section is devoted to reviewing some of these working models. We emphasize here the models, some of their features, their uses, and what they can do. We reserve for chapter 7 discussion of the evidence from human comprehension that they have modeled.

Learning Subcategorization and Frequency Information Connectionist models should be sensitive to statistical properties of text. Juliano and Tanenhaus (1994) constructed a connectionist model to illustrate how both verb subcategorization

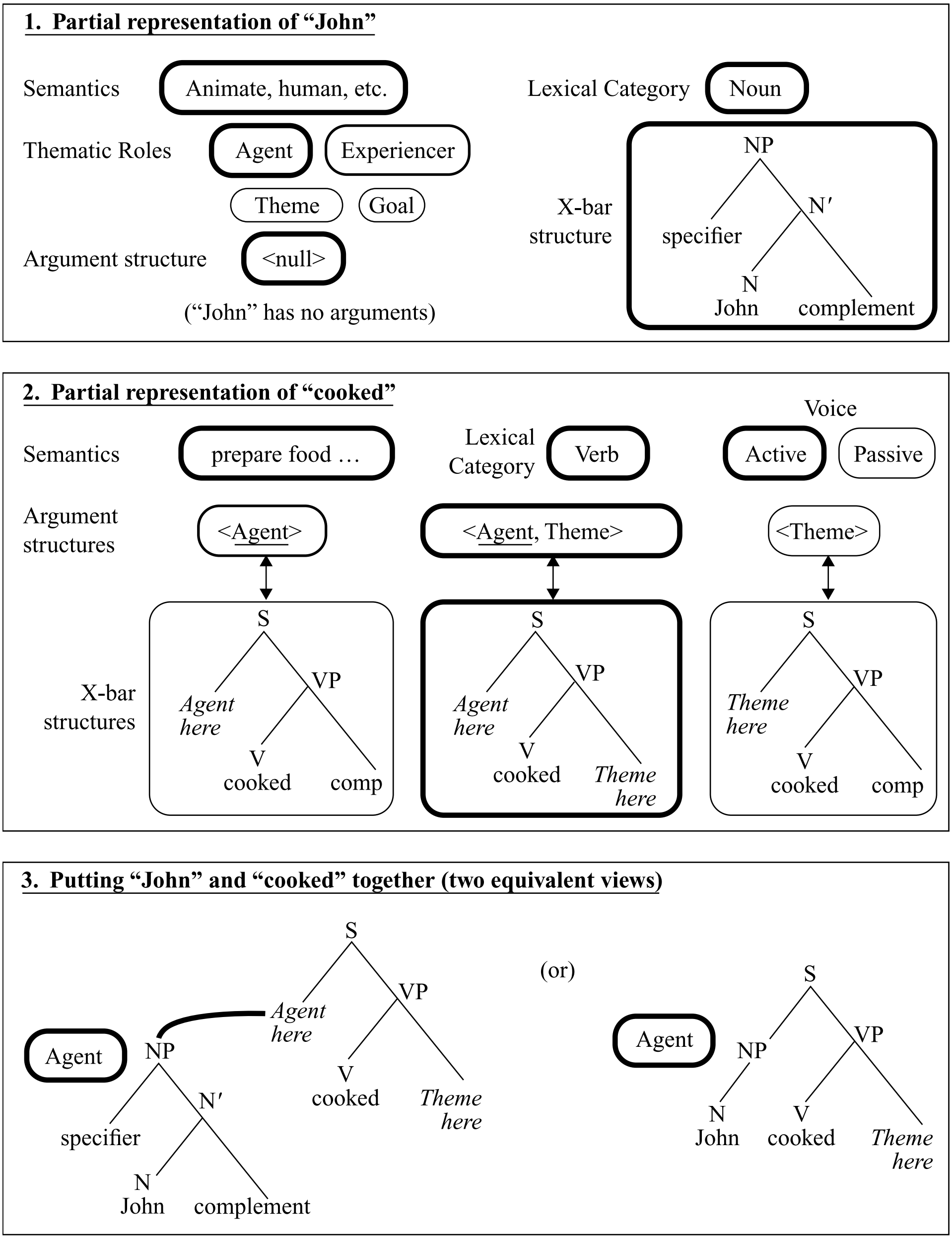

Figure 4.7
Representations in the processing of *John cooked* (reprinted with permission from MacDonald, Pearlmutter, and Seidenberg 1994, fig. 3). Partial representations of lexical information appear for individual words in parts 1 and 2. Part 3 combines possible combinations of these lexical representations into a tree structure.

information and verb frequency influence reading times in sentences like (53a) and (53b).

(53) a. The student read the book was stolen.
 b. The student implied the book was stolen.

Earlier studies (Juliano and Tanenhaus 1994; Trueswell, Tanenhaus, and Kello 1993) showed that verb frequency influences reading time. For example, reading times for *the* following the verb decreases as the frequency of use of verbs that require a sentential complement increases (*hinted*, *implied*, *thought*, *realized*; *hinted* and *implied* are low frequency, *thought* and *realized* are high frequency). But for verbs that require an NP complement, verb frequency has little effect on reading times for the word after the verb (*invited* is low frequency; *studied* and *gave* are high frequency).

Juliano and Tanenhaus (1994) presented sentences to a simple recurrent network in a word-prediction task. The network had 214 input units, 8 hidden units, and 8 output units. The input units consisted of 156 units for each of 156 verbs, 50 units for each word than could follow the verb, and 8 units that stored the latest activation of the hidden units. The sentences were from the Brown Corpus and the *Wall Street Journal* Corpus. Training involved presenting a verb and then using back-propagation to train the network to produce the appropriate type of phrase. The network received 13,051 training trials.

The network learned. For NP only verbs (*invited*, *studied*, *gave*), the model most often predicted a noun phrase. For S-bias verbs, the model most often predicted a sentential complement.

Immediate vs. Delayed Use of Meaning Connectionism can be used to construct detailed models of alternative theories of sentence comprehension. McRae, Spivey-Knowlton, and Tanenhaus (1998) examined the role of thematic fit between the initial noun and the initial verb in processing sentences with reduced relative clauses. In their studies, the initial noun was always animate, but in some cases, the initial noun was rated as a good agent for the verb, and in other cases a good patient.

(54) *Good agent*
 a. The cop arrested by the detective was guilty of taking bribes.

 Good patient
 b. The crook arrested by the detective was guilty of taking bribes.

Agent/patient-hood was determined by asking subjects to rate how likely is it for a cop/crook to arrest someone.

McRae and colleagues (1998) conducted two-word self-paced reading studies, and they constructed models to simulate immediate versus delayed use of thematic fit. The constraint-based model incorporated various constraints as soon as they became available:

1. Thematic fit on reading the initial verb
2. Relative frequency of the verb as past tense versus passive participle on reading the initial verb
3. The preposition *by*, which is a strong cue for a reduced relative on reading the verb + by frame (80% vs. 20%)
4. The relative frequency of main clauses versus reduced relatives on reading the initial verb (92% vs. 8%)
5. Thematic fit between the second noun and the initial verb on reading the second noun
6. The main verb, which is a definitive cue for a reduced relative on reading the main verb

The two-stage model applied the main-clause bias immediately, and the remaining constraints were applied one or two words after they conceivably could become available.

The two models were compared against actual human reading-time data. Mean reduction effects for humans closely followed the predictions of the constraint-based model and diverged from the predictions of the two-stage model. The human data showed that reduction effects were larger on verb + *by* for good patients than for good agents; there was a reversal on the second NP, and on the main verb there was a large reduction effect for good agents and no reduction effect for good patients.

The Normalized Recurrence Algorithm Connectionism can be used to develop sophisticated quantitative models of feedback. Spivey and Tanenhaus (1998) used connectionist modeling to reproduce conflicting results of studies of reading sentences with reduced relatives. For example, they found that discourse context eliminated the processing difficulty of reduced relatives, while Murrary and Liversedge (1994) did not. We review these and related results in chapter 7. For the moment, we consider the details of how their connectionist model of sentence processing works. (Readers familiar with connectionist methods for modifying associative strength may wish to move on to the following section on hybrid models.)

Spivey and Tanenhaus (1998) assumed for convenience a localist representation of various constraints and interpretations. That is, each of several constraints was represented by one node, and each of the alternative interpretations of a syntactically ambiguous sequence was represented by one node. The model was used to predict the reading times for the ambiguous verb (*selected*) in sentences like (55).

(55) The actress selected by the director believed that her performance was perfect.

At the point of reading *selected* two interpretations are possible. *Selected* may be interpreted as a past-tense verb and function as the main verb in the sentence, or it may be interpreted as a passive participle and function as the verb in a reduced relative clause. Several constraints bias the reader toward one or the other of these interpretations.

The model determines a provisional interpretation based on the strengths of each of the constraints. The model uses an algorithm called *normalized recurrence* to modify the activation of the constraint nodes, and then recomputes the activation of the provisional interpretation nodes. The model continues this cycle—compute activation of provisional interpretations, compute feedback, normalize—until a criterion is reached that stops the model. The interpretation node with the greatest activity at that point is the interpretation that is accepted. The number of cycles that the model goes through until it reaches criterion is by hypothesis related to difficulty in reading.

The diagrams in figure 4.8 illustrate this model. Figure 4.8 shows that four constraints influence the activity of the two interpretation nodes. First, there is the frequency with which *selected* has been experienced as a past-tense verb versus a passive participle. For the sake of illustration, suppose that analysis of a corpus showed that 40 percent of the uses of *selected* in the corpus were the past tense, and 60 percent were the passive participle. The initial activation of the nodes for the past tense and the passive participle will be .40 and .60 respectively (see figure 48a).

A second constraint is the bias toward interpreting an initial noun phrase + verb sequence as a main clause versus a reduced relative. Suppose a corpus shows that 85 percent of sentence-initial sequences of noun phrase + verb are main clauses and 15 percent are reduced relatives. The initial activation of the nodes for the main clause bias will be .85 for the main clause and .15 for the reduced relative (see figure 4.8a).

Third, discourse information influences the interpretation. Discourse may bias the interpretation of the ambiguous verb toward a past tense in an active-voice main clause, or toward a passive participle in a passive-voice reduced relative clause. In this case, if there are two actresses mentioned in the preceding discourse, a reduced relative clause may be expected in order to identify which of the two actresses is being referred to (see chapter 7 for more details). Suppose that a pretest showed that subjects who read the discourse and the initial noun phrase + verb sequence in the target sentence completed the sentence as a main clause in 33 percent of the trials, and as a reduced relative clause in 67 percent of the trials. This sets the initial activation of the discourse nodes to .33 for the main clause interpretation and .67 for the reduced relative interpretation (see figure 4.8a).

The last constraint that Spivey and Tanenhaus (1998) considered was parafoveal information from the word *by*. They assumed that in this eye-movement study, subjects obtained information from the word *by* when their eyes were focused on *selected*. Suppose that analysis of a corpus reveals that *by* following a verb ending in *-ed* introduces the agent in 85 percent of the cases, and it introduces some other

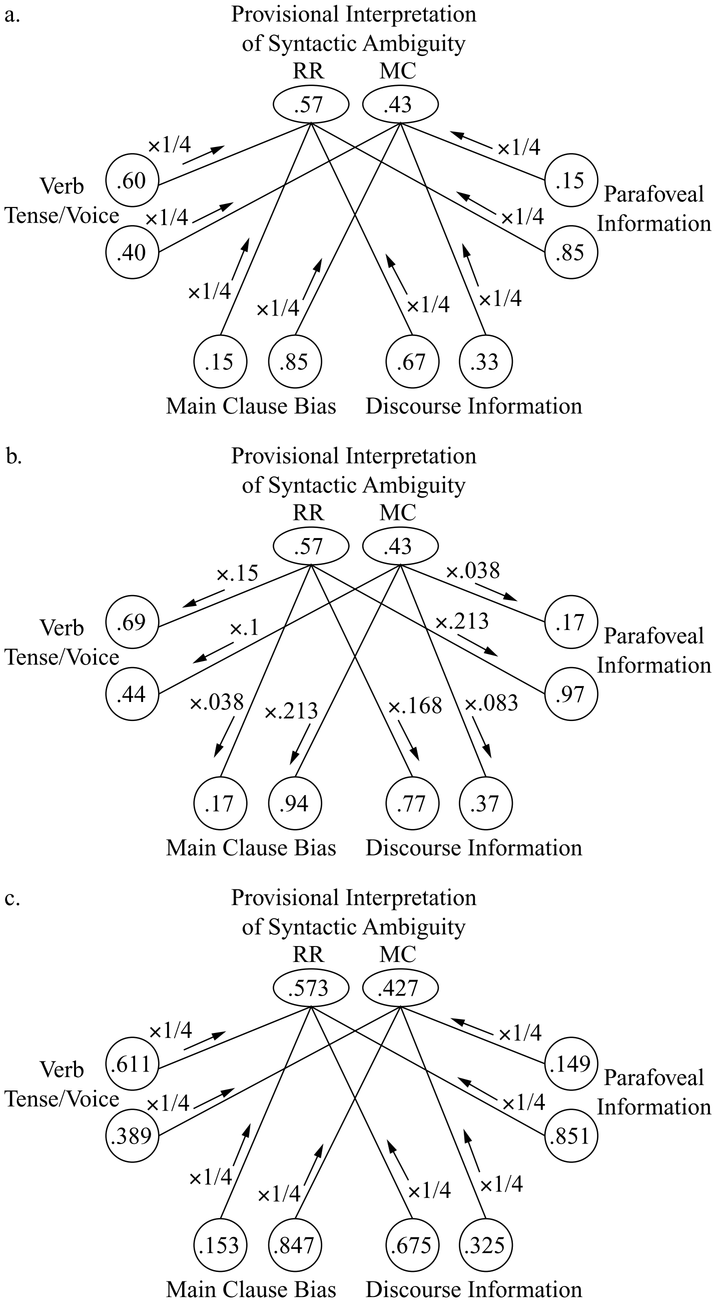

Figure 4.8
Spivey and Tanenhaus's normalized recurrence algorithm (reprinted with permission from Spivey and Tanenhaus 1998:1535, fig. 5). Part A shows how the initial activation of the reduced relative (RR) and main clause interpretations are the weighted activations of the four constraints. Part B shows that each weighted constraint is multiplied by the activation of the provisional interpretations. Part C shows the renormalization of the constraints to sum to 1.

construction, such as a locative phrase, in 15 percent of the cases. The initial activation of the parafoveal nodes will be .15 for the main clause and .85 for the reduced relative (see figure 4.8a).

Spivey and Tanenhaus assume that there are separate nodes that provisionally represent each of the two interpretations, main clause versus reduced relative. These provisional interpretations accept input from each of the four corresponding constraint nodes that support the interpretation. Because there are four constraints on each interpretation, Spivey and Tanenhaus assume that the weights on the connections between a constraint node and a provisional interpretation node are .25.

The initial activity of the provisional interpretation nodes is obtained by multiplying the weights by the initial activity of each of the constraint nodes. For example, for the reduced relative interpretation, the initial activation level is

.25[passive participle + (NP + V as reduced relative) + discourse bias toward reduced relative + *by*-bias toward passive]

$.25[.60 + .15 + .67 + .85] = 0.57$

as shown in figure 4.8a. For the main-clause interpretation, the initial activation is

.25[past tense + (NP + V as main clause) + discourse bias toward main clause + *by*-bias toward main clause], or

$.25[.40 + .85 + .33 + .15] = 0.43$

Once the initial values of the provisional interpretation nodes are determined, feedback to the constraint nodes is computed. That is, the activity of each constraint node changes to reflect the contributions of the other constraints. For example, a particular constraint node of the reduced relative interpretation (e.g., passive participle) will increase its activity level if all other constraint nodes support the reduced relative interpretation. A constraint node that supports the reduced relative interpretation will decrease its activity if all other constraint nodes support the main-clause interpretation. The amount of change in the activity of a constraint node is determined by multiplying the activity of the corresponding provisional interpretation by a weight. The weight is determined by multiplying the initial activity of the constraint by the initial weight on the connection between the constraint node and the provisional interpretation node. For example, the feedback weight for the passive-participle node is

(initial activity of constraint node) × (initial connection weight) = feedback weight

$.60 \times .25 = .15$

as shown in figure 4.8b. The feedback weight for the past-tense node is

$.40 \times .25 = .10$

Computing the new activity levels of the constraint nodes involves two phases. The first phase determines the amount of change in activity. The second phase normalizes this activity so that the activity levels of the two nodes for a particular kind of constraint add up to 1.0. In the first phase of feedback, the new activity level of each constraint node is determined by

$$[(\text{feedback weight}) \times (\text{activity of provisional node})] \\ + (\text{initial activity of constraint node})$$

Thus, the first phase feedback activation of the passive-participle node and the past-tense node are

$$[(.15) \times (.57)] + .60 = .69$$
$$[(.10) \times (.43)] + .40 = .44$$

as shown in figure 4.8b. Since these activation levels add to more than 1.0, the second-phase activation levels normalize the activation levels to 1.0. Normalizing the activation of the passive-participle and past-tense nodes involves the following computation:

$$(.69)/(.69 + .44) = .611$$
$$(.44)/(.69 + .44) = .389$$

as shown in figure 4.8c. These values add to 1.0, and they represent the activation of the constraints from verb-frequency information in the second cycle.

The procedure repeats so that new values of activation of the provisional nodes are computed, using the new activation levels of the constraint nodes and the weights on the forward connections. For example, the new activation levels of the reduced relative and main-clause nodes are

$$.25(.611 + .153 + .675 + .851) = .573$$
$$.25(.389 + .847 + .325 + .149) = .427$$

as shown in figure 4.8c. The activation of the reduced relative node has increased from .57 to .573, while that of the main-clause node has decreased from .43 to .427. These changes reflect the fact that three out of the four kinds of the constraints initially favored the reduced relative interpretation.

Intuitively, this interactive process captures the notion that ultimately, the winning analysis overtakes all the constraints. Furthermore, the less ambiguous or conflicted the initial analysis, the fewer cycles needed to arrive at a final interpretation. Spivey and Tanenhaus (1998) showed that this model with local representations and the normalized recurrence algorithm provided a good match with human reading-time data. In particular, the number of cycles that the model needed to reach criterion on a particular word was related to the amount of time that human participants spent reading the word. The model also matched the conflicting results from three studies

(Spivey and Tanenhaus 1998; Spivey-Knowlton, Trueswell, and Tanenhaus 1993; Murray and Liversedge 1994) when differences in materials were used to set the initial activity levels of specific constraints in different ways. Another factor that influenced the model's ability to duplicate the conflicting results was its use of differences in procedures, such as eye movement versus self-paced reading. For example, parafoveal information from *by* becomes available earlier in an eye-movement study than in a word-by-word self-paced reading study. The computations of this model are simple, yet when several potential constraints are included, its performance is impressive.

Arguments for Explicit Quantitative Models Deciding between competing theories of sentence comprehension requires explicit quantitative models that quantify sources of constraint and specify explicit mechanisms for relating information integration to processing time. Tanenhaus, Spivey-Knowlton and Hanna (2000) illustrate the difficulties of discriminating between alternative theories with examples like:

(56) *Hard*
   a. The horse raced past the barn fell.
   b. Sally warned the lawyer was greedy.

(57) *Easy*
   a. The land mine buried in the sand exploded.
   b. Sally said the lawyer was greedy.

The difficulties that people have in understanding the hard sentences seem to support a two-stage model like the garden-path theory (e.g., Frazier 1987a). The garden path that occurs as a result of the hard sentences suggests that comprehenders assign a structure based solely on syntactic-category information in order to reduce short-term memory load. Only after initial structural assignment does the sentence-processing system use lexical properties of words such as meaning and subcategorization information to revise the initial structural assignment. However, the relative ease of understanding the structurally identical sentences in the easy group suggests that information other than syntactic-category information, such as the subcategorization information of verbs, is used to determine the initial assignment of structure. In the absence of explicit quantitative models, it is a simple matter for advocates of two-stage models to dismiss the experimental evidence from the easy sentences by claiming that the experimental measures do not tap the earliest phase of comprehension, but instead tap the revision stage.

To advance our understanding of sentence-comprehension mechanisms, it is essential that we develop explicit theories that make predictions that can be proved false. Tanenhaus and colleagues demonstrate how detailed constraint-based models lead to simulations of human sentence processing. The results of human experiments on particular sentence materials are compared with the results of computer simulations that use identical materials. To the extent that the two sets of human and model results match up, the model is supported as an explanation of human sentence comprehension.

Tanenhaus et al. (forthcoming) review experiments and simulations on the processing of sentences with reduced relative clauses:

(58) a. The witness examined by the lawyer turned out to be unreliable.
b. The evidence examined by the lawyer turned out to be unreliable.

They identify the constraints that may operate to determine the initial structural assignment of *examined*:

1. Goodness of fit of the initial NP (*witness* vs. *evidence*) as agent or patient of the initial verb (*examined*)
2. Relative frequency of initial main clauses in English
3. Relative frequency of the initial verb (*examined*) as past tense versus passive participle
4. Relative frequency that the immediately following word (*by*) is used in an active (e.g., locative phrase) versus reduced relative construction
5. Lexical information provided by the second NP (*lawyer*)—for example, whether it is a good agent or patient of the initial verb
6. Bias from the main verb (*turned out*)
7. Whether the discourse context supports a reduced relative interpretation by mentioning two possible referents for the initial NP
8. Whether the discourse context supports a locative interpretation of the *by*-phrase in the target sentence

Tanenhaus et al. show how a constraint-based model that uses information as soon as it becomes available can mimic the results of human experiments. The model uses recurrent feedback and normalization to resolve ambiguities. The relative support that a constraint provides for a particular interpretation is related to the weight of the corresponding node. The activation of a provisional interpretation is calculated by determining the sum of its supporting constraint. The model calculates the activation of each provisional interpretation repeatedly until it settles on one interpretation. The length of time that the model requires to settle into a provisional interpretation is taken to indicate processing difficulty. The model generated results similar to those of a variety of experiments that investigated the constraints listed above (Ferreira and Clifton 1986; Hanna, Barker, and Tanenhaus 1995; Liversedge, Pickering, and Branigan 1995; Spivey-Knowlton, Trueswell, and Tanenhaus 1993; McRae, Spivey-Knowlton, and Tanenhaus 1998).

4.3.5 Dynamic Systems

Dynamic systems were developed to describe complex physical systems such as the weather or planetary systems. At any given moment, a dynamic system may be in any one of a number of states. The dynamic-systems approach uses a set of equations with several variables to describe how the system changes over time.

Tabor, Juliano, and Tanenhaus (1997) used the metaphor of dynamic systems to shed light on sentence comprehension. At any given moment, there is a preferred structural assignment for a sentence. To use the dynamic-systems metaphor, the possible structural assignments are arranged in a "representational space" that is organized according to similarity. As a sentence is heard, the system "gravitates" to a particular structural assignment, or "attractor" in the metric space. At any given moment, a number of variables influences which structural assignment or attractor is most strongly preferred. These variables include the relative frequency of a word in a particular syntactic category, the local context within the sentence, semantic information about words, discourse context, and so on. In cases of ambiguity, there is no strongly preferred assignment. That is, during temporary ambiguities, the processor is between two or more attractors. Additional information may lead the processor to move toward a particular attractor, just as a spaceship may be between planets, but some force moves the ship toward a particular planet.

Tabor, Juliano, and Tanenhaus (1997) use the sentence fragment *the insect examined ...* to illustrate the dynamic-systems metaphor for sentence comprehension. On receiving the three words in this fragment, the processor is between two attractors. For example, *examined* may have the structural role of the past tense of the main verb of the sentence, with *insect* as its subject, or it may have the role of a passive participle in a reduced relative clause. Factors that influence which of these attractors the processor is closer to include how frequently main clauses versus center-embedded reduced relatives are used, how frequently *examined* is used as a past tense versus a passive participle, and how likely it is that an insect examines something. For example, an insect may be more likely to be the object of the action *examine*, and so the processor initially may be closer to the reduced relative attractor than to the main-clause attractor. If the fragment contained a different noun that is more likely to examine something, such as *entomologist*, the processor initially might be closer to the main-clause attractor.

The researchers examined the usefulness of the dynamic-systems metaphor by conducting experiments on the processing of the word *that*, which is ambiguous between a demonstrative determiner and a complementizer, among other possibilities.

(59) *Demonstrative determiner*
   a. That cheap hotel was clean and comfortable to our surprise.
   b. The lawyer insisted that cheap hotel was clean and comfortable.

(60) *Complementizer*

a. That cheap hotels were clean and comfortable surprised us.

b. The lawyer insisted that cheap hotels were clean and comfortable.

Number information on the noun disambiguates *that*: if the noun is singular (*hotel*), *that* is a demonstrative determiner, but if it is plural (*hotels*), *that* is a complementizer.

Overall, *that* is used more often as a complementizer (70%) than as a demonstrative determiner (15%). However, when *that* appears at the beginning of a sentence, it is used more often as a demonstrative determiner (35%) than as a complementizer (11%). The relative usage of *that* reverses when it appears after the verb: the complementizer usage is more frequent than the determiner usage (93% vs. 6%).

Another factor that may influence the initial position of the processor in the representational space is the subcategorization requirements of the verb when *that* follows. For example, *insisted* is biased toward taking a sentential complement rather than an NP object complement, whereas *visited* is biased toward an NP object complement. Tabor et al. showed that these factors influence self-paced reading times after the disambiguating noun. They even found that reading times increased on the word *the* after verbs that require a sentential complement, suggesting that the NP object attractor exerted an influence on processing.

Tabor and Tanenhaus (1999) extend the dynamic model of Tabor, Juliano, and Tanenhaus to account for semantic effects in processing reduced relatives. They argue that syntactic structures arise as "emergent properties" of connectionist models, and that the dynamic models, which they call the *Visitation Set Gravitation* (VSG) model, demonstrate this. Thus, connectionist models induce syntactic categories from mere exposure to sentences. The VSG transforms the output of the simple recurrent network (SRN) into hypotheses about the structure of a sentence that is being parsed. The fact that the VSG model can identify the possible syntactic structures of a sentence is an advance over models that relied on a syntactically informed "oracle" to define error measures.

The VSG model can be described in terms of a representation set, metric space, attractors, starting points, trajectories, and basins of attraction. The representation set is the set of all representations that can be formed during the parse of a sentence. The metric space is the set in which distances between elements are defined; the metric space arranges states or representations according to similarity. A starting point is the position in metric space that is occupied when a word in the sentence is has been received. A trajectory is a path that the system follows over time. An attractor is a stable state that attracts trajectories that are nearby. A basin of attraction is the set of starting points that lead to a particular stable state. The change in the state of the system is described by a differential equation.

The VSG model works with a simple recurrent network that has input units, hidden units, and output units. There are recurrent connections in the hidden layer.

This SRN was trained to predict the next word of sentences. When its predictions became accurate, its weights became fixed and it was tested on new sentences. As the SRN was tested, the values of the hidden units were recorded. These values constitute the visitation set. As each word was presented, the corresponding hidden-unit states were recorded and defined as the visitation set. The hidden-unit states for each word of the sentence defined the starting points. The VSG system then changed according to its differential equation, and the system gravitated toward an attractor. The amount of time that the system took to approach an attractor was taken to correspond to human processing time.

Tabor and Tanenhaus (1999) use the VSG model to examine the distinction between semantic oddness and syntactic incongruity. The model uses both kinds of information immediately, but they have different consequences, which can be characterized as a "graded qualitative difference." The VSG model responded to syntactic violations by continuing its trajectory between attractors, whereas a semantic violation involved direct gravitation toward an attractor.

Taken together, the word of Tanenhaus and colleagues is the most impressive and empiriclly tested set of connectionist models for comprehension, at least defined as the problem of choosing between competing available analyses. The usual critical issues remain.

1. Does the model scale well, when advanced to deal with more than reduced relative sentences?
2. Does the model solve the grain problem? In particular, how is the "density" of the attractor space in Tabor and Tanenhaus constrained to form attractors that correspond to (and thereby "explain") syntactic categories?
3. Does the attractor space really replace grammatical structures and categories, and, as, in the other models we have discussed, does it "find" their behavioral doppelgängers, given that they are exposed already in the linguistic environment?

We do not wish to prejudge what can be. At the moment, the answer to each question may be negative, and seems so to us. But the inventiveness of the scheme, along with explicit recognition of the three problems, means we can also choose to be cautiously optimistic.

4.3.6 Hybrid Models

There have been recent attempts to integrate structural and associative approaches. In the models we review in this section structural information plays a central role in comprehension. This differs from the dynamic systems of the previous section, in which structure-like information allegedly emerges from the normal operation of a connectionist model.

Competitive Attachment The competitive-attachment model (Stevenson 1994) is a connectionist model that explicitly incorporates symbolic information from linguistic theory. The syntactic nodes proposed by government and binding theory are represented as processing units in a connectionist network. Figure 4.9 illustrates the processing network and the phrase structure tree for (61),

(61) The warden believes the report.

In this model each word activates the phrase structure properties that are associated with that word and potential attachments, including those that attach the word to the existing phrase structure network. For example, *believe* activates the categories V, V′, and VP, along with information about case, potential theta roles, and potential categories to be selected (e.g., object NP or sentential complement). Once attachment occurs, the activation of existing processing units is updated, and symbolic features are passed through the network. This processing continues until every processing unit in the network reaches a stable state. This stable state is either some minimum level of activation, or zero. When the activation level of an attachment node reaches zero, the node is disconnected from the phrase structure network. All possible attachments in a syntactic ambiguity are active to varying degrees until there is sufficient information to take the activation level of a potential attachment to zero. For example, receiving the word *will* after *report* establishes an attachment between the verb and a new inflection phrase, and takes the activation level of the connection between the verb and the noun phrase *the report* to zero.

Stevenson assumes that processing nodes gradually decay over time. This assumption provides several attractive properties. It allows the model to account for attachment preferences and filler-gap assignment without relying on strategies. For instance, the decay assumption explains why there is a preference for attaching *on the train* to *sleeping* rather than *saw* in the following:

(62) I saw the child who was sleeping on the train.

Since *sleeping* was attached more recently than *saw*, *sleeping* has higher activation and can compete more effectively for the phrase *on the train*. The decay assumption also provides an explanation for why attachment difficulty varies over a range of distances. For example, attachment becomes increasing harder in the order (63a), (63b), and (63c).

(63) a. I called the guy a rotten driver.
b. I called the guy who smashed my car a rotten driver.
c. I called the guy who smashed my brand new car a rotten driver.

The decay assumption also explains filler-gap preferences without relying on strategies. In processing a sentence like (64)

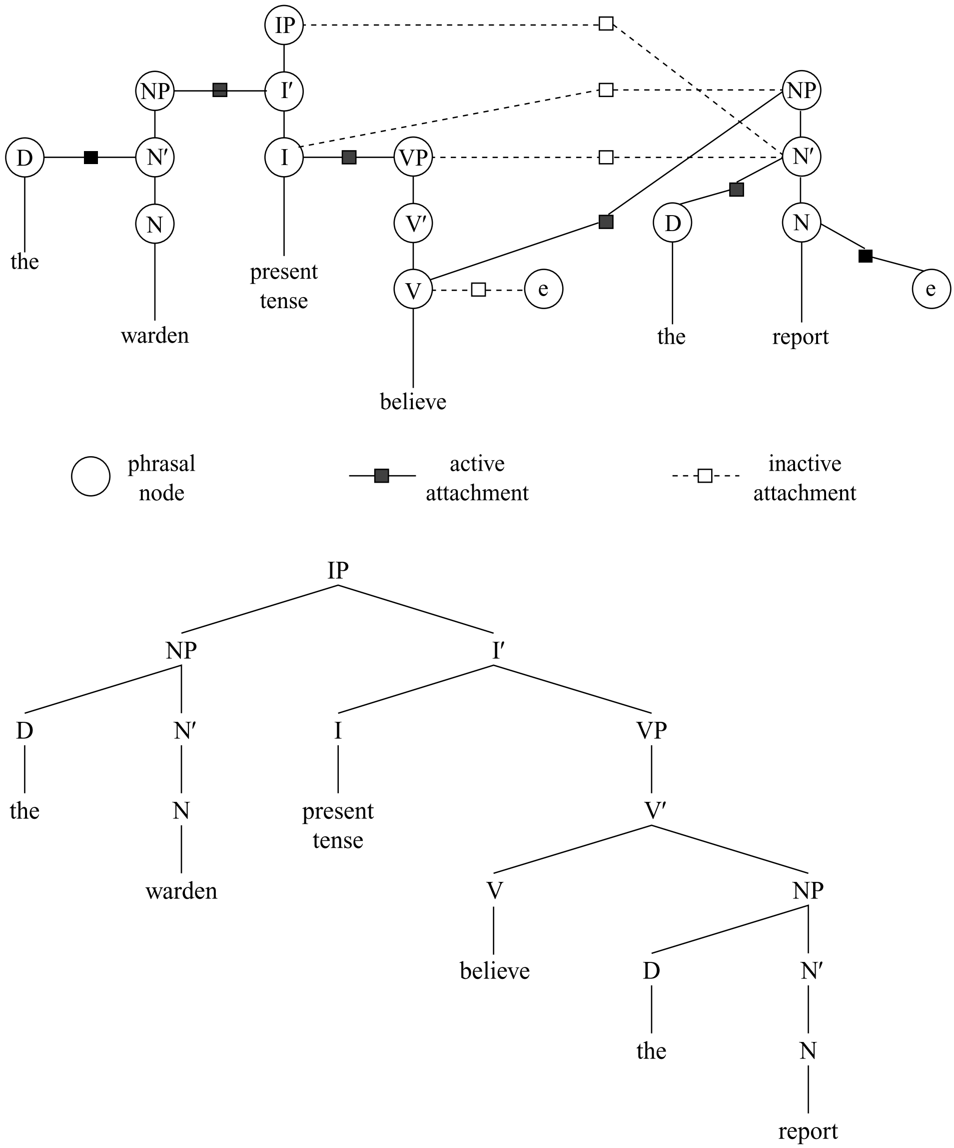

Figure 4.9
Stevenson's Competitive Attachment Model (adapted from Stevenson 1994:300, fig. 3.). Part (a) shows the network when *the report* has been attached. Part (b) shows the phrase structure tree that the network represents.

(64) a. Who did Maya kiss?
b. Who did Sara say that Maya kissed?

the word *who* establishes a binding node that is initially unfilled. When *kiss* is received, it establishes a binding node and sends its features throughout the network, establishing the binding relationship between *who* and *kiss*. The same process operates in a sentence like (64b). The difference between (64a) and (64b) is that this one establishes a chain that links the gap after *kissed* to a binding node that is linked to a gap after *say*, which is linked to a binding node that is linked to the filler *who*.

To summarize, the competitive-attachment model is an integration of a connectionist architecture and symbolic properties of language. It provides explanations of attachment and filler-gap preferences with a single mechanism. There is no need for strategies to explain these preferences.

Structure Frequency Jurafsky (1996) presents a model of sentence comprehension that relies on frequency of occurrence of possible structures. The model entertains multiple structures in parallel, and ranks them according to their prior probability of use. Structures that have very low probability are pruned—that is, no longer considered.

Jurafsky uses probabilities to establish the relative rankings of two linguistic properties: particular constituent structures, such as a noun phrase consisting of a determine and a noun, and particular arguments for a predicate. For example, the obligatory arguments for a predicate have a probability of 1.0 of being filled, while the probability of optional arguments being filled can range from 0.0 to 1.0. The probabilities are conditioned on lexical, syntactic, and semantic information.

A "beam search" algorithm is used to determine which structures are pruned. *Beam search* means that structures that fall within a certain percentage of the highest-ranked structure are retained, while those outside that percentage are dropped. Since a garden-path sentence occurs because the appropriate structure has been pruned, this sentence sets the highest value of beam width. A sentence in which two structures are entertained sets the lowest value of beam width. Jurafsky estimated the beam width to be between 3.8 and 5.6—that is, when a structure is 4–5 times more likely than another, both are retained.

The way the parser works can be illustrated with the following garden-path sentence:

(65) The complex houses married and single students and their families.

The initial interpretation is that *complex* is an adjective modifying *house*, but the correct interpretation is that *complex* is a noun and *houses* is the verb. Probabilities account for the garden path. The accompanying table shows the probability of the application of various rules for the two structures.

Adjective-Noun		*Noun-Verb*	
S → NP ...	[.92]	S → [NP[V ...	[.48]
NP → Det Adj N ...	[.28]	NP → Det N ...	[.63]
N → ROOT *s*	[.23]	N → *complex*	[.000029]
N → *house*	[.0024]	V → *house*	[.0006]
Adj → *complex*	[.00086]	V → ROOT *s*	[.086]

Since the nonlexical nodes do not differ much, Jurafsky omits them for simplicity in calculating the conditional probabilities of the two structures. The probabilities of the two structures are:

Adj-N: $.0024 \times .00086 \times .23 = 4.7 \times 10^{-7}$
N-V: $.0006 \times .000029 \times .086 = 1.4 \times 10^{-9}$

Combining these probabilities indicates that the adjective-noun structure is 267 times more likely than the noun-verb interpretation, much greater than the beam width. Thus, the less probable noun-verb interpretation is (incorrectly) pruned, and we experience a garden-path effect.

In other cases, two alternative structures that are similar in conditional probability are retained. The structure of the following sentence is similar to that of the preceding example:

(66) The warehouse fires a dozen employees each year.

However, there is little garden-path effect (Frazier and Rayner 1987). The probabilities of the two structures are:

NN: 4.2×10^{-5}
NV: 1.1×10^{-5}

Since the ratio of these probabilities is within the beam width (3.8), both interpretations are retained, and there is no garden-path effect.

4.4 Grains of Truth

This chapter has reviewed a wide range of approaches to sentence comprehension in contemporary psycholinguistics. Despite the diversity of approaches, there are some major themes and important insights that each approach contributes. We now summarize the themes and insights of contemporary approaches to sentence processing.

4.4.1 Statistical Models

Statistical models attempt to explain comprehension in terms of cues to meaning. Cues that are highly correlated with a particular meaning have high "cue validity."

Examples of models of sentence recognition that are based on cues with high validity are Tabor, Juliano, and Tanenhaus 1997, MacDonald, Pearlmutter, and Seidenberg 1994, and the earlier work of Bates and MacWhinney (1982), Bever (1970a), and Osgood (1963). The more recent models have developed impressive quantitative methods for weighting different kinds of statistically valid cues. These cues include knowledge of the likelihood of real-world events, such as the fact that dogs bite people more often than the reverse. They include knowledge of how frequently particular subcategorization frames are used—for example, that *believe* is used more often intransitively whereas *charge* is used more often transitively, as in

(67) a. The police believed the thief picked the lock.
b. The police charged the thief with burglary.

And they include knowledge of the frequency of use of sentence patterns or templates, such as the fact that a sequence of noun-verb-noun more often corresponds to agent-action-patient than to patient-action-agent. The probabilistic models that we reviewed showed that the comprehension system relies a great deal on statistical information.

As impressive as the probabilistic models are, though, statistical information cannot tell the whole story about comprehension. There are two senses in which statistical information falls short of a complete account of comprehension. First, statistical templates sometimes yield the incorrect meaning. For example, the simple strategy of interpreting a sequence of noun-verb-(noun) as agent-action-(patient) fails in five out of fourteen cases in the previous paragraph:

(68) a. ... that are based on cues ...
b. ... frames are used ...
c. ... *believe* is used ...
d. ... *charge* is used ...
e. ... that we reviewed ...

In each of these examples, the first noun is the patient, not the agent. By their very nature, statistical templates are guaranteed to give the incorrect meaning some of the time.

The second sense in which statistical templates fall short is that, even when they do provide the correct meaning, they do not provide a complete description of the structure of sentences. For example, in sentence (69),

(69) Several recent models have tried to integrate statistical information with comprehension.

the NVN template assigns the word *models* to the role of agent and the word *information* to the role of patient. These assignments are approximately correct, but the statistical template does not reveal that *to integrate statistical information with* *comprehension* is actually the object of *tried*. Nor does it describe the structure of the complex verb *have tried*, and so on. These structural details may not be necessary for a rough approximation to meaning, but they may be and can be part of the final representation of the sentence.

Statistical information provides, at best, a partial description of a sentence. Even when the most common template applies successfully, it does not provide all the details of a sentence. From the partial description that statistical information provides, the sentence-comprehension system can arrive at a complete description. The sentence-comprehension system must fill in details, just as the speech-perception system fills in details when a cough occurs at the * in (70a) to (70c) (see Warren and Warren 1970):

(70) a. The *eel was on the table.
b. The *eel was on the orange.
c. The *eel was on the axle.

The speech-perception system is able to provide the missing details by referring to a list of possible word patterns that fit the context (*meal*, *peel*, *wheel*, *deal*, *seal*, *kneel*, and so on). By referring to this list of possible word patterns, the speech-perception system can complete the rest of the pattern.

The value of this capability of filling in missing information is just as great for sentence recognition as it is for speech recognition. This becomes clear when one considers actual speech in natural settings. It is characteristic of such speech that there is noise at the level of the sentence: false starts, word substitutions, ungrammatical sentences, filled pauses, and so on (Osgood and Maclay 1967).

Pattern completion on the basis of a partial description works in the speech-perception system because the system has access to a list of possible patterns—that is, words. In the case of sentence recognition, the list of possible patterns is much larger: there is no limit to the number of possible sentences. For pattern completion to work in recognizing a sentence there has to be some way to access the list of possible sentence types. We propose that the sentence-comprehension system gains access to possible sentences through the grammar. It is the grammar that informs the comprehension system about the possible objects of perception.

4.4.2 Structural Models

There is a complementary "rock of truth" in contemporary structural models (Berwick and Weinberg 1986; Crocker 1996; Frazier and Clifton 1996; Gibson 1998; Gorrell 1995; Pritchett 1992). A parser that relies solely on knowledge of the grammar and the lexical categories of words (noun, verb, adjective, and so on) can provide much of the description of a sentence. In contrast to the statistical models, a parser equipped with the grammar and lexical categories of words may be able to provide (eventually) a complete description of a sentence. For example, various models that we reviewed in this chapter provide a description as well as a plausible explanation of the data for sentences like the following:

(71) a. The boy was kissed.
b. The horse raced past the barn fell.
c. Without her contributions will fail to appear.
d. Ian knows Thomas is a fool.
e. After the child sneezed the doctor left.
f. The reporter who the senator attacked admitted the error.

For a structural model to provide complete description of a sentence, it must have a procedure for determining when it has obtained all possible descriptions of a sentence. It also must have a way of continuing the computation of structure even after it has obtained one analysis for the sentence, since in some cases there is more than one analysis:

(72) a. Flying planes can be dangerous.
b. The shooting of the cowboys was disgraceful.
c. The duck is ready to eat.

Contemporary structural models often use nonstructural information to resolve ambiguity, generally after the parser has applied grammatical principles. Depending on the particular theory, this nonstructural information includes just the kinds of information that statistical models rely on: information about word meaning such as animacy, subcategorization information, information about events in the world such as the biting preferences of dogs and people, and information about common sentence patterns.

The various structural models have emphasized several important facts about comprehension: the insufficiency of lexical-category information, the importance of basic argument structure, the parser's ability to project structure, the limitations of processing capacity, and the dependence of reanalysis on degree of commitment to an already-computed analysis. We will consider these facts in turn.

Lexical Category Structural models agree that structural information by itself is not sufficient to produce a complete description of a sentence. The insufficiency of lexical-category information, for example, is clear in Crocker's model. In this model, a phrase structure module assembles a phrase structure using only lexical-category information. This phrase structure is checked by a thematic module, which has access to subcategorization information. This allows the phrase structure module to correct its initial error in (73).

(73) After the child sneezed the doctor left.

Basic Argument Structure Several of the structural models emphasize the central role of basic argument structure in comprehension. Frazier and Clifton (1996) ascribe importance to basic argument structure by limiting the principle of minimal attachment to argument positions. Pritchett's (1992) model emphasizes basic argument structure in its attempt to satisfy the theta criterion at every point during processing; every NP must be assigned a theta role, and every predicate must have all its obligatory theta positions filled.

Projection Several of the models assume that the parser can project information ahead. The nature of the projected information varies from model to model. Frazier and Clifton (1996) assume that the parser projects syntactic categories based on phrase structure rules. For example, the phrase *the boy* ... projects the minimal amount of structure necessary; in this case, the minimal structure is a verb phrase. Similarly, Gorrell (1995) assumes that verbs trigger the precomputation of syntactic categories, as when *what did Ian say* ... projects a noun phrase following *say*.

Processing Capacity The limit on processing capacity is a driving force behind most structural theories. Gibson's (1998) model relies most clearly on this assumption, in that both noun phrases that have not been assigned to a role and project roles that have not been assigned a noun phrase, have cost-limited processing capacity. In addition, limitations of processing capacity motivate the principle of minimal attachment (Frazier and Clifton 1996).

Degree of Commitment Most models assume some version of the idea that some kinds of reanalysis are easier than others. The central theme in predicting difficulty of reanalysis is degree of commitment to the analysis. Crocker (1996) makes this assumption most explicit, but it is apparent as well in other models. Gorrell (1995) distinguishes between a reanalysis that changes dominance and precedence relations from one that does not. Pritchett (1992) distinguishes between a reanalysis that involves reassignment to a new governing category from one that keeps an element in the current governing category.

4.5 Conclusion: Implications for an Integrated Model

We can use the grains of truth in contemporary structural models to sketch constraints on the architecture of the comprehension system. The comprehension system uses thematic requirements of verbs to project an argument structure, according to the most frequent use of a verb (as well as semantic information and plausibility).

Consider the process of understanding *He gave the dog a child*. On recognizing *gave* in *He gave* ..., the argument structure of an agent, patient, and a goal is activated. This structure is projected as the information that is needed to complete a syntactic/semantic unit.

(74) a. He gave the dog a bone.
b. He gave the dog to the child.

On receiving *the dog* in *He gave the dog . . .*, the probability of goals versus patients following immediately after *gave*, as in

(75) *Goal after verb*
a. He gave the dog a bone.

Patient after verb
b. He gave the dog to the child.

(as well as plausibility) will influence the initial assignment of *dog* to patient or goal. Receiving enough noun phrases to fill all of the necessary argument roles, in this case either *a bone* or *to the child*, will trigger a process of synthesizing a syntactic representation. This syntactic representation will be compared against the input to check for grammaticality. The comprehension system will fail when the generated syntax does not match the input. For example, having assigned *dog* and *child* the roles of patient and goal respectively, in our example input sentence

(76) a. He gave the dog a child.

the grammar may generate a sequence such as

(76) b. He gave the dog to a child.

When this syntactic representation is compared against the input sentence, a mismatch is detected and the system revises its assignment of noun phrases to argument roles. Instead of *dog* being patient and *child* being goal, the system assigns *dog* the role of goal and *child* the role of patient. The grammar generates the corresponding syntactic representation

(76) a. He gave the dog a child.

which matches the input.

As the comprehension system takes in the noun phrases to fill the projected argument roles, it must keep track of which noun phrases have been assigned to argument roles. It also must keep track of which roles are filled and which are not. Long or complex noun phrases will make it harder for the comprehension system to keep track of which argument role needs filling. Such a strain on memory capacity apparently motivates Heavy NP shift, such that long noun phrases tend to appear later in a sentence. Because of the increased demands on keeping track of which roles have been filled, the first sentence below is harder than the second:

(77) a. He gave the dog with long sharp teeth and short hair a bone.
b. He gave a bone to the dog with long sharp teeth and short hair.

These facts justify the assumption that keeping track of which argument roles have been filled can cause problems for the comprehension system. It appears that complexity of processing has greater effects on memory load than mere length. Shifting a long noun phrase to the end of a sentence does not greatly affect its difficulty. Thus, the following two sentences do not differ greatly in difficulty:

(78) a. He gave a book, a dog, a sled, and a railroad set to Bill.
b. He gave Bill a book, a dog, a sled, and a railroad set.

When an object noun phrase has a relative clause, however, processing difficulty is affected. Thus, the sentence below is harder than the two above, even though the number of words in the same:

(79) a. He gave a railroad set with an engine that produces real steam to Bill.

Shifting the complex object noun phrase to the end of the sentence makes the sentence easier, as in:

(79) b. He gave to Bill a railroad set with an engine that produces real steam.

These examples suggest that an object noun phrase with a relative clause in difficult because the relative clause triggers its own cycle of comprehension. The relative clause verb activates an argument structure, and the system searches for noun phrases that can fill those argument roles. When noun phrases have been received that can fill those roles, the grammar generates a syntax that is checked against the input to determine whether the input is grammatical based on the system's assignment of noun phrases to argument roles. It is not merely the number of noun phrases that must be assigned, but rather the completion of a semantic unit that strains the processing system. The completion of a semantic unit initiates checking of the generated syntax.

We can summarize the grains of truth of the contemporary models:

- Frequency of use of argument structures and sentence patterns influences the comprehension system's hypothesis about meaning, at least initially.
- Lexical-category information is not sufficient to produce a complete description of a sentence.
- Recognizing a verb allows the comprehension system to project the argument structure needed to complete a semantic unit.
- The grammar defines what is and what is not an acceptable sequence of words.
- It is harder to revise an analysis when the system has become more strongly committed to it.

Notes

1. It is not clear that the phrases in example 8 of (23) are accurately categorized as primary versus nonprimary. For example, Frazier and Clifton (1996) consider "purpose clauses" with a "to VERB ..." sequence as in (8a) to be arguments, and therefore subject to syntactic attachment onto the main verb phrase. They consider "rationale clauses" as in (8b) to be adjuncts, and subject to semantic association. These adverbial phrases, however, fit the usual criteria for adjunct phrases (Quirk et al. 1972):

(i) If an adverbial cannot appear initially in a negative clause, it is an adjunct.
  a. *Quickly* Nixon bought something.
  b. **Quickly* Nixon did not buy something.
  c. *Perhaps* Nixon bought something.
  d. *Perhaps* Nixon did not buy something.
  e. *To amuse us* Nixon bought something.
  f. ?*To amuse us* Nixon did not buy something.

This test suggests that *quickly* and *to amuse us* are adjuncts, *perhaps* is not.

(ii) If an adverbial can be contrasted with another adverbial in an alternative interrogation, it is an adjunct.
  a. Does he write to his parents *because he wants to* or does he write to them because he needs money?
  b. *Does he write to his parents *since he wants to* or does he write to them since he needs money?
  c. Did Nixon buy something *to amuse us* or did he by something to clean out his bank account?

This test suggests that *because he wants to* and *to amuse us* are adjuncts; *since he wants to* is not.

(iii) If an adverbial can be contrasted with another adverbial in an alternative negation, it is an adjunct.
  a. We didn't go to Chicago *on Monday*, but we did go there on Tuesday.
  b. Nixon didn't buy something *to amuse us*, but he did buy something to clean out his bank account.

This test suggests that *on Monday* and *to amuse us* are adjuncts.

2. The initial interpretation in many of these sentences is indeed governed by semantics. This fact becomes clear when we consider how the attachment preferences interact with purely syntactic phenomena. For example, there is no particular difficulty in processing the sentence depending on whether *yellow* modifies *broccoli* or *naked* modifies *John*. There is no preference for high attachment; rather, the parser appears to attach the modifier to the most plausible noun. However, when the sentence is made into a passive, there are clear differences in acceptability:

Secondary Predication, Passivized
(i) a. The broccoli was painted yellow/raw/*naked by John.
  b. The broccoli was painted by John *yellow/*raw/naked.

Serial order becomes important when the words appears in noncanonical order. In this case, the parser seems to follow the strategy of attaching the adjective to the most recent noun phrase. With active sentences, the parser is not constrained that way.

3. The separation of the processors into modules that deal with very specific kinds of information leads to difficulties for the parser. In particular, the phrase structure processor does not have access to subcategorization information. It raises the question of how the parser understands sentences when verbs like *sneeze* do take object-like constructions, as in:

(i) a. When the child sneezed his heart out ...
    b. When the child sneezed the goober into the handkerchief ...
    c. When the child sneezed his last sneeze ...

If the resolution to problem 1 is that the system is sensitive to frequency of particular subcategorizations, we are immediately on the slippery slope into a performance model that Crocker denies. Nonetheless, it would seem likely that if we compiled a list of verbs that can take objects but do so with differential probability, we would also find "graded" degrees of complexity at the garden-path site.

Second, the claim that subcategorization information is not accessed in assigning phrase structure may result in inconsistencies anyway. Consider:

(ii) a. The boy pushed the cat the ball.
     b. The boy pushed the cat to us.

Intuitions suggest that the first sentence above involves a small garden path at *the ball*. That is, intuitively, *the boy pushed the cat* is a complete sentence. But *the cat* has been initially miscategorized as the direct object. The subcategorization information in *pushed* will not resolve the local ambiguity, because *push* does not require an indirect object, and hence the direct-object assignment of *the cat* can end a well-formed sentence. Only when *the ball* is encountered is there information that *the cat* was miscategorized.

So far, Crocker could claim that this is all okay with his theory since the treatment of

(iii) a. The boy pushed the cat the ball.

is like that of

(iii) b. After the child sneezed the doctor ...

Notice, though, that Crocker ought to predict that the *sneeze* sentence is less of a garden path than the *pushed* sentence. This is because the *pushed* sentence cannot be noticed as a small garden path until arriving at the final phrase. The *sneeze* sentence, however, can be corrected right at the next word, based on subcategorization information in *sneeze*. Hence, the incorrect object assignment of *the cat* is more entrenched by the time recoding has to occur.

But now consider a verb that must have two arguments—for example, *give*. Why is the first sentence below no harder than the second?

(iv) a. The boy gave the cat the ball.
     b. The boy gave the cat to us.
     c. The boy donated the cat to us.

Crocker's model has the phrase structure processor making the same initially incorrect assignment, which also cannot be known to be incorrect until the words after the object, just as in *the boy pushed the cat the ball*. So, there ought to be the same kind of small garden path. But these isn't.

4. However, the patient-argument position, for example,

(i) The boy was hit.

is needed more than the agent argument position, for instance,

(ii) ? The boy hit.

This means that Crocker has no way of explaining the relative difficulty of passive.

5. Regarding subject raising, the deep structure attachment strategy is relevant. Crocker's theory incorrectly predicts that

(i) John seemed to be happy.

is harder than

(ii) John wanted to be happy.

since *John* in the *seemed* sentence does not occupy the deep structure position of *seem*. The *seemed* sentence, therefore, would require reanalysis.

6. The following examples show that subcategorization information is used in positing traces:

(i) Which cat did Bill push the ball?

(ii) Which cat did Bill push to us?

(iii) Which cat did Bill give the ball?

(iv) Which cat did Bill give to us?

Sentence (i) is harder than (iii). There is closure at *push* in (i) since *push* only requires a patient (*cat*), and hence a garden path when *the ball* is received. This garden path does not occur in (iii), because *give* requires a recipient. This suggests that the parser accesses subcategorization information.

(v) In what house_1 did you know Bill lives t_1?

(vi) In what house_1 did you say Bill lives t_1?

(vii) In what house_1 did you say t_1 Bill knows Enstein?

Crocker says that a trace is posited as soon as possible. So, for (vi) Crocker there is a trace posited at *say*. This is revised on receiving *lives*, so there should be a garden path ((vi) actually seems quite easy). The fact that semantics determines acceptability without any observable garden path is a problem for Crocker.