Chapter 5

The Dependency Locality Theory: A Distance-Based Theory of Linguistic Complexity

Edward Gibson

A major issue in understanding how language is implemented in the brain involves understanding the use of language in language comprehension and production. However, before we look to the brain to see what areas are associated with language processing phenomena, it is necessary to have good psychological theories of the relevant behavioral phenomena. Recent results have suggested that constructing an interpretation for a sentence involves the moment-by-moment integration of a variety of different information sources, constrained by the available computational resources (see, e.g., Ford, Bresnan, and Kaplan 1982; MacDonald, Pearlmutter, and Seidenberg 1994; Trueswell, Tanenhaus, and Garnsey 1994; Trueswell 1996; Tyler and Marslen-Wilson 1977; McClelland, St. John, and Taraban 1989; Pearlmutter and MacDonald 1992; Crain and Steedman 1985; Altmann and Steedman, 1988; Ni, Crain, and Shankweiler 1996; see Gibson and Pearlmutter 1998 and Tanenhaus and Trueswell 1995 for summaries). This chapter presents evidence for one theory of resource use in sentence comprehension: the *dependency locality theory* (DLT). If the evidence for a theory such as this one accumulates, it will then make sense to look for neural correlates of the theory (see Kaan et al. 1998; Harris 1998, for some initial attempts to find event-related potential measurements of brain activity corresponding to the components of the DLT).

An important part of a theory of sentence comprehension is a theory of how sentence structures are assembled—sentence parsing—as words are input one at a time. Two important components of sentence parsing consume computational resources:

1. Performing structural integrations: connecting a word into the structure for the input thus far.
2. Keeping the structure in memory, which includes keeping track of incomplete dependencies.

The DLT is a theory of human computational resources in sentence parsing that relies on these two kinds of resource use. One of the key ideas underlying the theory is *locality*, such that the cost of integrating two elements (such as a head and a dependent, or a pronominal referent to its antecedent) depends on the distance between the two. This idea is elaborated extensively in section 5.3. The remainder of this chapter is organized as follows. First, some empirical observations regarding the processing difficulty associated with unambiguous structures are presented in section 5.1. It is argued that a computational resource theory needs to account for these empirical observations. Some earlier theories of computational resource use are then discussed in section 5.2. The DLT is presented in detail in section 5.3. It is shown that the DLT accounts for complexity effects in unambiguous structures as well as preferences in ambiguous structures. A summary and conclusions are provided in section 5.4.

5.1 Evidence for Computational Resource Constraints: Nesting Complexity

One way to investigate the constraints affecting sentence comprehension is to explore the factors responsible for processing complexity in unambiguous structures. A general class of structures that are complex independent of any ambiguity are *nested* (or *center-embedded*) structures. A syntactic category *A* is said to be nested within another category *B* if *B* contains *A*, a constituent to the left of *A*, and a constituent to the right of *A*. Increasing the number of nestings soon makes sentence structures unprocessable (Chomsky 1957, 1965; Yngve 1960; Chomsky and Miller 1963; Miller and Chomsky 1963; Miller and Isard 1964). For example, the sentences in (1) are increasingly complex:[1,2]

(1) a. The reporter disliked the editor.
b. The reporter [$_{S'}$ who the senator attacked] disliked the editor.
c. #The reporter [$_{S'}$ who the senator [$_{S'}$ who John met] attacked] disliked the editor.

In (1a), no lexical material intervenes between the subject noun phrase (NP) *the reporter* and the verb on which it depends, *disliked*. In (1b), the relative clause (RC) *who the senator attacked* occurs between the NP *the reporter* and the verb *disliked.* This RC is therefore nested between the NP and the verb. In (1c), a second RC is nested between the subject NP *the senator* and the verb *attacked* of the first RC. This RC, *who John met*, is therefore doubly nested. The resulting structure is so complex that it is unprocessable for most people.

The difficulty associated with processing nested structures is probably caused by the quantity of resources they require during their processing. First, note that there is no local ambiguity in (1c), so the processing difficulty associated with this sentence is not related to ambiguity confusions. Second, note that the difficulty in understanding (1c) is not due to lexical frequency or plausibility, because sentence (2) contains the same words and expresses the same ideas as (1c), yet (2) is much easier to understand:

(2) John met the senator [$_{S'}$ who attacked the reporter [$_{S'}$ who disliked the editor]].

The RCs in (2) are not nested as they are in (1c), so (2) is not difficult to understand.

Multiply nested structures are complex across different structures in all languages. For instance, the Japanese examples in (3) are increasingly nested and are correspondingly increasingly difficult to understand:

(3) a. Ani-ga imooto-o ijimeta.
older-brother-nom younger-sister-acc bullied
'My older brother bullied my younger sister.'

b. Bebiisitaa-ga [$_{S'}$ ani-ga imooto-o ijimeta to] itta.
babysitter-nom older-brother-nom younger-sister-acc bullied that said
'The babysitter said that my older brother bullied my younger sister.'

c. #Obasan-ga [$_{S'}$ bebiisitaa-ga [$_{S'}$ ani-ga imooto-o ijimeta to] itta to] omotteiru].
aunt-nom babysitter-nom older-brother-nom younger-sister-acc bullied that said that thinks
'My aunt thinks that the babysitter said that my older brother bullied my younger sister.'

In (3a), the object NP *imooto-o* 'sister-acc' is nested between the subject NP *ani-ga* 'brother-nom' and the verb *ijimeta* 'bullied'. In addition to this nesting, the clause *ani-ga imooto-o ijimeta to* 'that the brother bullies the sister' is nested between the subject NP *bebiisitaa-ga* 'babysitter-nom' and the verb *itta* 'said' in (3b), so this structure is more complex. In addition to the two nested relationships in (3b), in (3c) the complex clause *bebiisitaa-ga . . . itta* 'the babysitter said . . .' is nested between the subject NP *obasan-ga* 'aunt-nom' and the verb *omotteiru* 'thinks', making this sentence even more complex—so complex that people have great difficulty understanding it at all. A less nested version of (3c) is provided in (4):

(4) [$_{S'}$ Bebiisitaa-ga [$_{S'}$ ani-ga imooto-o ijimeta to] itta to] obasan-ga omotteiru.
babysitter-nom older-brother-nom younger-sister-acc bullied that said that aunt thinks
'My aunt thinks that the babysitter said that my older brother bullied my younger sister.'

Japanese allows objects to occur before subjects. By placing the clausal object *bebiisitaa-ga . . . itta* 'the babysitter said . . .' of the verb *omotteiru* 'thinks' at the front of the sentence, this clause is no longer nested between the subject NP *obasan-ga* 'aunt-nom' and the verb *omotteiru* 'thinks'. Sentence (4) is therefore easier to understand than (3c).[3]

5.2 Previous Theories of Nesting Complexity

One of the earliest theories of nesting difficulty is that difficulty is indexed by the maximal number of incomplete syntactic dependencies that the processor has to keep track of during the course of processing a sentence (for related proposals see Yngve 1960; Chomsky and Miller 1963; Miller and Chomsky 1963; Miller and Isard 1964; Bever 1970; Kimball 1973; Hakuta 1981; MacWhinney 1987; Abney and Johnson 1991; Gibson 1991; Pickering and Barry 1991; Lewis 1993; Stabler 1994).[4] This hypothesis accounts for the increasing complexity of (1a) through (1c) as follows. In (1a), there is at most one incomplete syntactic dependency in the processing of the sentence. For example, immediately after processing the NP *the reporter*, there is one incomplete syntactic dependency: the NP is dependent on a verb to follow. This dependency is satisfied on processing the next word, *disliked*. Sentence (1b) is more complex, however, because processing this sentence requires passing through a processing state with more incomplete dependencies. In particular, there are three incomplete dependencies at the point of processing *the senator* in (1b): (1) the NP *the reporter* is dependent on a verb to follow it; (2) the NP *the senator* is dependent on a different verb to follow; and (3) the pronoun *who* is dependent on a verb to follow (and this ends up being the same verb that *the senator* depends on). Sentence (1c) is even more complex because there are five incomplete dependencies at the point of processing *John*: the same three incomplete dependencies discussed previously for (1b), plus two more: (1) the NP *John* is dependent on another verb to follow, and (2) the pronoun *who* is dependent on a verb to follow. Thus (1c) is the most difficult to understand of the three.

Sentence (2) is much easier to understand than (1c) because there is at most one incomplete dependency during the course of processing (2), far fewer than the maximum number of incomplete dependencies incurred during the processing of (1c). For example, at the point of processing *the senator*, there are no incomplete dependencies: the input up to this point is *John met the senator*, which is a complete sentence. On processing *who*, there is one incomplete dependency, in that this pronoun is dependent on a verb to follow. The target verb arrives as the next word, however, so the complexity does not increase. The rest of the sentence is processed similarly, with never more than one incomplete dependency. The Japanese nesting effects are accounted for similarly under the incomplete dependency hypothesis.

An extension of the incomplete dependency hypothesis is that complexity is indexed by the maximal number of incomplete dependencies of the same kind (Lewis 1993, 1996; Stabler 1994), where two syntactic dependencies are the same if the same case (such as nominative case for the subject of a verb, and accusative case for the object of a verb, and so on) is assigned in the relationship. According to this hypothesis, different kinds of incomplete syntactic dependencies do not interfere with one another, but similar incomplete syntactic dependencies simultaneously present at a processing state are difficult for the processor to keep track of. This extension of the incomplete dependency hypothesis is motivated by a variety of empirical observations regarding syntactic complexity (Lewis 1993, 1996; Stabler 1994).

According to the incomplete similar dependency hypothesis, the incomplete nominative and accusative case-assignment relationships at the point of processing the object NP *imooto-o* 'sister-acc' in (3a) do not interfere with one another. Thus the maximum complexity of processing (3a) is only one incomplete dependency, because there is never more than one incomplete syntactic dependency of the same kind at any processing state. On the other hand, in (3b), there are two incomplete nominative case assignment relationships at the point of processing the NP *ani-ga*, 'brother-nom', leading to a maximal complexity of two incomplete dependencies for this sentence structure. Sentence (3c) is even more complex under this hypothesis, because there are three incomplete nominative case-assignment relationships at the point of processing the most embedded subject *ani-ga* 'brother-nom'.

The incomplete similar dependency hypothesis also accounts for the English complexity contrasts observed earlier in a similar way. The maximal complexity of (1a) is one incomplete dependency, because there is at most one incomplete dependency of one kind at any point in the processing of this sentence. The maximal complexity of (1b) is two incomplete dependencies, because there are at most two incomplete dependencies of the same kind in the processing of this sentence: at the point of processing the embedded subject *the senator*, there are two incomplete dependencies involving nominative case-assignment. The maximal complexity of (1c) is three incomplete dependencies, because at the point of processing the most embedded subject *John*, there are three incomplete dependencies involving nominative case assignment.

5.2.1 Problems with Previous Theories of Nesting Complexity

These theories of nesting complexity reveal a number of empirical problems, one of which I will illustrate here. I will discuss other problems with the incomplete dependency approaches when I present evidence for the DLT. The problem for the incomplete dependency theories that I will concentrate on here is the lack of complexity of the examples in (5):

(5) a. A book [that some Italian [that I have never heard of] wrote] will be published soon by MIT Press. (Frank 1992)
b. The reporter who everyone that I met trusts said the president won't resign yet. (Bever 1974)

Sentences (5a) and (5b) are structurally similar to (1c) but much easier to understand.

(1) c. #The reporter [who the senator [who John met] attacked] disliked the editor.

All three examples contain an RC nested within an RC nested within the main clause of the sentence. For example, in (5a), the RC *that some Italian ... wrote* is nested between the subject NP *a book* and the VP *will be published*. And the RC *that I have never heard of* is further nested between the subject *some Italian* and the verb *wrote* of the outer RC. But (5a) and (5b) are much easier to understand than (1c).

There is an important difference between (5a) and (5b) on the one hand and (1c) on the other: (5a) and (5b) contain a pronoun (e.g., *I*, *you*) as the subject of the most embedded RC: *some Italian that I have never heard of* in (5a) and *everyone that I met* in (5b), whereas the most embedded subject is a proper name in (1c): *the senator who John met*. When the most embedded subject of nested RC structures is a pronoun, the structures are much easier to process (Bever 1970; Kac 1981).

Warren and Gibson (1999, also discussed in Gibson 1998) performed a questionnaire study evaluating whether this generalization was true. In this study, participants were asked to rate sentences for their perceived difficulty on a scale from 1 to 5, where 1 indicated a sentence that was easy to understand and 5 indicated a sentence that was difficult to understand. Warren and Gibson compared sentences like (1c) to sentences that differed minimally from (1c) by replacing the most embedded subject by a first- or second-person pronoun, as in (6), in which the proper name *John* is replaced with the pronoun *I*:

(6) The reporter [who the senator [who I met] attacked] disliked the editor].

A third condition was also compared, in which the most deeply embedded subject was replaced by an NP having the form of a definite description, such as *the professor*:

(7) The reporter [who the senator [who the professor met] attacked] disliked the editor].

The results of the questionnaire are displayed in the graph in figure 5.1. As can be seen from the graph, the structures with the embedded pronouns were rated significantly easier to understand than the other two kinds of structures. This effect was robust in both the participant and item analyses.

Complexity theories that rely on incomplete dependencies do not predict the observed complexity difference. Changing the content of the most embedded subject NP to a pronoun does not change the maximal number of incomplete syntactic dependencies in the structures. For example, there are maximally five incomplete syntactic dependencies in processing (1c), at the point of processing the most embedded subject *John*. If the most embedded subject is a pronoun, such as *I* in (6), the maximal number of incomplete dependencies is still five. Similarly, if only incomplete nominative case assignments are counted, there is the same maximal complexity in both (1c) and (6)—three in each—because the pronoun *I* takes part in the same kind of syntactic dependency relationship as *John*. As a result, these theories predict no dif-

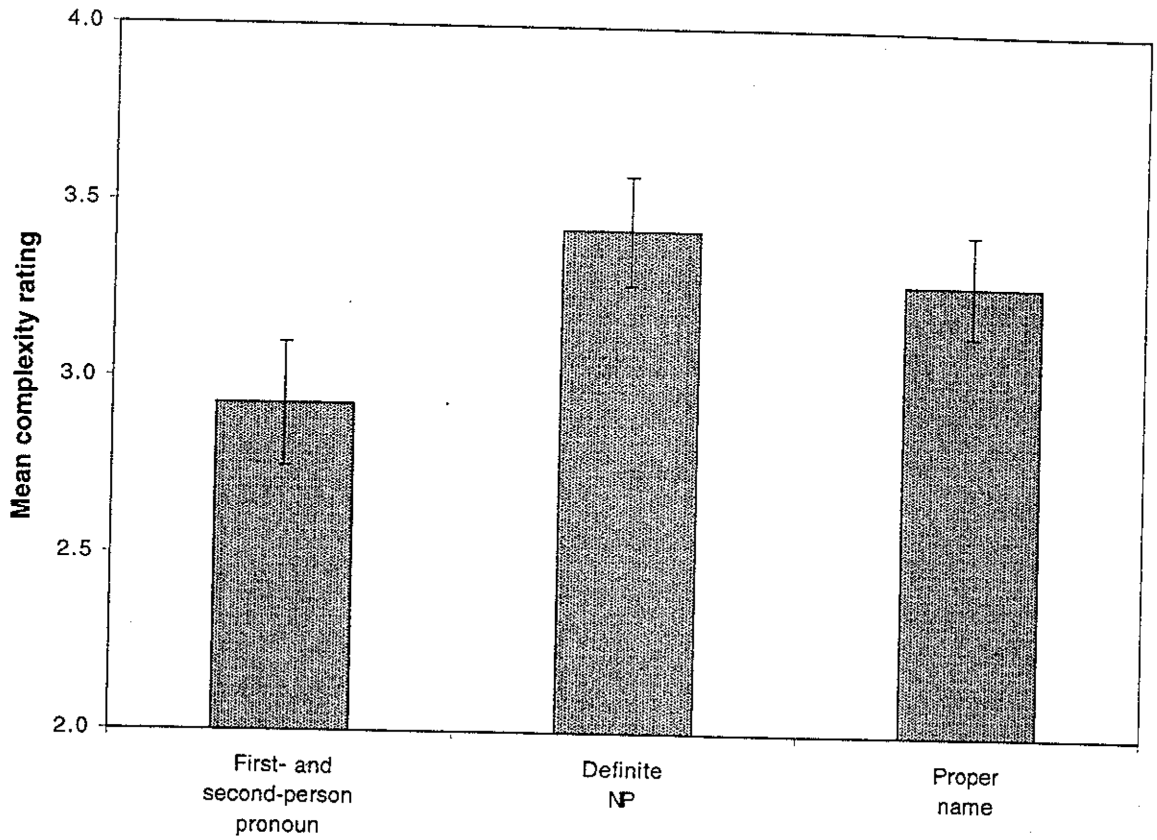

Figure 5.1
Complexity ratings for nested structures containing different kinds of NPs in the most embedded subject position (from Warren and Gibson 1999). The scale that participants used went from 1 (easy to understand) to 5 (hard to understand).

ference between (1c) and (6), but there is a large complexity difference between the two.

5.3 The Dependency Locality Theory

As a result of problems such as the one just exemplified, Gibson (1998) proposed a new theory of the use of computational resources in sentence comprehension: the *syntactic prediction locality theory* (SPLT). A variant of the SPLT—the *dependency locality theory* (DLT)—was also presented in Gibson (1998), along with some preliminary conceptual evidence in favor of the DLT over the SPLT. In addition, Gibson and Ko (1998) present empirical evidence from reading time studies supporting the DLT over the SPLT. Consequently, I will consider only the DLT here. There are two key insights in the DLT:

1. Resources are required for *two* aspects of language comprehension: (1) *storage* of the structure built thus far (as in earlier theories) and (2) *integration* of the current word into the structure built thus far. The integration aspect of resource use in sentence comprehension has been ignored in most resource theories preceding the DLT (and SPLT). The claim that both storage and integration consume resources is closely related to claims about resource use in other cognitive domains. In these domains, working memory resources are assumed to have both storage and processing/integration components (see Baddeley 1990; Just and Carpenter 1992; Anderson 1994; Lewis 1996).
2. The structural integration complexity depends on the *distance* or *locality* between the two elements being integrated (see other distance-based theories of linguistic complexity: Wanner and Maratsos 1978; Joshi 1990; Rambow and Joshi 1994; Hawkins 1990, 1994).

It turns out that many resource complexity effects can be explained using integration cost alone. As a result, I will first discuss the integration component of the DLT. I will then show how this component of the theory accounts for a number of complexity effects. Next, I will introduce the storage cost component of the theory.

5.3.1 Integration Cost

Following Gibson (1991), I will assume that maximal projections corresponding to the lexical entries for a newly input word *w* are constructed, with the speed of access dependent on the frequency of the relevant lexical entry (MacDonald, Pearlmutter, and Seidenberg 1994; Trueswell, Tanenhaus, and Garnsey 1994). Each of these maximal projections XP includes maximal projections of syntactic predictions for all the possible syntactic categories that can immediately follow *w* as the next word in a grammatical sentence. There are a number of components to the process of integrating XP and its semantic and discourse meaning into the discourse and syntactic structure(s) built thus far. First, there is a *structural integration* component, such that XP's syntactic category is matched with a syntactic expectation in the syntactic structure already built. This syntactic attachment will involve a head-dependent relationship between some category in XP (possibly the head *w*) and some projection of a head in the structure. The structural integration process also involves linking pronouns to their appropriate antecedents.

Following structural integration, there are also processes that interpret the resulting structural attachments (see Frazier 1978 for arguments that structural attachments precede contextual plausibility evaluations; see McElree and Griffith 1995 for evidence that this is the case). In particular, there is a process of discourse integration (e.g., constructing or accessing a discourse referent in the discourse model; see Kamp 1981 and Heim 1982) and one of evaluating the plausibility of the resultant discourse structure(s) in the current context, two processes that may run in tandem (or may be two parts of one more complex process). In this chapter, I will concentrate on the processes of structural integration and discourse processing.

It has been established the discourse processing literature that the difficulty of processing an NP depends on the accessibility of the referent of the NP in the discourse (Haviland and Clark 1974; Haliday and Hasan 1976; Garrod and Sanford 1977, 1982; see Garrod and Sanford 1994 for a summary). The less accessible the referent of an NP is in the discourse, the more resources are required to find or construct it (Warren and Gibson 1999). Focused entities or individuals, which are usually referred to with pronouns, are highly accessible, so they require a small quantity of resources to access. Nonfocused entities or individuals in the discourse require more resources to access. Such NPs are usually referred to using proper names and definite descriptions. Elements new to the discourse, which are usually introduced using indefinite NPs, require the most resources because they must be constructed in the discourse model. I will follow Gibson (1998) in assuming a simplified version of discourse-processing cost such that only the processing of new discourse referents consumes resources. A discourse referent is an entity that has a spatiotemporal location so that it can later be referred to with an anaphoric expression, such as a pronoun for NPs, or tense on a verb for events (Webber 1988). In particular, it is assumed that processing the head noun of an NP that refers to a new discourse object consumes substantial resources, and processing the head verb of a VP that refers to a new discourse event (also a discourse referent) consumes substantial resources, but processing other words does not consume substantial resources in the discourse processing component of structure building. This discourse-processing assumption is a simplification of Warren and Gibson's (1999) hypothesis, in that it only distinguishes costs for old and new referents, but no finer-grained distinctions are made according to accessibility.

As noted, it is assumed that the process of structural integration depends on the distance between the heads of the two projections being integrated together. The computational motivation for this hypothesis is that integrating a newly input maximal projection, XP, headed by h_2, with a previous syntactic category headed by h_1 (as in figure 5.2) involves retrieving aspects of h_1 from memory. In an activation-based framework, this process involves reactivating h_1 to a target threshold of activation. Because of the limited quantity of activation in the system, h_1's activation will decay as intervening words are processed and integrated into the structure for the input. Thus, the difficulty of the structural integration depends on the complexity of all aspects of the integrations that took place in the interim since h_1 was last highly activated. That is, the difficulty of the structural integration depends on the complexity of the structural integrations in the interim, as well as on the discourse integrations and the plausibility evaluations in the interim.

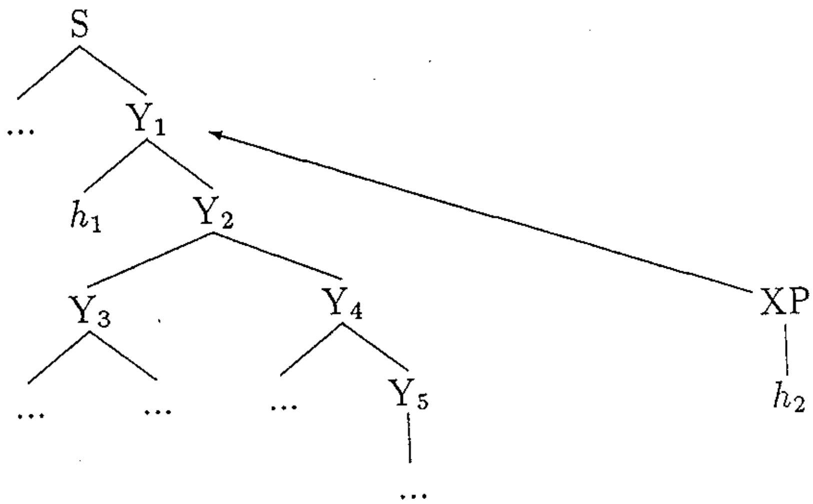

Figure 5.2
Structural integration of a maximal projection XP of a newly input head h_2 to an attachment site headed by a head h_1 in the structure for the input so far. Depending on the type of syntactic attachment and on the syntactic assumption, the structural integration may involve adjunction of a node in the structure built thus far.

In principle, this distance-based structural integration cost might be quantified in many ways. For simplicity, I will initially concentrate on the cost associated with building new discourse referents in the intervening region. (See Warren and Gibson 1999 for evidence for the more general proposal that discourse accessibility—not simply new vs. old referents—affects the distance-based structural integration cost.) Thus I will ignore the structural difficulty and contextual plausibility of the intervening integrations in computing the distance-based difficulty of a structural integration. To work out the specific predictions of the structural integration cost proposal, it is necessary to have a hypothesis about the relationship between the number of new discourse referents processed and the resulting structural integration cost. The structural integration cost apparently rises with the initial few intervening new discourse referents but then heads toward a maximal cost (Gibson 1998). For simplicity, however, I will assume a linear relationship between the number of new discourse referents and structural integration cost, such that one *energy unit* (EU) is expended for each new discourse referent in the intervening region. My assumptions about discourse processing and structural integration cost are summarized as follows:

(8) *DLT simplified discourse processing cost* (the cost associated with accessing or constructing the discourse structure for the maximal projection of the input word head h_2)
1 energy unit (EU) is consumed if h_2 is the head of a new discourse referent;
0 EUs otherwise.

(9) *DLT structural integration cost*
The structural integration cost associated with connecting the syntactic structure for a newly input head h_2 to a projection of a head h_1 that is part of the current structure for the input is dependent on the complexity of the computations that took place between h_1 and h_2. For simplicity, it is assumed that 1 EU is consumed for each new discourse referent in the intervening region.

To see how these assumptions apply, consider an example in which both h_1 and h_2 head phrases indicating discourse referents that were introduced by h_1 and h_2. Furthermore, suppose that two other discourse referents were introduced between h_1 and h_2. The cost of building the new discourse structure and connecting the phrase structure for h_2 to the phrase headed by h_1 would be 3 EUs, corresponding to 1 EU for constructing the new discourse referent, and 2 EUs for the structural integration cost, corresponding to the two intervening discourse referent heads.

In summary, the comprehension difficulty at a word in a sentence (e.g., as measured by reading times) is assumed to be determined by a combination of the following factors: (1) the frequency of the lexical item being integrated, (2) the structural integration cost at that word, (3) the storage cost at that word, (4) the contextual plausibility of the resulting structure, and (5) the discourse complexity of the resulting structure. In addition, there may also be some reanalysis difficulty, if the current word is not compatible with the most highly ranked structure built thus far (Gibson, Babyonyshev, and Kaan 1998). Furthermore, it is assumed that the overall intuitive complexity of a sentence depends to a large degree on the maximum intuitive complexity incurred at any processing state during the processing of a sentence. In the examples to be considered later, there are minimal lexical frequency and contextual plausibility differences throughout the processing of the sentences being compared, and there are no temporary ambiguities leading to reanalysis. So we may focus on structural integration, discourse complexity, and storage costs as the main contributors to complexity differences. I will initially consider the effect of structural integration and discourse complexity alone, ignoring storage cost. This is also an oversimplification, but the predicted patterns are similar when storage cost is also considered.

5.3.2 Accounting for Nesting Complexity Effects within the DLT

Consider the complexity contrast between the singly nested RC structure in (1b) and the doubly nested RC structure in (1c).

(1) b. The reporter who the senator attacked disliked the editor.
c. The reporter who the senator who John met attacked disliked the editor.

The maximal discourse and structural integration cost incurred during the processing of (1b) occurs at the point of processing *attacked*. At this point, one discourse referent is introduced—the event referent for the verb *attacked*—and three structural integrations take place:

1. The verb *attacked* is integrated as the verb for the subject NP *the senator*. No new discourse referents intervene. As a result, this integration step is cost free according to the simplified integration cost assumptions above.
2. An empty category to be coindexed with the RC pronoun *who* is integrated as the object of *attacked*. The attachment step is local, with no new discourse referents intervening.
3. The object-position empty category is coindexed with the preceding RC pronoun *who*. Two discourse referents were introduced in the interim—the NP *the senator* and the event referent *attacked*—leading to an integration cost of 2 EUs for this step.

The discourse and structural integration cost at *attacked* is therefore 3 EUs: 1 EU for the construction of the new discourse referent, and 2 EUs for the structural integrations. The discourse and structural integration cost at the following word *disliked* is also 3 EUs, corresponding to 1 EU for the construction of the event referent indicated by the verb *disliked*, and a structural integration cost of 2 EUs corresponding to the two new discourse referents—*the senator* and *attacked*—that separate the verb *disliked* and the subject NP *the reporter*, to which this verb connects. The total integration cost of 3 EUs, occurring at the verb *attacked* and at the verb *disliked*, is greater than at any other point in processing this sentence structure.

Now consider the doubly nested RC structure in (1c). The points of maximal total integration cost in this sentence also occur at the points of processing the verbs *attacked* and *disliked*, but the costs at these points are much higher here than in the singly embedded RC example in (1b). In particular, the total integration cost at *attacked* in (1c) is 7 EUs, corresponding to:

1. 1 EU for the construction of the event referent indicated by the tensed verb *attacked*
2. 2 EUs for the structural integration of the verb *attacked* to its subject NP *the senator*, corresponding to two discourse referents—*John* and *met*—that were introduced in the interim
3. 0 EUs for the integration of an empty category as object of the verb *attacked*
4. 4 EUs for the structural integration coindexing the object empty category with the preceding RC pronoun *who*, EUs because four new discourse referents—*the senator*, *John*, *met*, and *attacked*—separate the empty category and its coindexed pronoun

The maximal total integration cost in the doubly nested structure in (1c) is therefore substantially greater than that for the singly nested structure in (1b). Thus, according to the DLT, the cause of the processing difficulty associated with nesting complexity is simply that too many long distance structural integration steps take place at the same point in processing a nested structure.

It should be emphasized that aspects of the DLT have been oversimplified up to this point. Importantly, the structural integration cost function is not actually linear, so that a single long structural integration will never be as complex as multiple shorter integrations taking place at the same processing state. The nonlinearity of the cost function must be kept in mind when considering predictions of the DLT with respect to other linguistic structures. In addition, the distance function assumed thus far increments its cost for new discourse referents alone. Processing other kinds of elements also probably causes measurable increments in structural integration cost. Finally, the storage cost at a processing state also affects the processing load at that state. Despite these oversimplifications, the explanations provided by the DLT for nesting complexity effects are essentially the same when a more complete theory is considered. For example, although storage cost is important to account for some sentence processing effects (some of which will be discussed later), it is not critical for the comparisons just discussed, because the storage costs are the same at the locations that were compared (which will be verifiable below, once the storage cost component of the DLT is discussed).

In addition to accounting for general nesting complexity differences, the DLT accounts for the relative lack of difficulty associated with processing doubly nested structures like (6), in which the most embedded subject is a pronoun:

(6) The reporter who the senator who I met attacked disliked the editor.

Because referents for first- and second-person pronouns are already present in the current discourse, integrating across them consumes fewer resources than integrating across new discourse referents, according to the discourse-based DLT structural integration cost hypothesis. The point of maximal integration cost in (6) occurs at the point of processing the second verb *attacked*, just an in (1c), discussed earlier. However, the structural integration steps are less costly at this point in (6) than in (1c), because fewer new discourse referents are crossed in the integration steps in (6).[5] The distance-based DLT therefore provides a straightforward account of the observed contrast—a contrast not accounted for by earlier theories.

The DLT also explains a number of other nesting complexity contrasts not accounted for by earlier theories. For example, it accounts for the contrast between (1) embedding an RC within a complement clause (CC) of a noun and (2) the reverse embedding consisting of a CC within an RC:

(10) a. *Complement clause, then relative clause* (CC/RC)
The fact that the employee who the manager hired stole office supplies worried the executive.

b. *Relative clause, then complement clause* (RC/CC)
#The executive who the fact that the employee stole office supplies worried hired the manager.

The CC/RC embedding is much easier to understand than the RC/CC embedding (Cowper 1976; Gibson 1991, 1998; Gibson and Thomas 1998). The DLT integration hypothesis accounts for the CC/RC versus RC/CC contrast straightforwardly. The points of maximal integration cost in both structures occur at the verbs in each. In the CC/RC structure, the total integration cost for the most embedded verb *hired* is 3 EUs. The total integration cost for the verb *stole* is also 3 EUs, and the cost for the matrix verb *worried* is 6 EUs. In the RC/CC structure the total integration cost for the most embedded verb *stole* is only 1 EU, and the total integration cost at the matrix verb *hired* is 6 EUs, as in the CC/RC structure. It is at the point of processing the second verb *worried* in the RC/CC structure that the maximal integration cost occurs. The total integration cost at this point is 9 EUs, more than at any point in the CC/RC structure. The RC/CC structure is therefore more difficult to process than the CC/RC structure.

Another interesting nesting contrast accounted for by the DLT is that adding an extra new discourse referent at a nested location in a sentence leads to increased complexity. For example, (11b) is more complex than (11a) (Gibson and Nakatani 1998):

(11) a. The possibility that the administrator who the nurse supervised lost the medical reports didn't bother the intern from the maternity ward.
b. The possibility that the administrator who the nurse from the maternity ward supervised lost the medical reports didn't bother the intern.

In (11b) the prepositional phrase (PP) *from the maternity ward* is in a nested position, modifying the most embedded subject *the nurse*, whereas the same PP is in a nonnested position in (11a). According to the DLT, the points of highest integration cost in (11a) and (11b) are at the three verbal regions: *supervised*, *lost*, and *didn't bother*. These integration steps are all harder in (11b) than in (11a) because an extra new discourse referent intervenes in each case, and the contrast is accounted for.

5.3.3 Comparing DLT Integration Costs to Comprehension Times in Simple Relative Clause Structures

Although the DLT was initially developed to account for complexity effects in complex structures such as (1c) and (6), it accounts for processing effects in simple structures as well. One well-established complexity phenomenon to be explained by a computational resource theory is the greater complexity of an object-extracted RC as compared with a subject-extracted RC in a subject-verb-object language like English:

(12) *Subject extraction*
The reporter who sent the photographer to the editor hoped for a good story.

(13) *Object extraction*
The reporter who the photographer sent to the editor hoped for a good story.

In (12), the relative pronoun *who* is extracted from the subject position of the RC, whereas the same pronoun is extracted from the object position in (13). The object extraction is more complex by a number of measures, including phoneme monitoring, online lexical decision, reading times, and response accuracy to probe questions (Holmes 1973; Hakes, Evans, and Brannon 1976; Wanner and Maratsos 1978; Holmes and O'Regan 1981; Ford 1983; Waters, Caplan, and Hildebrandt 1987; King and Just 1991). In addition, the volume of blood flow in the brain is greater in language areas for object extractions than for subject extractions (Stromswold et al. 1996; Just et al. 1996), and aphasic stroke patients cannot reliably answer comprehension questions about object-extracted RCs, although they perform well on subject-extracted RCs (Caramazza and Zurif 1976; Caplan and Futter 1986; Grodzinsky 1989; Hickok, Zurif, and Canseco-Gonzalez 1993).

The DLT accounts for reading-time effects in these RC structures. Let us compare the integration costs predicted by the DLT at each word in (13) and (12) to actual reading times for participants in a self-paced reading experiment performed by Gibson and Ko (1998). The hypothesis is that DLT integration cost predicts reading times, when other factors such as temporary ambiguity, word length, and word frequency are controlled for. Again, this is an oversimplification even within the DLT, because reading times will also be affected by storage costs. But as long as storage costs are small, this simplification suffices. The word-by-word predictions of the DLT for the object-extracted RC structure in (13) are presented in table 5.1.

The DLT integration cost predictions for the object-extracted RC can be summarized as follows. Reading times are predicted to be fast for the first five words—*the reporter who the photographer*—in the object-extracted RC in (13), then slow on the embedded verb *sent*. Reading times should speed up again on the prepositional phrase *to the editor*, then slow down on the main verb of the sentence *hoped*, then speed up again on the final words *for a good story*.

A comparison between these predicted integration costs and actual reading times is presented in figure 5.3, based on data from thirty-two participants and sixteen items in a self-paced word-by-word reading experiment conducted by Gibson and Ko (1998). In this reading task, participants read sentences on a computer screen, at their own pace. At the beginning of a trial, a sentence is displayed on the screen with all nonspace characters replaced by dashes. When the participant presses the space bar, the first word of the sentence is displayed, replacing the corresponding dashes. When the participant presses the space bar a second time, the first word reverts to dashes,

Table 5.1
Word-by-word predictions of the DLT for the object-extracted RC structure in (13)

	Input word													
Cost type	The	reporter	who	the	photographer	sent	to	the	editor	hoped	for	a	good	story
New discourse referent	0	1	0	0	1	1	0	0	1	1	0	0	0	1
Structural integration	0	0	0	0	0	2	0	0	0	3	0	0	0	0
Total	0	1	0	0	1	3	0	0	1	4	0	0	0	1

Table 5.2
Word-by-word predictions of the DLT for the subject-extracted RC structure in (12)

	Input word													
Cost type	The	reporter	who	sent	the	photographer	to	the	editor	hoped	for	a	good	story
New discourse referent	0	1	0	1	0	1	0	0	1	1	0	0	0	1
Structural integration	0	0	0	0	0	0	0	0	0	3	0	0	0	0
Total	0	1	0	1	0	1	0	0	1	4	0	0	0	1

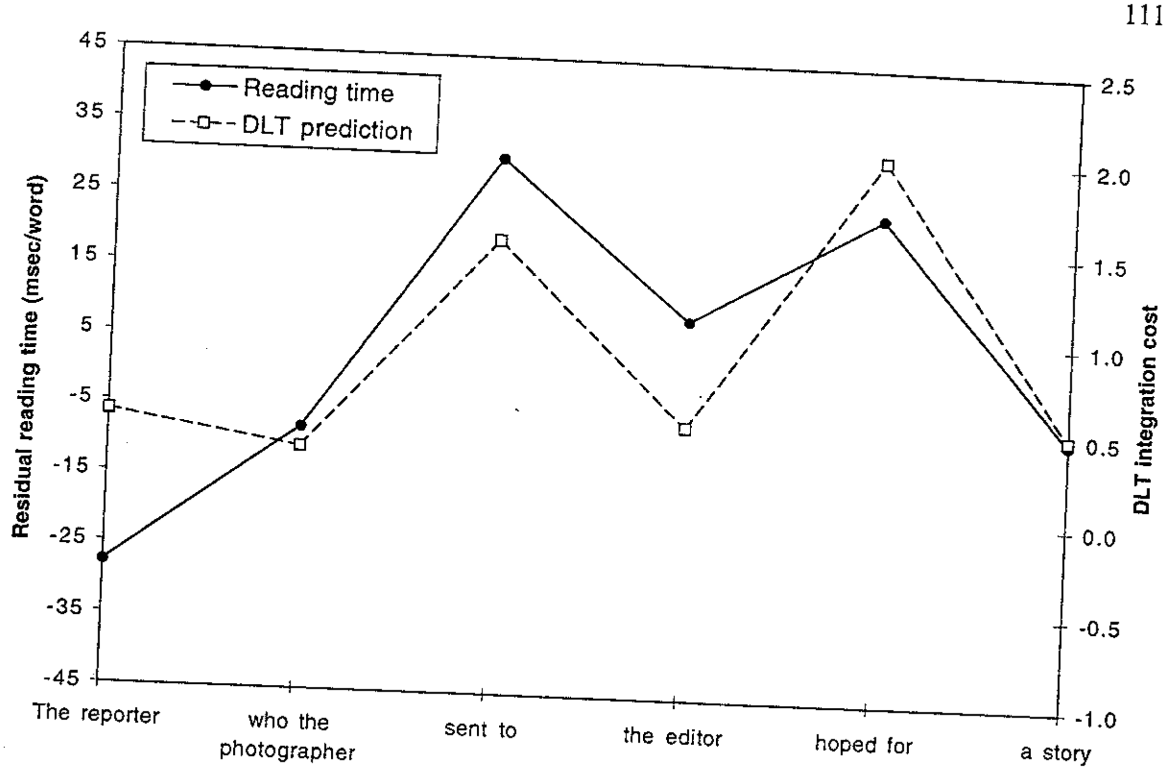

Figure 5.3
A comparison between residual reading times and locality-based integration costs in an object-extracted RC structure.

and the second word is displayed in place of the appropriate dashes. Each subsequent press of the space bar reveals the next word and removes the previous word. The computer records the time between each button-press, which represents the reading time for each word.

The reading-time data in the figure represent reading times normalized for length for each individual participant, computed by subtracting from raw reading times each participant's predicted time to read words of the same length, calculated by a linear regression equation across all sentences in the experiment (Ferreira and Clifton 1986). Thus a typical word will be read at 0 ms of normalized reading time (*residual reading time*), whereas words read quickly will have negative residual reading times, and words read slowly will have positive residual reading times. To reduce noise, the words are grouped in two- and three-word regions, and the average residual reading times are provided for each region. Because there is a reading-time spillover in self-paced reading, such that slow reading times are often reflected one or two words after a point of high complexity, locations of high predicted DLT integration cost were grouped with the following word. Other groupings were made according to constituent groupings in the remainder of the sentences. The reading times and DLT integration cost predictions are closely correlated for the object-extracted RC ($r = 0.79$, $r^2 = 0.63$, $F(1,4) = 6.85$, $p < 0.06$).

The word-by-word predictions of the DLT for the subject-extracted RC structure in (12) are presented in table 5.2.

The integration costs for this construction are the same as for the object extraction, except on the embedded verb *sent*. In the subject extraction, the cost of integrating the embedded verb *sent* is 1 EU, because although it indicates a new discourse referent, this is a local structural integration. Thus, in contrast to the object-extracted RC, reading times should be fast all the way through the subject-extracted RC until the main verb of the sentence—*hoped*—is encountered, at which point reading times should slow down. Reading times should then speed up again after this.

A comparison between the predicted integration costs and actual reading times for the subject-extracted RC structure is presented in figure 5.4, based on data from Gibson and Ko's (1998) self-paced reading experiment. As for the object-extracted RC structure, the reading times and DLT integration cost predictions are also closely correlated for the subject-extracted RC structure ($r = 0.77$, $r^2 = 0.60$, $F(1,4) = 5.99$, $p < 0.08$). Combining the subject- and object-extracted RC data yields a significant

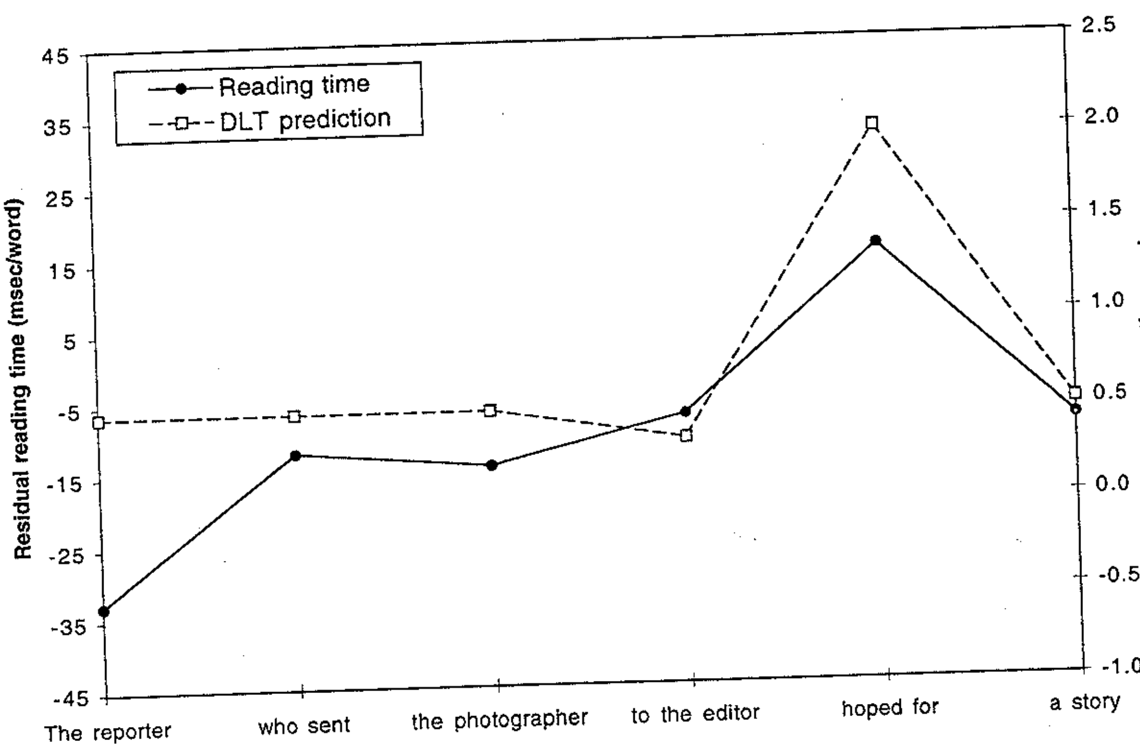

Figure 5.4
A comparison between residual reading times and locality-based integration costs in a subject-extracted RC structure.

correlation between reading times and DLT predictions, accounting for almost 60% of the variance in the RTs ($r = 0.77$, $r^2 = 0.59$, $F(1,4) = 14.7$, $p < 0.005$). Related reading-time data are provided by King and Just (1991), as well as in another experiment performed by Gibson and Ko (1998) and reported in Gibson (1998).

5.3.4 Crosslinguistic Support for the DLT

The evidence provided thus far in support of the DLT integration cost hypothesis has all come from the processing of English. However, the claims being made are about the nature of computational resource use in human language processing more generally, not just in English. Evidence from the processing of other languages is therefore crucial to the enterprise. Some Japanese evidence in support of the DLT is provided by Babyonyshev and Gibson (1999). They had participants rate the processing difficulty of a number of different structures, including the following two:

(14) Obasan-wa [bebiisitaa-ga [ani-ga naita] to itta] to omotteiru.
aunt-top babysitter-nom older-brother-nom cried that said that thinks
My aunt thinks that the babysitter said that my older brother cried.

(15) Obasan-wa [bebiisitaa-ga [ani-ga imooto-o ijimeta] to itta] to omotteiru.
aunt-top babysitter-nom older-brother-nom younger-sister-acc bullied that said that thinks
My aunt thinks that the babysitter said that my older brother bullied my younger sister.

Both (14) and (15) are doubly nested clausal structures. The difference between the two is that the most embedded clause in (14) is intransitive, containing a subject and a verb (*ani-ga naita* 'brother cried'), whereas the most embedded clause in (15) is transitive, containing a subject, an object, and a verb (*ani-ga imooto-o ijimeta* 'brother bullied sister'). The inclusion of the extra object NP, a new discourse referent, increases the distances between the subjects and the verbs for all three verbs, just as adding the prepositional phrase did so for the English examples discussed earlier (Gibson and Nakatani 1998). Thus the DLT predicts larger structural integration costs on each of the verbs, and correspondingly worse complexity ratings for the transitive structure in (15). The results from Babyonyshev and Gibson's (1999) difficulty rating experiment confirmed this prediction: the transitive structures were rated as significantly harder to understand than the intransitive structures.

Babyonyshev and Gibson present further evidence in support of the DLT from the processing of other Japanese constructions. Other crosslinguistic evidence relevant to the processing of unambiguous constructions is provided by Bach, Brown, and Marslen-Wilson (1986), who investigated nesting complexity in Dutch and German.

(See Gibson 1998 for a demonstration that this evidence supports the DLT. Much additional crosslinguistic support of the DLT with respect to ambiguity resolution is also described in Gibson 1998. Languages for which some evidence is described there include Spanish, German, Dutch, and Finnish.)

5.3.5 The DLT Storage Cost Component

I have concentrated on the integration cost component of the DLT so far. There is a second component of the theory: the storage cost component. According to the storage cost component, each syntactic head required to complete the current input string as a grammatical sentence is associated with a storage cost.[6] Under most syntactic theories (e.g., Bresnan 1982; Chomsky 1981, 1995; Pollard and Sag 1994) the minimal number of syntactic head categories in an English sentence is two: a noun for the subject and a verb for the predicate. The DLT storage cost hypothesis is given in (16):

(16) *DLT storage cost*
1 memory unit (MU) is associated with each syntactic head required to complete the current input as a grammatical sentence.

Consider (16) with respect to the object-extracted RC structure in (17):

(17)

	Input word								
	The	reporter	who	the	senator	attacked	disliked	the	editor
Storage cost (in MUs)	2	1	3	4	3	1	1	1	0

At the point of processing the sentence-initial determiner *the*, two syntactic heads are needed to form a grammatical sentence: a noun and a verb. There is therefore a cost of 2 MUs at this point. After processing *reporter*, only one head is needed to complete a grammatical sentence: a verb. The storage cost is therefore 1 MU here. When the pronoun *who* is processed, the rest of an RC must follow, in addition to the main verb of the sentence. The RC requires two more heads: a verb, and an empty category position in the RC to be associated with the RC pronoun *who*. Thus the total storage cost at this point is 3 MUs. For example, the sentence could be completed as *The reporter who slept left*. In this continuation, the three heads that are needed at the point of processing *who* end up being (1) the verb *slept*; (2) an empty category in subject position of this verb, which refers to the same individual as *who*; and (3) the verb *left*.

After processing the second instance of *the* (following *who*), four heads are needed to make a grammatical sentence: two verbs, an empty category position in the RC to be associated with the RC pronoun *who*, and noun for the determiner *the*. The noun *senator* satisfies the last of these requirements, leaving a cost of 3 MUs at this point. The verb *attacked* then satisfies the prediction for a verb in the RC, and an empty category can also be connected at this point, licensed in the object position of the verb *attacked*. The storage cost at *attacked* is therefore only 1 MU, corresponding to the prediction of the main verb. The main verb *disliked* is encountered next, satisfying this prediction. However, this verb requires an NP object to its right, resulting in a cost of 1 MU. The determiner *the* does not satisfy the prediction for a noun, but it does not add any additional predictions either, so the storage cost remains at 1 MU here. Finally, the noun *editor* is attached at the noun object of the verb *disliked*, completing a grammatical sentence. There is therefore no storage cost on completion of the sentence.

Now that the storage and integration cost components of the DLT have been specified, it is necessary to say how the two parts of the theory interrelate to provide a theory of comprehension times and intuitive complexity. The set of assumptions made by Gibson (1998) with respect to this issue are as follows: (1) integrations and storage access the same pool of resources (Just and Carpenter 1992; see Caplan and Waters 1999); (2) there is a fixed capacity of resources in the resource pool; and (3) each predicted syntactic head takes up a fixed quantity of resources. As a result, the more resources that are required in storage, the slower integrations occur.

It is also possible that storage costs might not all consume a fixed quantity of resources. Alternatively, fewer resources might by used for each additional syntactic prediction stored as more predictions are stored. For example, 1 storage unit might be used to store one prediction, but only 1.5 units to store two predictions, and only 1.75 units for three. The result of such a system might be that, as more predictions are stored, some are stored less well. Consequently, the likelihood that all predictions will be recalled may decrease as more are stored.

There are of course many other possibilities for the relationship between storage and integration cost. Currently, there is no empirical data that decides among these possibilities. We will therefore not make a commitment here, beyond assuming that larger quantities of either storage or integration cost cause slower integration times, and that intuitive complexity is determined by the maximal integration time in the parse of a sentence.

5.3.6 Applying the DLT to Ambiguity Resolution

The evidence put forward thus far in support of the DLT has come from the processing of unambiguous structures. This section considers some predictions that the DLT makes when used as part of a metric in ambiguity resolution. The DLT ambiguity resolution claim is given in (18):

(18) *Ambiguity resolution hypothesis*
In choosing among ambiguous structures, two of the factors that the processor uses to evaluate its choices are DLT storage and structural integration cost (in

addition to informational constraints, such as lexical frequencies, plausibility, and context).

Consider the ambiguity resolution hypothesis with respect to an ambiguous example like (19):

(19) The bartender told the detective that the suspect left the country yesterday.

The adverb *yesterday* can be linked to either the local verb *left* or to the more distant verb *told*. There is no memory cost difference between the two resultant structures, but there is a structural integration cost difference: the attachment to *left* crosses only one new discourse referent, whereas the attachment to *told* crosses four new discourse referents. Thus the local integration to *left* is strongly preferred. (For earlier accounts of locality preferences in ambiguity resolution, see Kimball 1973; Frazier and Fodor 1978; Gibson 1991; Stevenson 1994; Gibson et al. 1996.[7] See Gibson 1998 for discussion of further evidence of locality preferences in ambiguity resolution.)

Consider now the DLT ambiguity resolution hypothesis with respect to the ambiguity in (20):

(20) The evidence examined by the lawyer turned out to be unreliable.

There is a temporary ambiguity at the word *examined* between a past-tense main verb (MV) interpretation and a past-participle reduced relative (RR) interpretation. The past-tense and the past-participle readings are roughly equally frequent in the Brown corpus (Francis and Kučera 1982). Plausibility strongly favors the RR structure, because it is plausible for evidence to be examined, but it is not plausible for evidence to examine something (Ferreira and Clifton 1986; Trueswell, Tanenhaus, and Garnsey 1994). There is no structural integration cost difference between the two potential attachments, because both are local. There is potentially a small memory cost difference between the two, favoring the MV structure. In particular, one syntactic head is needed to complete the MV structure as a grammatical sentence: a noun in the object position of the transitive verb *examined*. One or possibly two syntactic heads are needed to complete the RR structure as a grammatical sentence: (1) the main verb of the sentence, and possibly (2) a modifier for the reduced relative clause, because single-word reduced relative clauses are not very acceptable. Thus the DLT memory costs are either balanced for this ambiguity or slightly favoring the MV structure. Weighing all the applicable factors together, the RR structure is preferred, primarily because of the large plausibility difference (Trueswell, Tanenhaus, and Garnsey 1994).

In contrast to the ambiguity resolution hypothesis proposed here, in which resource use is a factor, recent lexicalist "constraint-based" processing theories claim that ambiguity resolution is determined exclusively by lexical frequencies and plausibility (MacDonald, Pearlmutter, and Seidenberg 1994; Trueswell, Tanenhaus, and Garnsey 1994), with no effect of resource use. Evidence for this claim comes from a number of reading-time studies of a few types of English temporary ambiguities, including the MV/RR ambiguity. However, the ambiguities on which this claim is based are all similar to the MV/RR ambiguity in that there is minimal or no resource use difference between the potential structures in the temporary ambiguities in question. It is therefore dubious to conclude that resource use plays no role in ambiguity resolution based on these ambiguities, because resource use does not differ very much in these cases.[8]

To test the resource use hypothesis, it is necessary to compare the resolution of an ambiguity with a small resource complexity difference with the resolution of ambiguities with larger resource complexity differences, while controlling for lexical frequency and plausibility. Gibson, Grodner, and Tunstall (1997) did exactly this by exploring the MV/RR structure embedded within a relative clause, as in (21):

(21) The witness who the evidence examined by the lawyer implicated seemed to be very nervous.

The items were constructed using Trueswell, Tanenhaus, and Garnsey's MV/RR items as a base, so that plausibility factors highly favored the RR interpretation of the ambiguous verb *examined*. However, syntactic complexity as measured by the DLT storage cost strongly favors the MV reading in this variant of the MV/RR ambiguity. In particular, only one syntactic head is required to complete the implausible MV structure as a grammatical sentence: the main verb of the sentence. In contrast, three or four syntactic heads are needed to complete the RR structure grammatically: (1) the main verb of the sentence, (2) a verb for the relative clause, (3) an empty NP to be associated with the RC pronoun *who*, and possibly (4) an adverbial modifier for the single-word reduced relative clause. The storage cost difference between the two structures is therefore two or three syntactic heads, two more than in the MV/RR ambiguities explored earlier. This difference is much larger than in any of the ambiguities that motivated the lexically based ambiguity resolution hypothesis.[9] If the DLT storage cost is being used in ambiguity resolution online, people should have more difficulty following the RR reading of the verb *examined* in (21) than they will in a control case like (20). To test this hypothesis, Gibson and colleagues compared the reading times of sentences like (21) to closely related examples like (23a), with a small storage complexity difference between the MV and RR readings. They also tested unambiguous control sentences for each type of structure, as in (22) and (23b) respectively.

(22) *Large storage cost difference, unambiguous control*
The witness who the evidence that was examined by the lawyer implicated seemed to be very nervous.

(23) a. *Small storage cost difference, ambiguous*
The witness thought that the evidence examined by the lawyer implicated his next-door neighbor.

b. *Small storage cost difference, unambiguous*
The witness thought that the evidence that was examined by the lawyer implicated his next-door neighbor.

As in the simple MV/RR ambiguities in (20), the storage complexity at the point of processing *examined* in (23a) is one syntactic prediction for the MV structure and one or two syntactic predictions for the RR structure, resulting in a smaller storage cost difference for this MV/RR structure.

The participants' residual reading times for each of these conditions are plotted in figure 5.5. The first result of interest to the DLT is that reading times for the region *the evidence examined by the lawyer* were significantly faster in the small storage cost conditions than in the large storage cost conditions. Although the same structural integrations are being performed in all these conditions, there is a larger syntactic storage load in the large storage conditions than in the small storage conditions, leading to longer reading times in these conditions, as predicted.

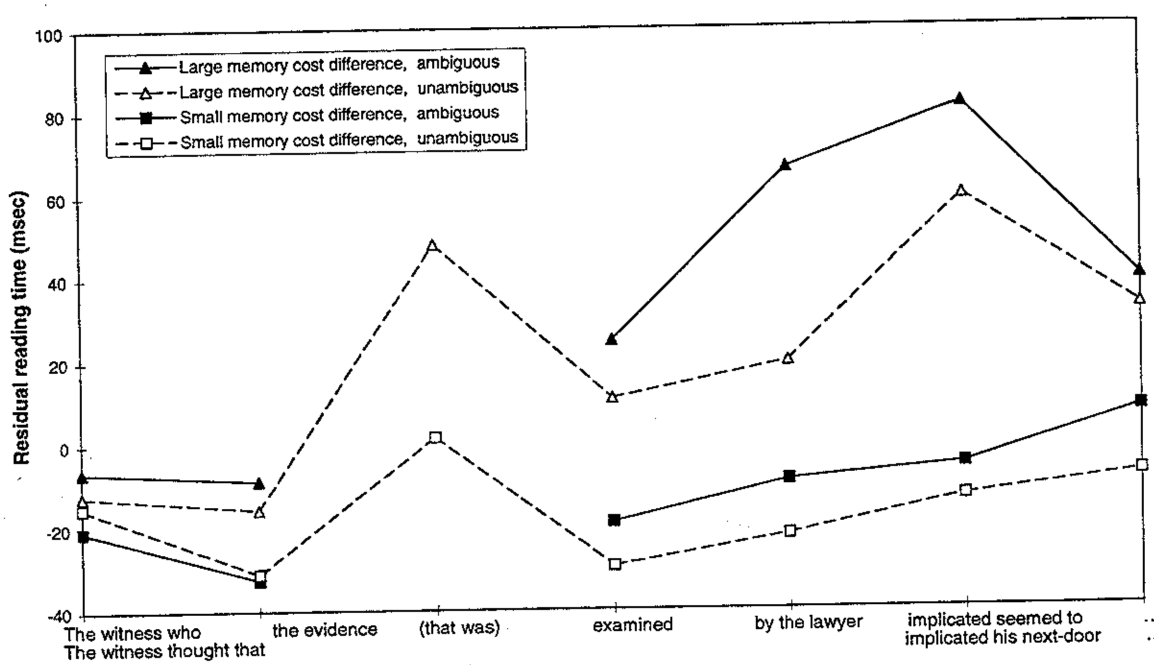

Figure 5.5
Residual reading times for sixty subjects taking part in a self-paced, word-by-word, moving-window reading experiment involving four conditions that crossed DLT storage complexity difference (high storage complexity difference, low storage complexity difference) with ambiguity (ambiguous, unambiguous).

Second, the experimental participants read the disambiguating prepositional phrase more slowly in the ambiguous large storage condition than in the disambiguated large storage condition, as predicted by the DLT applied to ambiguity resolution. Crucially, there was also a significant interaction of storage cost difference (large, small) and ambiguity (ambiguous, unambiguous) in the disambiguating PP, such that the reanalysis effect was significantly larger in the large storage conditions than in the small storage conditions (where there was a numerical but nonsignificant difference in reading times). This pattern of results is consistent with the hypothesis that people initially follow the main-verb reading in (21), because of its much lower syntactic complexity and in spite of its implausibility. On the other hand, people use the plausibility information to perform the disambiguation in (23a), because the syntactic complexity does not strongly favor either interpretation. Thus syntactic storage cost complexity as measured by the DLT appears to be an independent factor involved in ambiguity resolution that is not reducible to frequency and plausibility.

5.4 Summary and Conclusions

A theory of linguistic complexity has been proposed here that associates (1) increasing structural integration cost with the distance of attachment, and (2) storage cost with predicted syntactic categories. This theory—the dependency locality theory—provides a unified theory of a large array of disparate processing phenomena, including the following:

1. Online reading times of subject- and object-extracted relative clauses
2. The complexity of doubly nested relative clause constructions
3. The greater complexity of embedding a sentential complement within a relative clause than the reverse embedding in both English and Japanese
4. The lower complexity of multiply embedded structures with pronominals in the most embedded subject position in both English and Japanese
5. The high complexity of certain two-clause constructions
6. The greater complexity of nesting clauses with more arguments in Japanese
7. Ambiguity effects—syntactic complexity effects independent of plausibility and frequency
8. Numerous other effects not discussed here (see Gibson 1998), including:
   a. Heaviness effects (Bever 1970; Hawkins 1990, 1994; Wasow 1997)
   b. The greater complexity of center-embedded constructions as compared with cross-serial constructions (Bach, Brown, and Marslen-Wilson 1986)
   c. Gap-positing preferences in temporarily ambiguous structures
   d. Argument-attachment preferences in temporarily ambiguous structures

Understanding how the language comprehension mechanism uses computational resources will continue to be a fundamental area of computational psycholinguistic research in the coming years, as it has been over the past forty years.

Notes

I would like to thank the participants at the conference in addition to the following people for helpful comments on earlier drafts of this work: Dan Grodner, Carson Schütze, and Tessa Warren. Any remaining errors are my own. Funding for some of the work reported here was provided by NSF grant SBR-9729037 "Empirical Investigations of Locality Effects in Linguistic Complexity" and by the MIT/JST joint international Mind Articulation Project.

1. Although the phrase structure hypotheses implicit here are standard across most syntactic theories (Chomsky 1981, 1995; Bresnan 1982; Pollard and Sag 1994), some of the assumptions are less universally accepted. For example, the inventory of categories and their relationship to one another are debatable (Pollard and Sag 1994; Chomsky 1995; Steedman 1996), as is the implicit claim that there exist empty-category positions mediating long distance dependencies (Pickering and Barry 1991; Steedman 1996). The specific assumptions made with respect to these controversial issues are for convenience only, so that we have a consistent notation to discuss sentence meaning.

2. Sentences that cause extreme processing difficulty are prefixed with the symbol #.

3. One factor that can contribute to the processing complexity of nested structures but that is orthogonal to the factors to be investigated here is semantic similarity. Center-embedded RC structures like (1c) are easier to comprehend if the NPs come from distinct semantic classes and if the roles assigned by the following verbs are also compatible with distinct semantic classes, so that it is easy to guess who is doing what to whom (Stolz 1967; Schlesinger 1968; King and Just 1991). For example, (i) is easier to comprehend than (1c) (Stolz 1967):

(i) ?#The vase that the maid that the agency hired dropped on the floor broke into a hundred pieces.

Although semantic-role disambiguation improves the acceptability of these kinds of structures, a complexity theory based on semantic role interference alone is insufficient to explain many complexity effects. For example, although (i) is easier to comprehend than (1c), it is still very complex, and this complexity needs to be accounted for. Furthermore, including an additional pragmatically distinguishable nested RC makes the structure virtually unprocessable, similar to or more complex than (1c):

(ii) #The vase that the maid that the agency that the lawyer represented hired dropped on the floor broke into a hundred pieces.

Hence factors other than semantic similarity or interference are responsible for the complexity of nested structures like (ii).

4. A class of models of nesting complexity that will not be discussed here are connectionist models (e.g., Kempen and Vosse 1989; Elman 1991; Weckerly and Elman 1992; Miikkulainen 1996; Christiansen and Chater 1999). The goal for these models is to have the complexity phenomena fall out from the architecture of the processor. This kind of model, with a basis in neural architecture, may eventually provide an architectural explanation of the approach proposed here. However, because these types of models are still quite novel, they have not yet been applied to a wide range of phenomena across languages.

5. The maximal integration cost for (6) is 5 EUs, as compared to 7 EUs for (1c).

6. This hypothesis is very similar to the incomplete dependency hypothesis outlined in section 5.2. There is currently no empirical evidence relevant to deciding between indexing storage cost difficulty in terms of predicted categories or incomplete dependencies: either hypothesis suffices for the data that we know of thus far. Experiments are being run that will help decide between these and other possibilities.

7. Factors other than locality also affect modifier attachment preferences crosslinguistically. See Cuetos and Mitchell 1988; Frazier and Clifton 1996; Gibson et al. 1996; Gibson, Pearlmutter, and Torrens 1999; and Hemforth, Konieczny, and Scheepers, forthcoming, for evidence and theories about what other factors affect modifier attachment preferences.

8. To be fair, MacDonald and colleagues' and Trueswell and associates' claims center around demonstrating that a particular structure-based ambiguity resolution hypothesis, Minimal Attachment (Frazier 1978), is incorrect. They are less explicit about ruling out other possible syntactic complexity metrics.

9. It should be noted that although storage cost favors the MV structure in this ambiguity, structural integration cost actually favors the RR structure, because there is an extra integration between the object position of *examined* and the pronoun *who* in the MV structure. Gibson (1998) proposes that, in conflicts between minimizing storage and minimizing structural integration cost, storage cost is minimized. The motivation for this assumption is that storage cost is effectively potential structural integration cost—structural integration cost that will be expended later—and processing additional material cannot lower this cost, only increase it, leading to larger structural integration costs downstream. Thus by minimizing storage cost, the parser minimizes structural integration costs over the course of the sentence.

References

Abney, S. P., and M. Johnson. 1991. Memory requirements and local ambiguities of parsing strategies. *Journal of Psycholinguistic Research* 20:233–250.

Altmann, G., and M. Steedman. 1988. Interaction with context during human sentence processing. *Cognition* 30:191–238.

Anderson, J. R. 1994. *Learning and memory*. New York: Wiley.

Babyonyshev, M., and E. Gibson. 1999. The complexity of nested structures in Japanese. *Language* 75:423–450.

Bach, E., C. Brown, and W. Marslen-Wilson. 1986. Crossed and nested dependencies in German and Dutch: A psycholinguistic study. *Language and Cognitive Processes* 1:249–262.

Baddeley, A. 1990. *Human memory*. Needham Heights, Mass.: Allyn and Bacon.

Bever, T. G. 1970. The cognitive basis for linguistic structures. In *Cognition and the development of language*, ed. J. R. Hayes, 279–362. New York: Wiley.

Bever, T. G. 1974. The ascent of the specious, or there's a lot we don't know about mirrors. In *Explaining linguistic phenomena*, ed. D. Cohen, 173–200. Washington, D.C.: Hemisphere.

Bresnan, J. 1982. *The mental representation of grammatical relations*. Cambridge, Mass.: MIT Press.

Caplan, D., and C. Futter. 1986. Assignment of thematic roles by an agrammatic aphasic patient. *Brain and Language* 27:117–135.

Caplan D., and G. S. Waters. 1999. Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences* 22:77–126.

Caramazza, A., and E. Zurif. 1976. Dissociation of algorithmic and heuristic processes in sentence comprehension: Evidence from aphasia. *Brain and Language* 3:572–582.

Chomsky, N. 1957. *Syntactic structures*. The Hague: Mouton.

Chomsky, N. 1965. *Aspects of the theory of syntax*. Cambridge, Mass.: MIT Press.

Chomsky, N. 1981. *Lectures on government and binding*. Dordrecht: Foris.

Chomsky, N. 1995. *The minimalist program*. Cambridge, Mass.: MIT Press.

Chomsky, N., and G. A. Miller. 1963. Introduction to the formal analysis of natural languages. In *Handbook of mathematical psychology*, vol. 2, ed. R. D. Luce, R. R. Bush, and E. Galanter, 269–321. New York: Wiley.

Christiansen, M. H., and N. Chater. 1999. Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science* 23:157–205.

Cowper, E. A. 1976. Constraints on sentence complexity: A model for syntactic processing. Ph.D. dissertation, Brown University, Providence, R.I.

Crain, S., and M. Steedman. 1985. On not being led up the garden path: The use of context by the psychological parser. In *Natural language processing: Psychological, computational, and theoretical perspectives*, ed. D. Dowty, L. Karttunen, and A. Zwicky, 320–358. Cambridge, England: Cambridge University Press.

Cuetos, F., and D. C. Mitchell. 1988. Cross-linguistic differences in parsing: Restrictions on the use of the late closure strategy in Spanish. *Cognition* 30:73–105.

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

Ferreira, F., and C. Clifton, Jr. 1986. The independence of syntactic processing. *Journal of Memory and Language* 25:348–368.

Ford, M. 1983. A method for obtaining measures of local parsing complexity throughout sentences. *Journal of Verbal Learning and Verbal Behavior* 22:203–218.

Ford, M., J. Bresnan, and R. Kaplan 1982. A competence-based theory of syntactic closure. In *The mental representation of grammatical relations*, ed. J. Bresnan, 727–796. Cambridge, Mass.: MIT Press.

Francis, W. N., and H. Kučera. 1982. *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin.

Frank, R. 1992. Syntactic locality and tree-adjoining grammar: Grammatical, acquisition, and processing perspectives. Ph.D. dissertation, University of Pennsylvania, Philadelphia.

Frazier, L. 1978. On comprehending sentences: Syntactic parsing strategies. Ph.D. dissertation, University of Connecticut, Storrs.

Frazier, L., and C. Clifton, Jr. 1996. *Construal*. Cambridge, Mass.: MIT Press.

Frazier, L., and J. D. Fodor. 1978. The Sausage Machine: A new two-stage parsing model. *Cognition* 6:291–325.

Garrod, S., and A. J. Sanford. 1977. Interpreting anaphoric relations: The integration of semantic information while reading. *Journal of Verbal Learning and Verbal Behavior* 16:77–90.

Garrod, S. C., and A. J. Sanford. 1982. The mental representation of discourse in a focused memory system: Implications for the interpretation of anaphoric noun phrases. *Journal of Semantics* 1:21–41.

Garrod, S. C., and A. J. Sanford. 1994. Resolving sentences in a discourse context: How discourse representation affects language understanding. In *Handbook of psycholinguistics*, ed. M. A. Gernsbacher, 675–698. San Diego: Academic Press.

Gibson, E. 1991. A computational theory of human linguistic processing: Memory limitations and processing breakdown. Ph.D. dissertation, Carnegie Mellon University, Pittsburgh.

Gibson, E. 1998. Linguistic complexity: Locality of syntactic dependencies. *Cognition* 68:1–76.

Gibson, E., M. Babyonyshev, and E. Kaan. 1998. Sentence reanalysis: Generalized incremental processing within a parallel processing framework. Poster presented at the Eleventh CUNY Sentence Processing Conference, Rutgers, N.J.

Gibson, E., D. Grodner, and S. Tunstall. 1997. Evidence for a syntactic complexity metric in ambiguity resolution independent of lexical effects. Poster presented at the Tenth CUNY Sentence Processing Conference, Santa Monica, Calif.

Gibson, E., and K. Ko. 1998. An integration-based theory of computational resources in sentence comprehension. Paper presented at the Fourth Architectures and Mechanisms in Language Processing Conference, University of Freiburg, Germany.

Gibson, E., and K. Nakatani. 1998. The lexical and locality basis for linguistic complexity. Ms., MIT, Cambridge, Mass.

Gibson, E., and N. Pearlmutter. 1998. Constraints on sentence comprehension. *Trends in Cognitive Science* 2:262–268.

Gibson, E., N. Pearlmutter, E. Canseco-Gonzales, and G. Hickok. 1996. Recency preference in the human sentence processing mechanism. *Cognition* 59:23–59.

Gibson, E., N. Pearlmutter, and V. Torrens, V. 1999. Recency and lexical preferences in Spanish. *Memory and Cognition* 27:603–611.

Gibson, E., and J. Thomas. 1998. The complexity of English nested structures. Ms., MIT, Cambridge, Mass.

Grodzinsky, Y. 1989. Agrammatic comprehension of relative clauses. *Brain and Language* 31:480–499.

Hakes, B., J. Evans, and L. Brannon. 1976. Understanding sentences with relative clauses. *Memory and Cognition* 4:283–296.

Hakuta, K. 1981. Grammatical description versus configurational arrangement in language acquisition: The case of relative clauses in Japanese. *Cognition* 9:197–236.

Haliday, M. A. K., and R. Hasan. 1976. *Cohesion in English*. London: Longman.

Harris, A. 1998. Electrophysiological indices of syntactic processing difficulty. Ph.D. dissertation, MIT, Cambridge, Mass.

Haviland, S. E., and H. H. Clark. 1974. What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior* 13:512–521.

Hawkins, J. A. 1990. A parsing theory of word order universals. *Linguistic Inquiry* 21:223–262.

Hawkins, J. A. 1994. *A performance theory of order and constituency*. Cambridge, England: Cambridge University Press.

Heim, I. 1982. The semantics of definite and indefinite noun phrases. Ph.D. dissertation, University of Massachusetts, Amherst.

Hemforth, B., L. Konieczny, and C. Scheepers. Forthcoming. Syntactic attachment and anaphor resolution: Two sides of relative clause attachment. In *Architectures and mechanisms for language processing*, ed. M. Crocker, M. Pickering, and C. Clifton, Jr. Cambridge, England: Cambridge University Press.

Hickok, G., E. Zurif, and E. Canseco-Gonzalez. 1993. Structural description of agrammatic comprehension. *Brain and Language* 45:371–395.

Holmes, V. 1973. Order of main and subordinate clauses in sentence perception. *Journal of Verbal Learning and Verbal Behavior* 12:285–293.

Holmes, V. M., and J. K. O'Regan. 1981. Eye fixation patterns during the reading of relative clause sentences. *Journal of Verbal Learning and Verbal Behavior* 20:417–430.

Joshi, A. K. 1990. Processing crossed and nested dependencies: An automaton perspective on the psycholinguistic results. *Language and Cognitive Processes* 5:1–27.

Just, M. A., and P. A. Carpenter. 1992. A capacity theory of comprehension: Individual differences in working memory. *Psychological Review* 99:122–149.

Kaan, E., A. Harris, E. Gibson, and P. Holcomb. 1998. The P600 as an index of syntactic integration difficulty. Paper presented at the Fourth Architectures and Mechanisms in Language Processing Conference, University of Freiburg, Germany.

Kac, M. B. 1981. Center-embedding revisited. In *Proceedings of the Third Annual Conference of the Cognitive Science Society*, 123–124. Hillsdale, N.J.: Erlbaum.

Kamp, H. 1981. A theory of truth and semantic representations. In *Formal methods in the study of language*, ed. J. Groenendijk, T. M. V. Janssen, and M. Stokhof. Amsterdam: Mathematisch Centrum.

Kempen, G., and T. Vosse. 1989. Incremental syntactic tree formation in human sentence processing: A cognitive architecture based on activation decay and simulated annealing. *Connection Science* 1:273–290.

Kimball, J. 1973. Seven principles of surface structure parsing in natural language. *Cognition* 2:15–47.

King, J., and M. A. Just. 1991. Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language* 30:580–602.

Lewis, R. 1993. An architecturally-based theory of human sentence comprehension. Ph.D. dissertation, Carnegie Mellon University, Pittsburgh.

Lewis, R. 1996. A theory of grammatical but unacceptable embeddings. *Journal of Psycholinguistic Research* 25:93–116.

MacDonald, M. 1997. Lexical representations and sentence processing: An introduction. *Language and Cognitive Processes* 12:121–136.

MacDonald, M., N. Pearlmutter, and M. Seidenberg. 1994. The lexical nature of syntactic ambiguity resolution. *Psychological Review* 101:676–703.

MacWhinney, B. 1987. The Competition Model. In *Mechanisms of language acquisition*, ed. B. MacWhinney, 249–308. Hillsdale, N.J.: Erlbaum.

McClelland, J. L., M. St. John, and R. Taraban. 1989. Sentence comprehension: A parallel distributed processing approach. *Language and Cognitive Processes* 4:SI287–335.

McElree, B., and T. Griffith. 1995. Syntactic and thematic processing in sentence comprehension: Evidence for a temporal dissociation. *Journal of Experimental Psychology: Learning Memory, and Cognition* 21:134–157.

Miikkulainen, R. 1996. Subsymbolic case-role analysis of sentences with embedded clauses. *Cognitive Science* 20:47–73.

Miller, G. A., and N. Chomsky. 1963. Finitary models of language users. In *Handbook of mathematical psychology*, vol. 2, ed. R. D. Luce, R. R. Bush, and E. Galanter, 419–491. New York: Wiley.

Miller, G. A., and S. Isard. 1964. Free recall of self-embedded English sentences. *Information and Control* 7:292–303.

Ni, W., S. Crain, and D. Shankweiler. 1996. Sidestepping garden-paths: Assessing the contributions of syntax, semantics, and plausibility in resolving ambiguities. *Language and Cognitive Processes* 11:283–334.

Pearlmutter, N., and M. MacDonald. 1992. Plausibility and syntactic ambiguity resolution. *Proceedings of the Fourteenth Conference of the Cognitive Science Society*, 498–503. Hillsdale, N.J.: Erlbaum.

Pickering, M., and G. Barry. 1991. Sentence processing without empty categories. *Language and Cognitive Processes* 6:229–259.

Pollard, C., and I. A. Sag. 1994. *Head-driven phrase structure grammar*. Chicago: University of Chicago Press.

Rambow, O., and A. Joshi. 1994. A processing model for free word-order languages. In *Perspectives on sentence processing*, ed. C. Clifton Jr., L. Frazier, and K. Rayner, 267–301. Hillsdale, N.J.: Erlbaum.

Schlesinger, I. M. 1968. *Sentence structure and the reading process*. The Hague: Mouton.

Stabler, E. P. 1994. The finite connectivity of linguistic structures. In *Perspectives in sentence processing*, ed. C. Clifton, Jr., L. Frazier, and K. Rayner, 303–336. Hillsdale, N.J.: Erlbaum.

Steedman, M. 1996. *Surface structure and interpretation*. Cambridge, Mass.: MIT Press.

Stevenson, S. 1994. Competition and recency in a hybrid network model of syntactic disambiguation. *Journal of Psycholinguistic Research* 23:295–322.

Stolz, W. S. 1967. A study of the ability to decode grammatically novel sentences. *Journal of Verbal Learning and Verbal Behavior* 6:867–873.

Stromswold, K., D. Caplan, N. Alpert, and S. Rauch. 1996. Localization of syntactic comprehension by positron emission tomography. *Brain and Language* 52:452–473.

Tanenhaus, M. K., and J. C. Trueswell. 1995. Sentence comprehension. In *Speech, Language, and Communication*, ed. J. Miller and P. Eimas, 217–262. San Diego: Academic Press.

Trueswell, J. C. 1996. The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language* 35:566–585.

Trueswell, J. C., M. K. Tanenhaus. and S. M. Garnsey. 1994. Semantic influences on parsing: Use of thematic role information in syntactic disambiguation. *Journal of Memory and Language* 33:285–318.

Tyler, L. K., and W. D. Marslen-Wilson. 1977. The on-line effects of semantic context on syntactic processing. *Journal of Verbal Learning and Verbal Behavior* 16:683–692.

Wanner, E., and M. Maratsos. 1978. An ATN approach in comprehension. In *Linguistic theory and psychological reality*, ed. M. Halle, J. Bresnan, and G. Miller, 119–161. Cambridge, Mass.: MIT Press.

Warren, T., and E. Gibson. 1999. The effects of discourse status on intuitive complexity: Implications for quantifying distance in a locality-based theory of linguistic complexity. Poster presented at the Twelfth CUNY Sentence Processing Conference, CUNY, New York.

Wasow, T. 1997. Remarks on grammatical weight. *Language Variation and Change* 9:81–105.

Waters, G. S., D. Caplan, and N. Hildebrandt. 1987. Working memory and written sentence comprehension. In *Attention and performance*, vol. 12: *The psychology of reading*, ed. M. Coltheart, 531–555. Hillsdale, N.J.: Erlbaum.

Webber, B. L. 1988. Tense as discourse anaphor. *Computational Linguistics* 14:61–73.

Weckerly, J., and J. L. Elman. 1992. A PDP approach to processing center-embedded sentences. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*, 414–419. Hillsdale, N.J.: Erlbaum.

Yngve, V. H. 1960. A model and an hypothesis for language structure. *Proceedings of the American Philosophical Society* 104:444–466.