Journal of Experimental Psychology:
Learning, Memory, and Cognition
1995, Vol. 21, No. 1, 134–157

0278-7393/95/$3.00

Syntactic and Thematic Processing in Sentence Comprehension: Evidence for a Temporal Dissociation

Brian McElree and Teresa Griffith
University of California, Irvine

The time courses for processing constituent structure relations, subcategorization restrictions, and thematic role relations during sentence comprehension were measured with reaction time and speed–accuracy trade-off variants of a grammaticality judgment task. Thematic role processing was found to be delayed by as much as 100 ms relative to the time when constituent structure and subcategorization information were processed. These data suggest a model of sentence comprehension in which the construction of a syntactic representation temporally leads the construction of a more embellished thematic representation. Serial and parallel variants of such a model are discussed.

What is the underlying architecture of the cognitive processes that enable sentence comprehension? A recent, influential model of sentence processing postulates that the first stage in sentence comprehension involves computing a syntactic representation based strictly on knowledge of constituent structure (phrase–structure relations) and the major syntactic categories (e.g., verb, noun, adjective, and so forth) of the words processed in the string (Clifton, Speer, & Abney, 1991; Frazier, 1978, 1987, 1989; Frazier & Rayner, 1982).

Lexical information associated with individual words is substantially richer than a simple specification of a word's syntactic category. A number of researchers have suggested that the lexical representations for verbs (as well as other major categories) include a specification of permissible syntactic frames or environments in which the verb can occur. These frames may be formally described as argument structures (Grimshaw, 1990; Shapiro, Zurif, & Grimshaw, 1987, 1989), functional structures (Kaplan & Bresnan, 1982), licensing relations (Abney, 1989), or thematic role structures (Carlson & Tanenhaus, 1988; Stowe, 1989). For example, in the argument structure formalism of Grimshaw, the lexical representation for the stative verb *love* specifies a two-place predicate or argument structure in which the surface subject of the verb fills the thematic role of an *experiencer* and the direct object of the verb encodes the thematic role of *theme*. The claim of Frazier and colleagues is that although in principle thematic role information as well as other (e.g., semantic and pragmatic) information could be used to immediately assign, or aid in the assignment of, syntactic relations, such information is not processed until an initial syntactic structure has been first computed. Frazier and colleagues proposed a two-stage model in which more detailed thematic role information is used to embellish the initial representation, interfacing the syntactic representation with semantic and discourse representations, or to revise the initial syntactic analysis if it proves incorrect. This second stage of processing was referred to as the *thematic processor* (Rayner, Carlson, & Frazier, 1983). Mitchell (1987a, 1987b, 1994; Adams, Clifton, & Mitchell, 1991) has argued for a similar two-stage model that consists of an initial processor (termed the *director* or *assembler*), which constructs a tentative analysis based on category and order information, followed by a second processor (termed a *monitor*), which uses lexical constraints to evaluate the initial analysis.

In contrast with these two-stage models, a number of researchers have claimed that immediate parsing decisions are not based on strict category information only. Minimally, it has been argued that immediate parsing decisions are guided by lexical information that specifies the permissible or preferred thematic or argument frames associated with the verb(s) in a sentence (e.g., Carlson & Tanenhaus, 1988; Ford, Bresnan, & Kaplan, 1982; Shapiro, Nagel, & Levine, 1993; Stowe, 1989). For example, Carlson and Tanenhaus argued that access to a verb's representation in the mental lexicon activates the thematic grid(s) associated with that verb. These grids specify potential argument roles licensed by the verb and are conceptualized as open slots that can be filled by previously processed and to be processed elements. Thus, processing of the verb *love* will activate a grid that specifies an experiencer slot for the clause–initial noun phrase (NP) and a theme slot that can be filled by the first NP following the verb. Similar views have been advanced by Ford et al. (1982) and Shapiro et al. (1993).

Recent constraint–satisfaction approaches to sentence processing present a more radical departure from the two-stage model (Bates & MacWhinney, 1989; McClelland, St. John, & Taraban, 1989; MacDonald, 1994; MacDonald, Pearlmutter, & Seidenberg, 1993; Taraban & McClelland, 1988; Trueswell & Tanenhaus, in press; Trueswell, Tanenhaus, & Garnsey, 1994). In general, these approaches claim that lexical, semantic, and even discourse information may rapidly be used to guide the assignment of a syntactic structure. Extant constraint–

Brian McElree, Cognitive Sciences Department, University of California, Irvine; Teresa Griffith, Linguistics Department, University of California, Irvine.

This research was supported by a University of California Faculty Research Grant (92/93-90-FRG) awarded to Brian McElree. We thank Barbara A. Dosher and John Trueswell for helpful comments and suggestions.

Correspondence concerning this article should be addressed to Brian McElree, Cognitive Sciences Department, Social Science Tower, University of California, Irvine, California 92717. Electronic mail may be sent via Internet to bdm@agni.ss.uci.edu.

satisfaction approaches differ greatly, however, in the types of information that are thought to constrain sentence interpretation and when each type of information is thought to be actively processed. At one extreme is the distributed processing architecture proposed by McClelland et al. (1989) in which syntactic processing is given no autonomous status, but rather "there is but a single integrated system in which syntactic and other constraints are combined in the connection weights" (p. 329). Other constraint–satisfaction approaches have posited that syntactic information may be represented independently of thematic role and other conceptual information but make different assumptions about when each type of information is recruited in the course of processing a sentence. Bates and MacWhinney suggested that while syntactic and conceptual (i.e., semantic and pragmatic) information provide two distinct cues or constraints in sentence processing, both types of information are used immediately in an interactive fashion. In contrast, Trueswell, Tanenhaus, and Kello (1993; Trueswell & Tanenhaus, 1994; see also MacDonald et al., 1993) suggested that syntactic processing (in particular, the assignment of syntactic arguments) may actually precede thematic and semantic processing, but that the latter are used to rapidly select among competing syntactic analyses.

A number of recent experiments have attempted to evaluate the two-stage model, contrasting it with the lexical guidance/constraint–satisfaction models. We do not present an exhaustive review of this literature in this article (see Ferreira and Henderson, 1991, and Trueswell et al., 1993, for partial reviews) but rather focus on a subset of the experiments and the resulting issues that have been raised by this debate. One consequence of the assumption that the first stage in parsing utilizes knowledge of constituent structure and category information only is a high degree of local structural ambiguity: At various points, a sentence may be compatible with more than one structural representation. Frazier and colleagues have argued that local ambiguities at the first stage of processing are resolved (at least temporarily) by a set of general processing strategies. The most notable strategy instantiates a principle of minimal attachment, which mandates that when there is more than one way of attaching a constituent to a phrase marker, the "simplest" structure is preferred (in which simplest is defined as the structure with the fewest number of nodes intervening between the new node and the existing phrase marker; Frazier, 1987). Evidence for the two-stage model consists of demonstrations that the initial analysis of a sentence is controlled by strategies like minimal attachment, despite the fact that thematic (as well as semantic or discourse) information inherent in the text may ultimately force a different, nonminimal attachment analysis.

Illustrative of this style of research is an experiment by Ferreira and Clifton (1986) in which eye movements during the reading of sentences such as Examples 1 and 2 were examined:

The detective (that was) examined by the lawyer turned out to be unreliable. (1)

The evidence (that was) examined by the lawyer turned out to be unreliable. (2)

Two variants of each type of sentence were used. One variant included an overt relative clause marker, *that was,* to unambiguously denote the beginning of a relative clause. This variant forced a past participle reading of the verb *examined* with the initial NP assigned the thematic role of theme. The critical comparison concerned the reduced relative clause version of each sentence, in which the relative marker was eliminated. There, the verb *examined* was temporarily ambiguous between a relative clause reading, requiring a past participle form of the verb, and a main clause reading, requiring a simple past tense form of the verb with the thematic role of *agent* rather than theme assigned to the sentence–initial NP. Note that in Example 2, however, the sentence–initial NP *the evidence* is inanimate and hence inappropriate as an agent for the verb. If thematic role information were used to guide the initial structural analysis of the sentence, then it should unambiguously force a relative clause analysis with a past participle reading of the verb. A minimal attachment strategy, however, predicts that the simpler, main clause analysis with a past tense verb reading will be preferred regardless of the animacy of the initial NP. The subsequent phrase, *by the lawyer,* ultimately resolves the ambiguity in favor of the nonminimal attachment, relative clause reading. Ferreira and Clifton found longer fixation times on the *by the lawyer* phrase for both reduced relative forms as compared with sentences with overt relative clause markers. Hence, they argued that the initial analysis of the sentence was guided by the purely syntactic strategy of minimal attachment and that thematic role information was not used immediately to assign a structural representation. In short, minimal attachment induced a temporary garden path that could have been avoided if thematic information guided the initial parse.

Studies with similar logic have purported to show that other information that could potentially constrain an initial analysis does not appear to override a minimal attachment strategy. These have included experiments examining possible pragmatic and discourse constraints (Rayner et al., 1983), as well as verb subcategorization restrictions and verb subcategorization preferences (Ferreira & Henderson, 1990; Mitchell, 1989). Although recent studies have tended to confirm that discourse constraints do not override a minimal attachment strategy (Mitchell, Corley, & Garnham, 1992; Rayner, Garrod, & Perfetti, 1992), the situation is less clear with respect to more local lexical information. A number of initial demonstrations have failed to replicate when slightly different materials or different experimental procedures were used. For example, the Ferreira and Clifton (1986) result has been reexamined by Trueswell et al. (1994) and Burgess (1991). Both studies corrected some possible artifacts in the materials and used slightly different methods of presentation. These studies demonstrated that, contra Ferreira and Clifton's findings, the animacy of the initial NP did reduce the fixation or reading time in the ambiguous reduced relative structures, indicating that thematic role information does aid in assigning an initial syntactic representation (see Boland and Tanenhaus, 1991, for a review of the impact of other types of lexical information).

If additional studies continue to demonstrate that thematic role and other lexical information aid in the initial analysis of a sentence, then the particular two-stage model proposed by

Frazier and colleagues (Clifton et al., 1991; Frazier, 1978, 1987, 1989; Frazier & Rayner, 1982) will have to be abandoned. Following the approach of McClelland et al. (1989), one conclusion that might be drawn from the failure of this two-stage model is to reject the assumption of many psycholinguists that sentence processing includes a set of computations that uniquely serve to generate a syntactic structure for a sentence. That is, it might be argued that there is no direct experimental evidence that indicates that sentence processing includes a set of computations in which syntactic structure, as opposed to thematic, semantic, and discourse structure, is strictly at issue.

We argue that such a conclusion is premature. Studies such as Trueswell et al. (1994) and Burgess (1991; as well as related studies such as Taraban and McClelland, 1988) serve only to call into question strategies such as minimal attachment. In point of fact, proponents of structurally based strategies could argue that these studies serve simply to delimit situations in which such strategies are deployed; specifically, that attachment strategies might be operative only in situations in which other constraints are weak. Even if structural strategies were completely abandoned, a variant of the two-stage architectural assumption proposed by Frazier and colleagues may still provide a veridical description of the sentence processing mechanism. Note that structural strategies such as minimal attachment are necessary only if structural processing is serial—that is, just in case structural processes are capable of computing only one structural representation at a time. This assumption has been questioned on a number of grounds (see, for example, MacDonald et al., 1993; McElree, 1993; Trueswell et al., 1993, 1994; Waltz & Pollack, 1984). Alternatively, an initial stage of syntactic processing may project more than one structural representation of an ambiguous structure. Lexical, semantic, and even discourse constraints may subsequently filter out all but the appropriate analyses (MacDonald et al., 1993; Trueswell et al., 1993, 1994; Waltz & Pollack, 1984).

An adequate model of sentence processing must ultimately specify the means by which various sorts of ambiguity are resolved. However, examining the processing of ambiguous constructions is not the only experimental means of uncovering component processes in sentence comprehension. Some recent studies have examined the processing of lexical information in unambiguous sentence structures. Shapiro et al. (1987, 1989) estimated processing load by measuring secondary task load (lexical decision times for an unrelated word) immediately after processing a verb. This work purported to show that the complexity of a verb, defined in terms of the number of potential argument structures, could immediately affect processing load, thereby indirectly indicating that argument or *thematic* role information is processed immediately after encountering a verb (see, however, Schmauder, 1991). Other studies (Clifton, Frazier, & Connine, 1984; McElree, 1993; Shapiro et al., 1993) have used a rapid grammaticality judgment task (as discussed below) to demonstrate that the structural preferences of verbs have an early impact in sentence processing. Unfortunately, none of these studies explicitly contrasted the processing of syntactic and thematic role information: Although these studies indicate that lexical information is recruited early in processing, they provide no means of estimating the relative time courses of syntactic and thematic role processing. Our experiments sought to dissociate syntactic and thematic processing and to measure the respective time course of each.

Contrasting the Time Course of Syntactic and Thematic Processing

Consider what distinguishes the simple grammatical sentence in Example 3 from the ungrammatical strings (denoted by asterisks) in Examples 4–6.

Some people love books. (3)

*Some people agree books. (4)

*Some people amuse books. (5)

*Some people hastily books. (6)

Examples 3–5 all share a common syntactic structure, which, glossing over a number of important details, might be represented as $[_S[NP][_{VP}V\,NP]\,]$. Example 4 differs from Example 3 in terms of the subcategorization restrictions (e.g., see Chomsky, 1965) on the matrix verb *agree.* Example 4 is unacceptable because the verb *agree* is intransitive and as such cannot incorporate a direct object NP. Example 5, unlike Example 4, uses a verb that does subcategorize for an object NP, but which nevertheless is unacceptable because of the implicit thematic roles associated with the verb.[1] In Grimshaw's (1990) analysis, the verb *amuse* has a two-place argument structure in which the argument role internal to the verb phrase (VP) specifies the entity that is the experiencer of the psychological state of being amused, while the argument role external to the VP, namely the surface subject of the clause, specifies the theme or entity–object that induces the psychological state. Example 5 is unacceptable because the NP in the experiencer role, *books,* does not possess the requisite properties to fill this thematicrole.[2] Finally, Example 6 is a case of ill-formed constituent

[1] In recent formal theories (e.g., Chomsky & Lasnik, 1991), subcategorization restrictions are represented implicitly in the thematic role structures associated with a verb. For example, an intransitive verb like *agree* in Example 4 is simply a verb that does not assign a thematic role to an object NP. (An NP in the object position of such a verb will be excluded by more general principles of the grammar, namely, the θ criterion, see Chomsky & Lasnik, 1991). In subsequent discussion, we use the term *subcategorization restrictions/violations* as a descriptive label to denote cases in which the verb does not license an object NP. We reserve the label *thematic role violations* for cases like Example 5 in which an unacceptable NP occurs in the (required) object position.

[2] The unacceptability of Example 5 may reduce in this case to the simple fact that *books* are inanimate entities and hence cannot fill the role of an experiencer. This in turn might suggest that a single selective restriction feature like +/−ANIMACY (Chomsky, 1965) may serve to determine whether an NP can fill various thematic roles. Although possible, the details of such a proposal have not been worked out. We take a more conservative approach, suggesting that it may be a number of features or properties of a noun that determine whether it can fulfill a particular thematic role (see also the General Discussion section).

structure simply because the VP lacks a verb. As such, Example 6 lacks a well-defined syntactic representation such as $[_S[NP][_{VP}V\,NP]]$.

We used ill-defined sentence strings like Examples 4–6, along with a set of appropriate (control) grammatical structures like Example 3, to contrast the processing of syntactic and thematic role information. Our goal was to measure the time it takes to determine that structures like Examples 4–6 are ill formed. For example, if processes that compute a syntactic structure temporally lead those that compute a thematic structure, then we should observe that ill-formed constituent structures like Example 6 can be detected sooner than structures with ill-formed thematic representations like Example 5. Such a prediction clearly follows from the strict serial two-stage architecture proposed by Frazier and colleagues and perhaps also from constraint–satisfaction architectures in which establishing a syntactic representation is thought to precede establishing thematic role representations (e.g., Trueswell & Tanenhaus, 1994; Trueswell et al., 1993).[3] Additionally, the model of Frazier and colleagues (Clifton et al., 1991; Frazier, 1978, 1987, 1989; Frazier & Rayner, 1982) assumes that the initial syntactic stage of processing uses knowledge of constituent structure and major category information only; more specific lexical information, such as verb subcategorization restrictions, enters into computations at the second stage of processing only. A recent study by Trueswell et al. has questioned this assumption, demonstrating that subcategorization information can be used early in processing. We contrasted structures like Example 4 that are ungrammatical because of subcategorization restrictions with structures like Examples 3 and 5 as a means of determining the relative time in processing where subcategorization information is at issue.

We used a rapid grammaticality judgment task to measure the time taken to detect the respective violations. This task required participants to read the sentence string and at some point make a rapid yes–no judgment as to whether the string was an acceptable English sentence. In our application, the strings were read one word at a time and the judgment was made after reading the final word in the string (*books* in Examples 3–6). Hence, critical contrasts like Examples 3–6 were acceptable up to the final word in the string, at which point they became ill formed as a consequence of the various constraints outlined above. We used two different variants of the task, specifically a standard reaction time (RT) variant of the task and a more involved response–signal speed–accuracy trade-off (SAT) variant.

Reaction Time (RT) Measures

In the RT version of the grammaticality judgment task, participants judged whether a string was acceptable as quickly but as accurately as possible immediately after reading the final word in a sentence (e.g., *books* in Examples 3–6). Barring factors outlined below, we expected that differences in the time course of processing various types of information would be directly reflected in mean RT, RT accuracy, or both. For example, if the architecture proposed by Frazier and colleagues were essentially correct, then mean RT for detecting violations of constituent structure of the sort in Example 6 would be shorter than for detecting the thematic role violation in Example 5. Likewise, if subcategorization information were used in the second stage only, then mean RT for constructions like Example 6 would be shorter than for strings with subcategorization violations like Example 4. Any observed differences in mean RT for strings like Examples 4 and 5 would indicate the relative time course of the use of each type of information. In all cases, we would expect RT accuracy to pattern with mean RT appropriately: Increases in mean RT would be accompanied by higher error rates.

Unfortunately, although predictions concerning mean RT and RT accuracy seem relatively straightforward, an RT task is of limited value in assessing properties of the time course of processing. RT methods are subject to speed–accuracy tradeoffs that, in the extreme, could conceal differences between conditions or, in less extreme situations, could obscure the relative magnitude of the underlying differences in time course. Even in the highly implausible case that a speed–accuracy trade-off is inoperative in the task of interest, differences in mean RT and RT accuracy may not reflect differences in the time course of processing. Mean RT and RT accuracy typically vary with the strength of evidence supporting a response, independent of underlying differences in the time course of processing. This fact has been extensively documented in various types of memory judgments (Dosher, 1982, 1984; McElree & Dosher, 1989, 1993; Wickelgren, 1977) and is a central feature of explicit models of RT (e.g., Ratcliff's [1978] diffusion or random walk model). Directly relevant to present concerns, McElree (1993) demonstrated that RT differences in a rapid grammaticality judgment task for processing a verb in a preferred (more frequent) versus less preferred syntactic environment were due to differences in the strength (or probability) of retrieving syntactic information associated with the verb rather than to inherent differences in the time course for processing preferred and less preferred information.

Any potential differences in the strength of evidence supporting an ungrammatical (i.e., a no) response among contrasts of the type in Examples 4–6 could affect both mean RT and RT accuracy, thereby blocking the simple interpretation that differences in mean RT and RT accuracy reflect differences in the underlying time course of processing. Consider, for illustrative purposes, the contrast between Examples 4 and 5. Although in each case the verb is followed by an inappropriate direct object NP, the verb *amuse* in Example 5 subcategorizes for a direct object NP, whereas the verb *agree* in Example 4 is strictly intransitive, disallowing a direct object NP. There may be more evidence to support an ungrammatical response in the

[3] Without a sufficiently well-specified computational model, however, it is difficult to know with certainty whether a massively parallel constraint–satisfaction model of the type envisioned by Trueswell and colleagues would indeed predict substantial temporal differences between Examples 5 and 6.

case in which no NP is possible (Example 4) than in the case in which an (albeit different) NP is possible (Example 5).

Despite the limitations of the RT task, we nevertheless report results from an RT task in Experiment 1 to document differences in mean RT and RT accuracy for contrasts of the sentence forms in Examples 4–6. We do so primarily to show that the pattern of data from the more common RT task is compatible with the slightly different and more extensive method of measuring time course, namely the SAT procedure. The SAT procedure overcomes many of the inherent limitations of the RT method.

Speed–Accuracy Trade-Off (SAT) Measures

The response–signal SAT procedure interrupts the judgment in question with a cue to respond, typically a brief (e.g., 50-ms) tone, after varying amounts of processing time. Participants are required to respond at the tone, irrespective of whether processing is complete. Accuracy (usually d') is thereby measured as a function of processing time. By presenting the tone across an appropriate range of times (typically 10–3,000 ms), the full time course of processing can be measured. Full time course data typically show a period of chance performance, followed by a period of increasing accuracy, and finally by a period of asymptotic performance in which accuracy does not increase as further processing time is allowed. Generally, three parameters suffice to describe time course functions: (a) an asymptotic parameter reflecting the ultimate level of accuracy, (b) an intercept parameter reflecting the discrete point in time when the function departs from chance, and (c) a rate of rise parameter that describes the rate at which accuracy grows from chance to asymptote.

Asymptotic performance is a measure of the overall strength of evidence supporting a particular judgment. Stronger evidence will result in higher accuracy, which will be reflected in higher estimates for the asymptotic parameter. Differences in the time course of processing are reflected in the intercepts and rates of rise to asymptote, independent of the asymptotic levels. The intercept and rate jointly describe the *dynamics* or speed of processing, reflecting either the rate of continuous information accrual or the distribution of finishing times of a discrete or quantal process (Dosher, 1976, 1979, 1981, 1982; Meyer, Irwin, Osman, & Kounois, 1988; Ratcliff, 1988).

Examination of the empirical estimates for the dynamics parameters in grammaticality judgments of constructions like Examples 4–6 provides a strong test of the architecture of the sentence processing mechanism. Briefly, the two-stage serial model proposed by Frazier and colleagues (Clifton et al., 1991; Frazier, 1978, 1987, 1989; Frazier & Rayner, 1982) predicts that judgments of constructions that are ill formed as a consequence of either subcategorization or thematic role violations will be associated with a longer intercept or a slower rate of rise (or both) than constructions with ill-formed constituent structures. This follows directly from the claim that subcategorization and thematic role information are not consulted until after an initial constituent structure has been computed. Similar predictions may also follow from variants of the Frazier architecture that relax the strong serial assumption. For example, dynamics differences could result from a mechanism in which the processing of syntactic, subcategorization, and thematic role information is arranged in a cascaded (e.g., McClelland, 1979) or parallel fashion if, for example, the processing of one type of information starts before another or if one type of information is processed at a faster rate than another. We focus initially on the serial architecture proposed by Frazier and colleagues (Clifton et al., 1991; Frazier, 1978, 1987, 1989; Frazier & Rayner, 1982). SAT predictions from this type of model are presented in more detail in Experiment 2, after we present the basic pattern of RT data. We discuss cascaded and parallel variants after the SAT results have been presented (see the General Discussion section).

Experiment 1: Reaction Time Measures

In this experiment, we used a standard RT variant of the grammaticality judgment task to contrast the processing of syntactic, subcategorization, and thematic role information. Table 1 presents sample materials used to effect this comparison. The critical contrasts involved ungrammatical strings ill formed as a result of violations of thematic role relations, subcategorization restrictions, and constituent structures. To generalize the results beyond a specific type of construction, we used active and passive variants of each of the three types of constructions.

Thematic role violations were constructed around a set of psych-causative verbs, such as *alarmed, amused, offended,* and so forth, in the active constructions and a set of psych-stative verbs, such as *loved, admired, deplored,* and so forth, in the passive constructions (see Grimshaw, 1990, for a detailed discussion of the properties of these verbs). Following Grimshaw, we assumed that psych-causative verbs assign the thematic role of theme to the sentence–initial NP and the role of experiencer to the direct object NP. (Active) psych-stative verbs display the reverse assignment of thematic roles, in that the thematic role of experiencer is assigned to the sentence–initial NP and the role of theme to the direct object NP. In passive sentence frames, the mapping of psych-stative verbs'

Table 1
Example Materials Used in Experiments 1 and 2

Condition	Ungrammatical	Grammatical (control)
Active		
Thematic violations	Some people alarm books.	Some books alarm people.
Subcat. violations	Some people agree books.	Some people agree rarely.
Category violations	Some people rarely books.	Some people rarely agree.
Passive		
Thematic violations	Some people were loved by books.	Some books were loved by people.
Subcat. violations	Some people were agreed by books.	Some people were agreed with rarely.
Category violations	Some people were rarely with agreed.	Some people were rarely agreed with.
Control	Some books love people.	Some people love books.
Control	Some books were alarmed by people.	Some people were alarmed by books.

Note. Subcat. = subcategory.

thematic roles to surface structure is reversed once again, yielding a thematic role assignment that parallels the (active) psych-causative verbs. Thematic role violations were constructed in each case by using a direct object NP, for example, *books* in Table 1, that lacks the requisite properties to serve as an experiencer.

Our primary interest was to contrast thematic role violations with subcategorization and constituent structure violations. Active and passive subcategorization violations were constructed by using a set of strictly intransitive verbs, such as *smiled, agreed, winced,* and so forth. Strict intransitive verbs, of course, disallow a direct object NP and are thus ungrammatical when followed either by a simple noun in the active constructions or by a *by*-NP phrase in the passive constructions. Violations of constituent structure, which we label as *category violations,* were constructed around a set of adverbs, such as *rarely, boldly, intensely,* and so forth. In the active constructions, ill-formed constituent structures were generated by following the adverb with a noun (*books*) rather than a verb. In the passive variants, ill-formed constituent structures were generated by following the adverb with a preposition, such as *with, by, over,* and an intransitive verb.

A critical feature of the ungrammatical conditions was that all relevant (noncontrol) strings became ungrammatical at approximately the same point. In the case of the active ungrammatical contrasts, the thematic role, subcategorization, and category violations were all ill formed at the final word in the string (e.g., *books* in Table 1). In the case of the passive ungrammatical conditions, the relevant strings were ill formed at the last or next to last word in the string. The passive–thematic role and category violations were, like the active strings, ill formed at the final word. The passive–subcategorization violations were ill formed at the start of the *by*-phrase, the next to last word in the string. We used a *by*-phrase here to provide a more minimal contrast with the passive–thematic role violations, which also used a two-word *by*-phrase. Approximately equating the point at which the strings become ungrammatical enabled us to directly contrast when each respective type of information was processed. Participants were presented the strings one word at time (200 ms per word) and were required to assess the acceptability of the string at the final word. (The final word was followed by a period to clearly mark the end of a string.) The instructions to the participants emphasized that their (speeded) judgments should reflect whether or not the string was a "well-formed, meaningful sentence in English." RT and RT accuracy as a function of the nature of grammatical violation were the measures of primary interest.

We included an equal number of grammatical control strings in the experimental set. To a first approximation, these sentences were grammatical analogues of the ungrammatical strings. (For example, grammatical variants of the thematic role violations were constructed by reversing the surface positions of the two NPs, *books* and *people.*) Additional controls (the control conditions in Table 1) were added to decouple the nature of the initial NP and the type of judgment. In particular, ungrammatical strings were constructed with a sentence-initial inanimate NP, *books* in Table 1, so that subjects could not predict an ungrammatical response from the nature of the initial NP. Moreover, because these ungrammatical controls were ill formed at the main verb (rather than after the verb), they prevented the participants from focusing exclusively on material after the main verb.

Method

Participants. The participants were 18 undergraduates from the University of California, Irvine who participated to gain course credit in an undergraduate psychology course. All participants were native English speakers with normal or corrected vision. Eight participants served in one 1-hr session, 4 participants served in two 1-hr sessions, and 6 participants served in three 1-hr sessions.

Materials. We used the materials for this experiment in the SAT study reported in Experiment 2. The SAT procedure requires substantially more trials and hence more materials than the RT task. Ten sets of materials were constructed, with each set serving as the materials for one (RT and SAT) session. Each set consisted of 448 strings, made up of 224 grammatical and 224 ungrammatical strings, representing 28 instances of each of the 16 conditions illustrated in Table 1.

The sets consisted of the following. Twenty-eight psych-stative verbs, 28 psych-causative verbs, 28 intransitive verbs, and 28 adverbs were generated. Ten sets of plural NP pairs (e.g., *people–books* in Table 1) were then generated. We combined the verbs–adverbs with the 10 sets of NP pairs or, where appropriate (see Table 1), with one of the members of an NP pair to yield 10 sets of materials. The combination was such that each NP occurred approximately equally often in each of the 16 conditions, which in turn ensured that any inherent differences among the NPs (i.e., differences in frequency, word length, and so forth) were distributed equally across the 16 conditions. Sample materials from the corpus are presented in the Appendix.

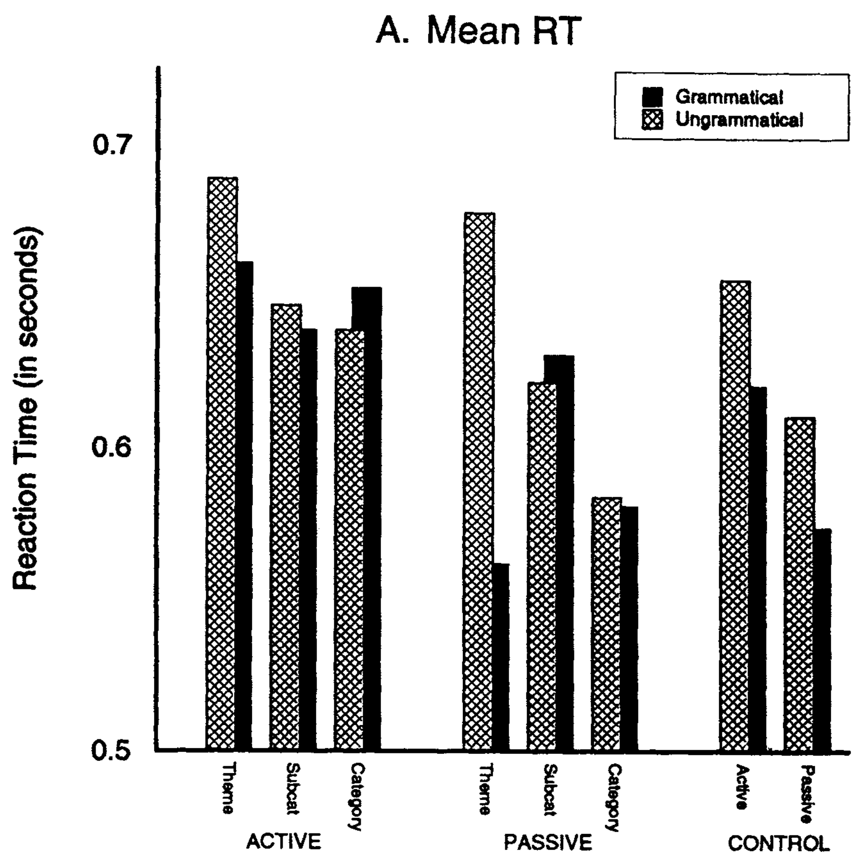
Procedure. Stimulus presentation, timing, and response collection were all carried out on a personal computer. A trial began with a fixation point (a small filled square) presented for 500 ms in the center of an otherwise clear screen. The words of a string were presented one after another in the center of the screen in a normal mixture of uppercase and lowercase characters. Each word remained on the screen for 200 ms. A period was appended to the final word of the string to clearly indicate to participants that the presentation of the string was complete and that a response was required.

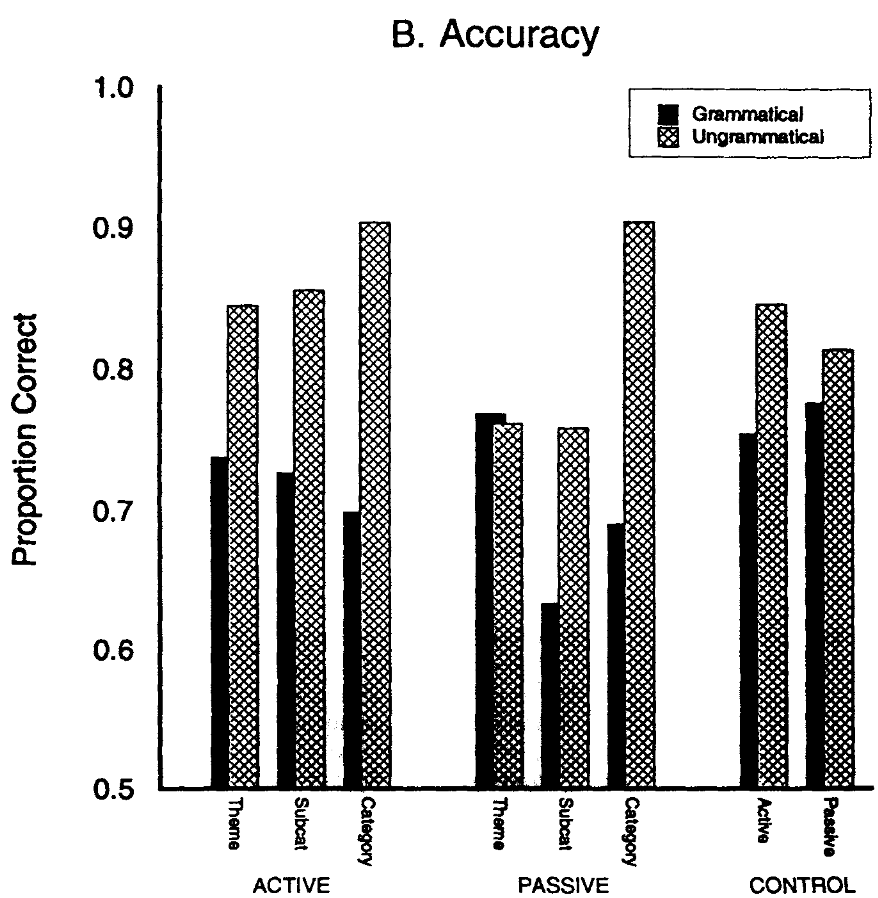
We instructed participants to read the strings as they would normally read any text and to decide "as quickly but as accurately as possible" after the final word was presented whether the string was a well-formed, meaningful English sentence. The first session was preceded by 20 practice trials. Participants were randomly assigned to one of the 10 sets of materials, with the constraint that no set was repeated if the participant served in more than one session. The sentences within a set were presented to the participants in a random order.

Results and Discussion

Grammatical strings were correctly classified, on average, 72.4% of the time, with a mean latency of 615 ms. Ungrammatical strings were correctly classified 83.7% of the time, with a mean latency of 641 ms. The average correct RT and proportion correct by condition for grammatical and ungrammatical strings are presented in Figures 1A and 1B.

In Figure 1 and for subsequent analyses, we computed average RT by including correct responses only, excluding trials in which response times fell below or exceeded 2.5


A. Mean RT
Grammatical
Ungrammatical
Reaction Time (in seconds)
0.7
0.6
0.5
Theme
Subcat
Category
Theme
Subcat
Category
Active
Passive
ACTIVE
PASSIVE
CONTROL


B. Accuracy
Grammatical
Ungrammatical
Proportion Correct
1.0
0.9
0.8
0.7
0.6
0.5
Theme
Subcat
Category
Theme
Subcat
Category
Active
Passive
ACTIVE
PASSIVE
CONTROL

standard deviations of the participant's average correct response time (approximately 3.3% of trials). RT and RT accuracy were analyzed with separate analysis of variance (ANOVA) procedures that (a) averaged over individual sentences and treated participants as the random and repeated measures variable (denoted by F_1) and (b) averaged over participants and treated individual verbs–adverbs as the random and repeated measures variable (denoted by F_2).

The comparisons of interest concern differences in mean RT and RT accuracy for the three different types of violations and, to a lesser extent, potential differences between the corresponding grammatical (control) conditions. Consider first the data for the active sentence frames. Inspection of Figure 1 suggests that there were no large differences among the three types of active grammatical strings. A one-way ANOVA (with sentence type as the main within-subject/item variable) found no reliable difference in either mean RT (both Fs $\approx$ 1) or RT accuracy (both Fs $\approx$ 1). In contrast, there were large and quite reliable differences among the three types of ungrammatical strings both in mean RT, $F_1(2, 34) = 6.43, p < .004, MSE = 0.0131; F_2(2, 54) = 15.44, p < .000, MSE = 0.0531$, and in RT accuracy, $F_1(2, 34) = 3.39, p < .046, MSE = 0.0174; F_2(2, 54) = 5.18, p < .009, MSE = 0.0184$. Our a priori interest was to contrast detection time and accuracy among the three types of ungrammatical strings. We performed separate ANOVA procedures to statistically compare performance among the three types of violations. Detection of thematic role violations was reliably slower than detection of subcategorization violations, $F_1(1, 17) = 8.48, p < .01, MSE = 0.0160; F_2(1, 27) = 9.97, p < .004, MSE = 0.0446$, and category violations, $F_1(1, 17) = 8.16, p < .011, MSE = 0.0227; F_2(1, 27) = 26.23, p < .000, MSE = 0.1027$. The observed accuracy for detecting thematic role violations was also lower than the accuracy for detecting either subcategorization or category violations; however, only the difference between thematic role and category violations was significant: $F_1(1, 17) = 5.85, p < .027, MSE = 0.0312; F_2(1, 27) = 7.75, p < .01, MSE = 0.0302$. The small 1% to 2% advantage for detecting subcategorization as compared with thematic role violations was nonsignificant in both subject and item analysis (both Fs $\approx$ 1). We found no reliable differences in the RTs for detecting subcategorization and category violations (both Fs $\approx$ 1). Accuracy, however, was found to be reliably higher for detecting category as compared with subcategorization violations, although only in the item analysis: $F_1(1, 17) = 2.9, p < .107, MSE = 0.0198; F_2(1, 27) = 8.50, p < .007, MSE = 0.0247$.

The data for passive sentence frames display a similar, although not identical pattern. In this case, there was some evidence to suggest that the judgments of the three grammatical strings differed in accuracy, $F_1(2, 34) = 5.11, p < .012, MSE = 0.0832; F_2(2, 54) = 15.53, p < .000, MSE = 0.1360$, and in RT, although only in the subject analysis, $F_1(2, 34) = 5.27, p < .01, MSE = 0.0226; F_2(2, 54) = 2.37, p < .103, MSE = 0.0200$. This difference can be localized to the relatively slower and error-prone judgments of the grammatical analogue of the subcategorization violations (e.g., "*Some people were agreed with rarely*"). This difference may be related to the placement of an adverb after a complex verb sequence, but is of no consequence to our purposes. Of critical importance are the judgments of ungrammatical strings. As with the active sentence frames, there were reliable overall differences among the three types of passive ungrammatical strings both in mean RT, $F_1(2, 34) = 10.7, p < .000, MSE = 0.0397; F_2(2, 54) = 19.01, p < .000, MSE = 0.1453$, and in RT accuracy, $F_1(2, 34) = 6.8, p < .003, MSE = 0.1254; F_2(2, 54) = 27.09, p < .000, MSE = 0.1972$. The latency for detecting thematic role violations was longer than the latency for detecting either subcategorization violations, $F_1(1, 17) = 9.59, p < .007, MSE = 0.0283; F_2(1, 27) = 7.31, p < .012, MSE = 0.0800$, or category violations, $F_1(1, 17) = 20.5, p < .000, MSE = 0.0784; F_2(1, 27) = 51.4, p < .000, MSE = 0.2904$. Detection accuracy was lower for thematic role as compared with category violations: $F_1(1, 17) = 14.5, p < .001, MSE = 0.1835; F_2(1, 27) = 52.7, p < .000, MSE = 0.3373$. However, accuracy rates were nearly identical for thematic role and subcategorization violations (both Fs $\approx$ 1). Figure 1 shows an average RT advantage of about 70 ms for detecting category as compared with subcategorization violations. This difference was significant in the item analysis but was not replicated in the subject analysis: $F_1(1, 17) = 2.87, p < .109, MSE = 0.0125; F_2(1, 27) = 10.34, p < .003, MSE = 0.0656$. The corresponding difference in accuracy was significant by both analyses: $F_1(1, 17) = 9.81, p < .006, MSE = 0.1927; F_2(1, 27) = 52.74, p < .000, MSE = 0.3373$.

To summarize, in both the active and passive sentence frames, the data clearly indicate that the detection of thematic role violations was slower, more error prone, or both than the detection of either subcategorization or category violations. These data are certainly consistent with the notion that the processing of thematic role relations is delayed relative to the processing of category and subcategorization information. As for any potential differences between subcategorization and category violations, the data are less clear. There was some evidence that category violations were more accurately detected than subcategorization violations in the passive constructions, but there was little clear evidence to support the claim that category violations are detected sooner than subcategorization violations. However, as with any speeded judgment task, participants may trade differences in speed for accuracy. Indeed, accuracy in this task was not high, averaging 78% across both grammatical and ungrammatical strings. Although some portion of these errors may reflect limited processing as a consequence of the rapid presentation rate (200 ms per word), the concomitantly fast (correct) response times (628 ms averaging across grammatical and ungrammatical strings) are certainly indicative of a speed–accuracy trade-off in which participants emphasized speed at the expense of accuracy. Without explicit control over one of these dimensions, we cannot reliably determine the cause of the observed differences. The same point, of course, applies to the differences observed for thematic role

Figure 1 (opposite). Mean correct reaction time (A) and mean proportion correct (B) for grammatical (solid bars) and ungrammatical (hatched bars) strings in Experiment 1. Subcat. = subcategorization.

violations. In the next experiment, we used SAT methodology to examine these differences in more detail.

Experiment 2: Speed–Accuracy Trade-Off Measures

The response–signal speed–accuracy trade-off procedure controls potential speed–accuracy trade-offs and provides a means of measuring the full time course of processing.

Figure 2A illustrates the application of the SAT procedure to the grammaticality judgment task. A sentence string was presented to participants to read, one word at a time. Participants were cued to respond with a tone presented at one of seven temporal points spanning from 10 to 3,000 ms after presentation of the final word in the string. Figure 2B shows a typical SAT function (response accuracy, here d', as a function of processing time) generated by this procedure. Full time-course SAT functions furnish estimates of when performance departs from chance (an intercept), the ultimate level of accuracy reached (an asymptote), and a speed at which accuracy grows from chance to asymptote (a rate).

Figure 3 illustrates two different SAT predictions, both consistent with the RT and RT accuracy differences observed in Experiment 1 but with substantially different theoretical imports. In Figure 3 and the subsequent discussion, we frame the discussion in terms of differences between detecting thematic role violations and syntactic violations; the latter is meant to include both subcategorization and category violations. We do so for expository convenience and because the most reliable differences observed in the RT task were between thematic role violations and the two other (syntactic) violations.

Figure 3A illustrates a case in which the only difference between thematic role and syntactic violations; is in asymptotic accuracy, with the latter violations resulting in higher detection rates than the former. The detections of thematic role and syntactic violations have the same underlying time course: Both functions have identical dynamics instantiated as equal intercepts and rates of rise to asymptote. (Consistent with prior SAT studies [e.g., Dosher, 1976, 1982, 1984; McElree, 1993; McElree & Dosher, 1989, 1993; Wickelgren, 1977], in this experiment it was assumed that the form of SAT function was approximated by an exponential approach to a limit; as discussed below.) The dotted line in Figure 3A illustrates that the difference between two SAT functions that differ only in terms of their asymptotic levels is itself a monotonic function that rises to the level of the differences in asymptotes. Differences in asymptotic performance can occur if the evidence supporting one type of violation is stronger than the evidence supporting another. We noted previously that constructions with thematic role violations may provide weaker evidence in support of an ungrammatical response than constructions with subcategorization violations because in the former (but not in the latter) an object NP is licensed. There are other differences between the ungrammatical strings that may affect the strength of evidence for a particular response. For example, the thematic role violations have a well-defined syntactic representation, specifically the common, perhaps even canonical structure $[_S[NP][_{VP}V\ NP]]$ (Bever, 1970). Category violations, in contrast, have no comparable representation. The well-defined syntactic representation for the thematic role violations may provide some evidence in favor of a grammatical response that competes with the evidence for an ungrammatical response stemming from the processing of thematic relations. The net effect may be to reduce the overall evidence for an ungrammatical response, resulting in a lower overall level of accuracy. Asymptotic performance reflects the overall level of (un)acceptability for a string. Systematic differences in asymptote, in turn, may provide a means of determining how various sources of information contribute to an overall assessment of the (un)acceptability of a string. However, if differences in the SAT functions are restricted to asymptote alone, then the implication is that the different sources of information are processed at the same time.

Figure 3B illustrates a case in which the detection of thematic role and of syntactic violations have different underlying time courses. Specifically, it illustrates the case in which thematic role processing occurs in a second stage following an initial stage of syntactic processing, assuming that the stages are linked in a strict serial fashion. In this instantiation of the two-stage serial model, the detection of any syntactic violation (e.g., the category violations in Table 1) occurs in the first stage of processing, whereas thematic role violations are detected in the second stage of processing. (In this simulation, each processing stage was assumed to be independently and identically exponentially distributed; see Luce [1986] and Townsend and Ashby [1983], for a general discussion of the use of the exponential distribution in modeling serial processing stages and McElree [1993] and McElree and Dosher [1993] for some specific applications.) The two-stage serial model predicts that the intercept will be longer and the rate of rise to asymptote will be slower for detecting thematic role violations than for detecting syntactic violations. This difference in dynamics is clearly illustrated in the dotted curve that plots the differences between the two functions. In contrast to the monotonic difference function in Figure 3A, this function shows a large difference early in processing that is eliminated later in processing. The large difference early in processing reflects the fact that the well-defined syntactic structure for the thematic role violations passes the first stage of processing, resulting in a high false-alarm rate as compared with the syntactic violations, which are correctly rejected at the first stage of processing.

Dynamics differences of the form illustrated in Figure 3B would provide strong support for the claim that thematic role processing is delayed relative to syntactic processing. It is important to note that, although we have derived predictions for SAT dynamics under a strict serial architecture, dynamics differences of the form illustrated in Figure 3B may also result from various parallel architectures. We explore architectural variants in the General Discussion section, after the empirical data have been presented.

Finally, as stated, the panels in Figure 3 illustrate two distinct predictions: Figure 3A illustrates a pure effect on (asymptotic) accuracy arising from potential differences in the type of evidence implicit in the ungrammatical strings, whereas Figure 3B illustrates a pure effect on processing dynamics (speed) arising from inherent differences in the time course of processing. Of course, it is possible to observe a mixture of both effects, and indeed such mixtures have been reported

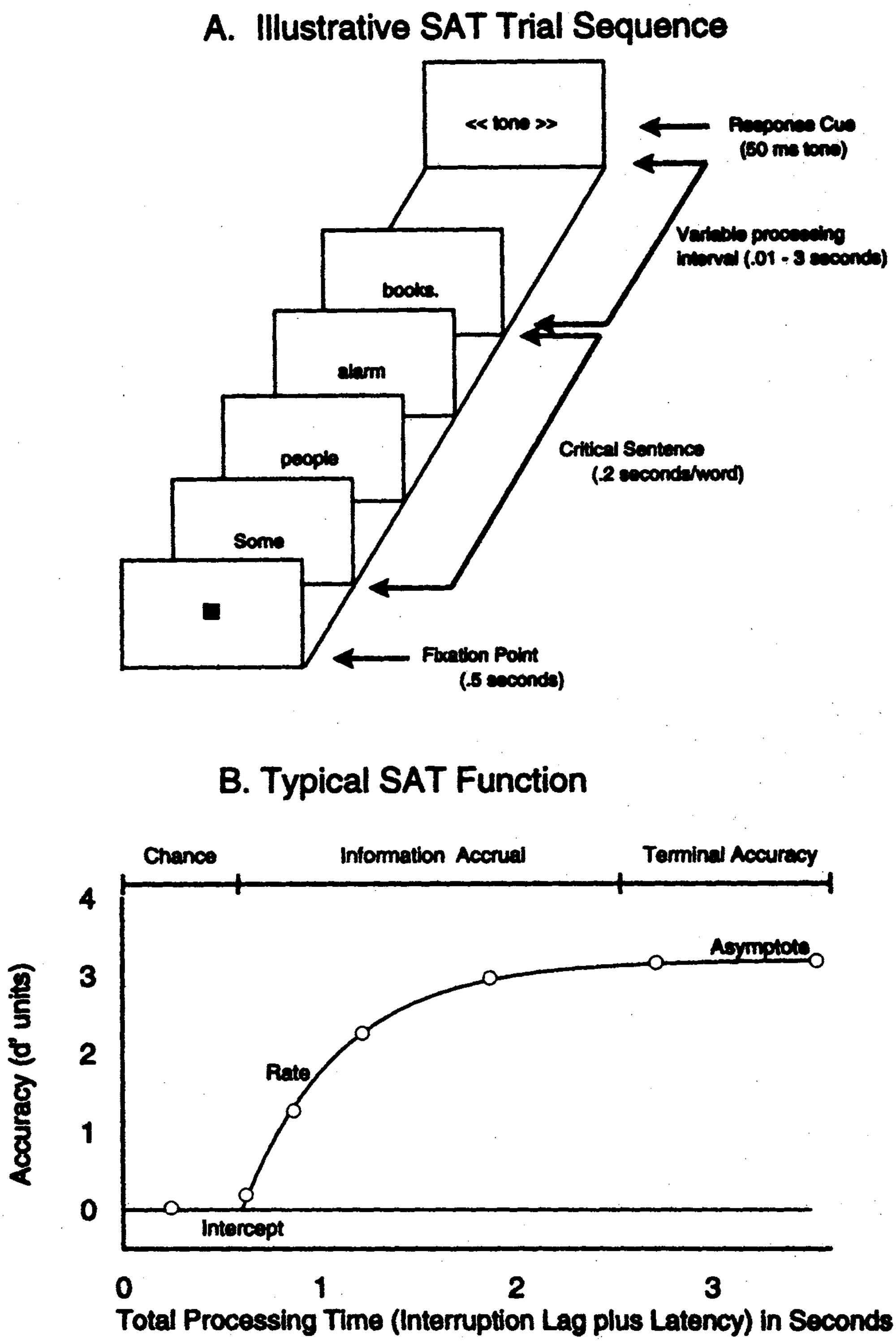


Figure 2. A sample trial sequence illustrating the response-signal speed–accuracy trade-off (SAT) variant of the rapid grammaticality judgment task (A) and (B) a typical (hypothetical) SAT function generated from this procedure.

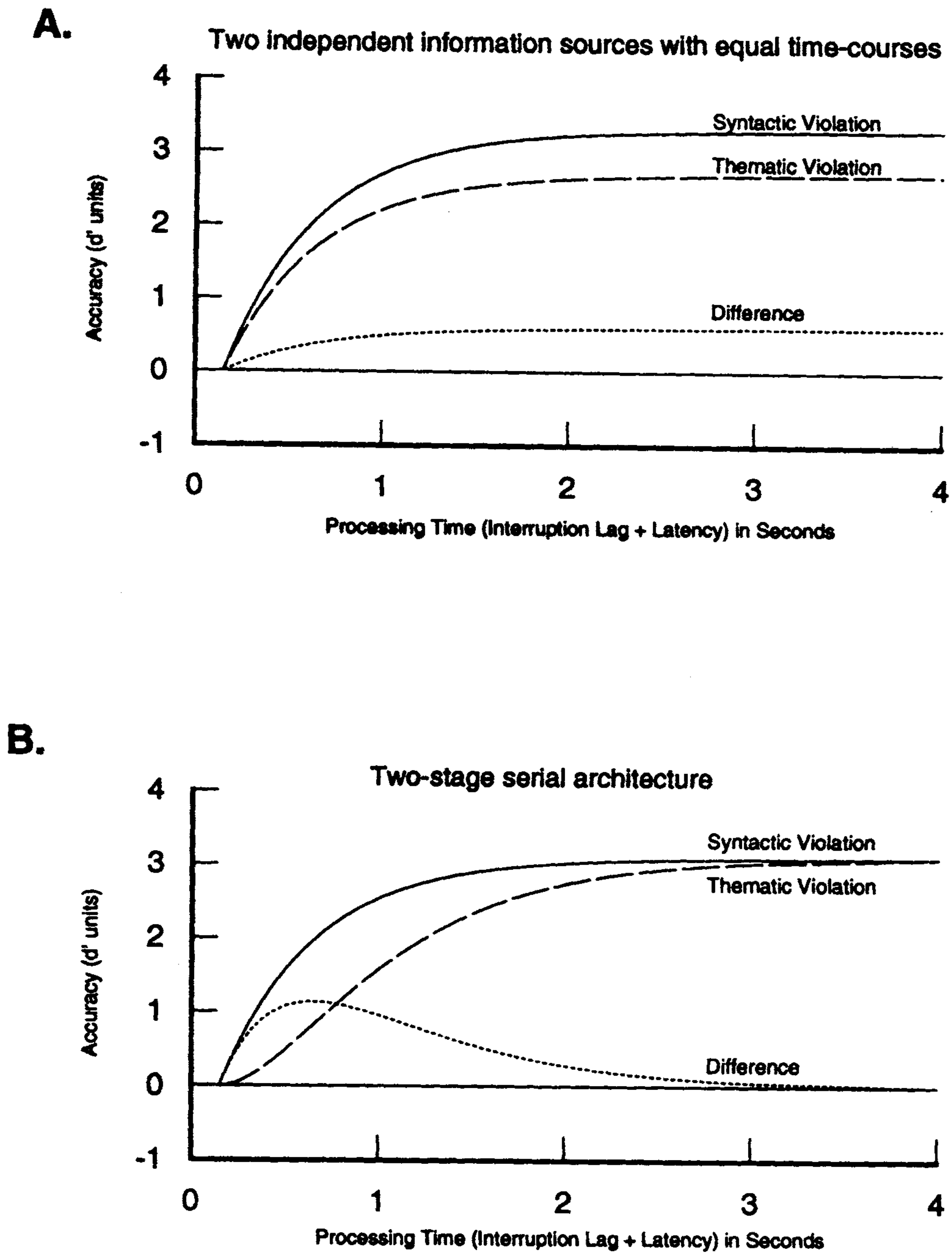


Figure 3. Speed–accuracy trade-off predictions for the grammaticality judgment task. (A) Illustration of the case in which syntactic and thematic role violations differ in terms of asymptotic accuracy only. (B) Illustration of predictions from the two-stage serial model (see text).

(McElree & Dosher, 1993). However, a number of SAT studies have documented pure effects on accuracy (e.g., Dosher, 1982, 1984; McElree, 1993; McElree & Dosher, 1989) and on processing dynamics (e.g., Dosher, 1981; Dosher & Rosedale, 1989). The empirical dissociation between accuracy and dynamics motivates a theoretical treatment in which such effects arise from fundamentally different aspects of processing (e.g., McElree & Dosher, 1989, 1993; Ratcliff, 1978). Nevertheless, a mixture effect in which accuracy directly varies with processing speed at least raises the possibility that both effects arise from the same underlying source. Note, however, that any (as yet unspecified) model that posits that both components of the SAT function arise from the same source would not be able to account for studies (cited above) that have demonstrated independent, dissociable effects. Consequently, the most viable models are those that give separate theoretical treatment to the speed and accuracy components of the SAT function.

Method

Participants. The participants were 6 graduate and undergraduate students affiliated with the Cognitive Sciences Department at the University of California, Irvine. All were native English speakers with normal or corrected vision. None of the participants had served in Experiment 1, and all participants were paid for their participation.

Materials and procedure. The 10 sets of sentences used in Experiment 1 also served as materials for this experiment. An SAT trial is illustrated in Figure 2A. The procedure was the same as described in Experiment 1 up until the presentation of the final word in a string. In the SAT procedure, a 50-ms (1000 Hz) tone was presented at 14, 157, 300, 557, 800, 1,500, or 3,000 ms after the final word. Participants responded yes or no by pressing one of two designated keys as soon as the tone sounded. After a response was made, visual feedback on the latencies to respond to the tone was given. We informed participants that latencies (to respond to the tone) longer than 300 ms were too long and that latencies shorter than 100 ms were anticipations. All participants had an initial 1-hr practice session. Each participant served in 10 approximately 1-hr sessions. In each session, participants used 1 of the 10 sets of materials. The assignment of materials to session was randomized across the 6 participants.

Data analysis. For each participant, the proportion correct for each condition (sentence type) at each response lag was used to derive a standard d' measure (equal variance Gaussian model). For grammatical conditions, the d' measure was constructed by scaling the proportion of hits in either an active or passive condition against a common false-alarm rate estimated from pooling all the active or passive ungrammatical conditions. Likewise, the d' measure for ungrammatical conditions was constructed by scaling the correct rejection rate in an active or passive condition against a common miss rate estimated from pooling all active or passive grammatical conditions. This scaling ensures that comparisons of different ungrammatical conditions reflect inherent differences among these conditions rather than differences among the grammatical conditions. Similarly, comparisons of different grammatical conditions reflect inherent differences among those conditions rather than differences among ungrammatical conditions. We adjusted perfect performance in any condition by a minimum-error correction that replaced a zero error rate with an error rate of 0.5%. This correction ensures that, given the sample size, the resulting d' values are in fact measurable (Macmillan & Creelman, 1991).

Standard SAT functions can quantitatively be summarized by the exponential approach to a limit shown in Equation 1 (cf. Dosher, 1976, 1979; Reed, 1973, 1976; Wickelgren, 1977).

$$d'(t) = \lambda(1 - e^{-\beta(t-\delta)}), t > \delta \quad \text{else } 0. \tag{1}$$

In Equation 1, the parameter λ serves to describe the asymptotic level of performance. The δ and β parameters jointly estimate the dynamics of the function. The former describes the intercept or the discrete point in time at which performance begins to rise from chance. The β parameter indexes the speed or rate at which performance rises from chance to asymptote. Analysis of the SAT function in terms of Equation 1 allows an independent assessment of information availability (λ) and the dynamics of information accrual (β and δ).[4]

Equation 1 was fit to the empirical SAT functions by using an iterative hill climbing algorithm (Reed, 1976), which is similar to STEPIT (Chandler, 1969). This algorithm minimized the squared deviations of predicted values from observed data. We used sets of competitive fits, varying the three parameters of Equation 1, to isolate differences between various SAT functions. The quality of the respective fits was assessed by three criteria as follows: (a) The value of an R^2 statistic,

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(d_i - \hat{d}_i)^2/(n-k)}{\sum_{i=1}^{n}(d_i - \bar{d})^2/(n-1)}, \tag{2}$$

where d_i represents the observed data values, $\hat{d}_i$ indicates the predicted values, $\bar{d}$ is the mean, n is the number of data points, and k represents the number of free parameters (Reed, 1973). This R^2 statistic is the proportion of variance accounted for by the fit, adjusted for the number of free parameters (k). (The same equation for adjusted r^2 is often used in multiple linear regression [e.g., Judd & McClelland, 1989].) (b) Evaluation of the consistency of parameter estimates across the 6 participants. (c) Most important, examination of whether a fit yielded systematic (residual) deviations that could be accounted for with more parameters.

Results and Discussion

Sets of competitive fits, systematically varying the three parameters of Equation 1, were applied to the individual participant data to isolate differences in the time courses of processing the various constructions. Consistent patterns across participants were summarized with analyses and graphs of the average (over participants) data.

Grammatical conditions. Differences among the active and passive grammatical conditions were of no intrinsic theoretical interest. Nevertheless, we performed sets of competitive fits on these constructions to (a) check the applicability of the d' scaling method and to (b) test whether participants were performing the task in an appropriate manner.

Figure 4A presents the average (over participants) empirical SAT data (symbols) and best fitting exponential function (line) for the three different types of active (control) constructions (see Table 1). There was no evidence across participants that judgments of the three active grammatical constructions systematically differed in the time course of processing, in terms of asymptotic performance (λ), SAT intercept (δ), or SAT rate (β). Consequently, the three average SAT functions were best summarized with a simple 1-λ, 1-β, and 1-δ variant of Equation 1 (R^2 = .962). This fit is shown as the single smooth function (line) in Figure 4A. The data from 3 participants (M.M., P.K., and S.H.) directly reflected the pattern in the average data and, consequently, were also best fit by a simple 1-λ, 1-β, and

[4] The empirical functions were also fit with a descriptive equation that derives from a time-bounded diffusion process as developed by Ratcliff (1978): $d'(t) = \lambda/\sqrt{1 + \nu^2/(t - \delta)}, t > \delta$, else 0. This equation, like the exponential Equation 1, has three parameters: λ represents the asymptotic accuracy level, δ is the intercept, and ν^2 is the combined random walk variance term formed by the ratio of drift rate variance (S^2) over item relatedness variance (η^2). ν^2, like β in Equation 1, indexes the speed with which accuracy rises from chance to asymptote. Fits of this equation showed the same pattern as fits of Equation 1. Hence, the conclusions drawn here are not restricted to the exponential form but generalize to at least one other three-parameter equation that provides an adequate description of time-course data. To conserve space, we reported and discussed the results for Equation 1 only.

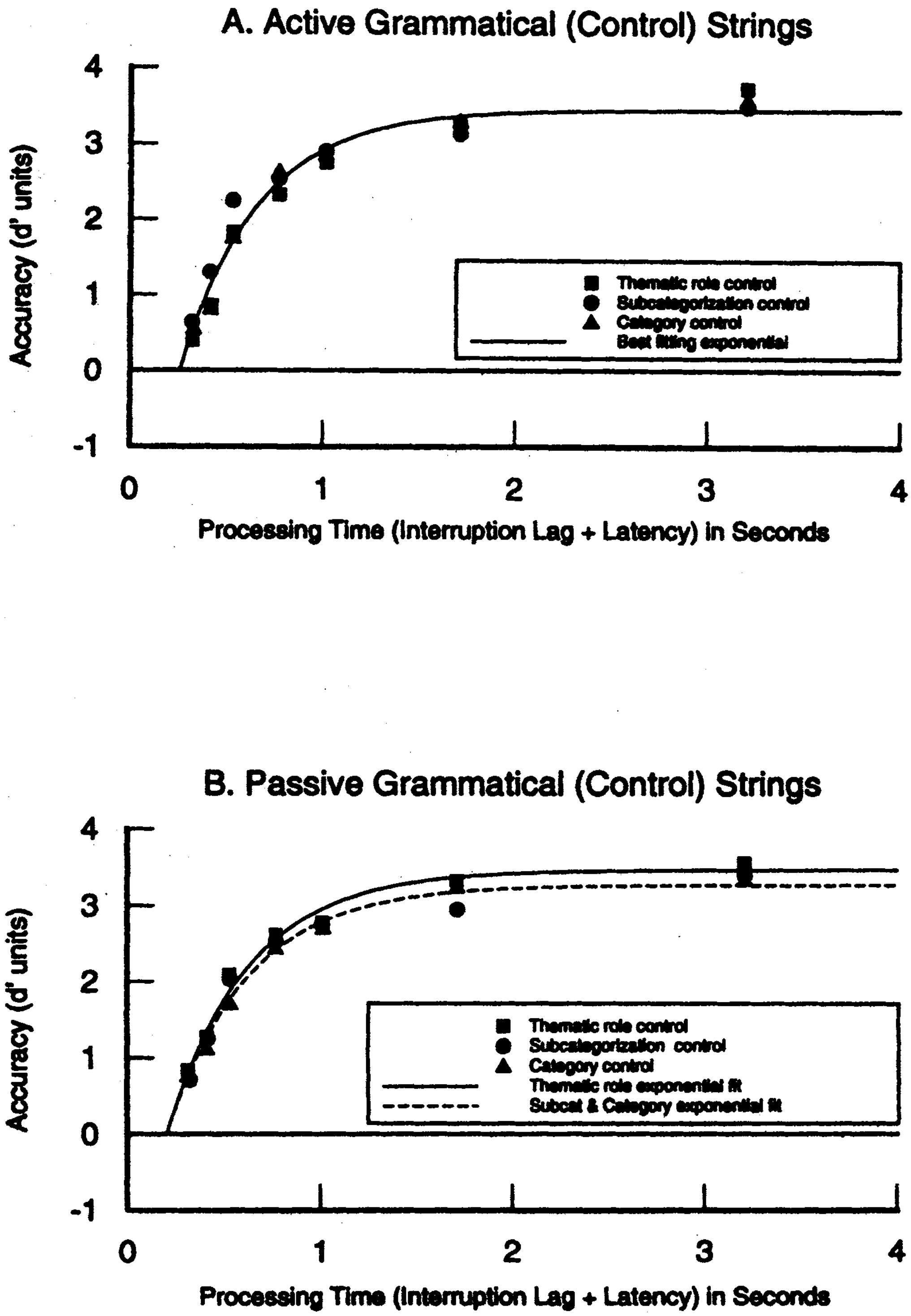


Figure 4. Average (over participants) d' accuracy (symbols) as a function of processing time for judgments of active (A) and passive (B) grammatical strings. Smooth curves in each panel show the best fits of Equation 1 with (the average) parameters listed in Table 2. Subcat. = subcategorization.

1-δ variant of Equation 1 (R^2 values of .891, .959, and .822, respectively). Allocation of additional parameters reduced the adjusted R^2 value for these participants and did not yield any stable differences in parameter estimates. The remaining participants were better fit by a 3-λ, 1-β, and 1-δ variant of the exponential that produced slightly higher R^2 values than the 1-λ, 1-β, and 1-δ variant (improvements in R^2 ranged from .008 to .016). However, this more embellished fit served only to capture idiosyncratic differences in asymptotic performance across the three conditions. These differences are documented in Table 2, which presents the parameter estimates of the best fitting model for the average and individual participant data.

Figure 4B presents the average empirical SAT data (symbols) and best fitting exponential functions (lines) for the three different types of passive grammatical (control) constructions (see Table 1). The average data in this case were best fit by a 2-λ, 1-β, and 1-δ variant of Equation 1 (R^2 = .979) because asymptotic accuracy was higher for the passive [*NP V by* NP] structure (e.g., "Some books were loved by people") than for the two variants with a more complex VP (e.g., "Some people were agreed with rarely" or "Some people were rarely agreed with"). The difference in asymptote was small, estimated at 0.2 d' units. The fit of the average data was representative of 3 participants (B.K., J.K., and S.H.), who were likewise best fit with a 2-λ, 1-β, and 1-δ model (R^2 values of .865, .854, .829, respectively, with differences in asymptote that ranged from 0.38 to 0.73 d' units). No comparable difference in asymptote was found for the other 3 participants (C.S., M.M., and P.K.) as these subjects were best fit with a simple 1-λ, 1-β, and 1-δ variant of Equation 1 (R^2 values of .955, .911, and .969, respectively). No participant showed any indication of dynamics differences among the various grammatical constructions, as evidenced by the fact that models that allotted more intercept (δ) or rate (β) parameters to the various conditions failed to produce any consistent trends across participants and resulted in lower adjusted R^2 values. The parameter estimates for the best fitting models for the average and individual participant data are shown in Table 2.

In the RT experiment, judgments of one of the passive grammatical constructions, specifically constructions of the "Some people were rarely agreed with" type, were slower and more error prone than the other two passive grammatical constructions. The SAT data suggest that these RT differences probably reflected differences in asymptotic accuracy rather than in dynamics. Moreover, the differences in asymptote appeared to be idiosyncratic, as no consistent pattern emerged across participants. The grammatical constructions were of no intrinsic interest; they served only as controls and as a means of estimating a miss rate for d' scaling of the ungrammatical constructions. However, because there were potential differ-

Table 2
Exponential Parameter Estimates

		Participants					
Parameter	Average	B.K.	C.S.	J.K.	M.M.	P.K.	S.H.
Active grammatical							
λ Thematic role control	3.46	2.71	3.58	3.66	3.51	3.93	3.69
λ Subcat. control	″	2.32	4.11	3.36	″	″	″
λ Cat. control	″	2.26	4.31	3.17	″	″	″
Common β	2.49	1.98	2.50	5.14	1.66	2.84	2.11
Common δ	.257	.270	.310	.300	.182	.247	.282
R^2	.962	.866	.968	.745	.891	.959	.822
Passive grammatical							
λ Thematic role control	3.48	2.56	3.83	3.68	3.51	4.49	3.17
λ Subcat. control	3.28	1.84	″	3.27	″	″	2.77
λ Cat. control	″	″	″	″	″	″	″
Common β	2.36	1.81	2.61	3.53	1.89	2.13	2.00
Common δ	.204	.251	.224	.235	.169	.164	.174
R^2	.979	.865	.955	.854	.911	.969	.829
Active ungrammatical							
λ Thematic role violations	3.45	2.52	4.16	3.05	3.34	4.28	3.22
λ Subcat. violations	″	2.81	″	3.32	″	″	″
λ Cat. violations	″	1.94	″	3.69	″	″	″
β Thematic role violations	2.04	1.00	1.66	7.78	1.73	1.76	2.82
β Subcat. & cat. violations	3.09	1.76	2.53	″	3.37	2.40	4.81
Common δ	.259	.316	.263	.341	.268	.209	.321
R^2	.976	.829	.983	.920	.942	.964	.831
Passive ungrammatical							
λ Thematic role violations	3.28	2.15	3.57	3.32	3.32	4.31	3.30
λ Subcat. violations	3.41	1.96	4.13	″	3.65	4.06	3.36
λ Cat. violations	3.12	2.21	3.55	″	3.31	3.73	2.66
β Thematic role violations	2.41	2.26	2.64	2.42	1.73	2.83	1.86
β Subcat. & cat. violations	″	″	″	4.05	″	″	″
δ Thematic role violations	.291	.381	.339	.180	.271	.279	.264
δ Subcat. & cat. violations	.155	.187	.225	″	.313	.123	.224
R^2	.973	.912	.947	.675	.922	.938	.914

Note. Subcat. = subcategorization; Cat. = category. Ditto marks indicate that the value of interest is jointly estimated by the parameter above; see the text for details (pp. 145–150).

ences among the grammatical constructions, we used a single miss rate for each participant in the d' scaling of the ungrammatical conditions, derived by pooling the participant's data across all active or passive grammatical conditions.

Ungrammatical conditions. Figure 5 presents the average empirical SAT data (symbols) and best fitting exponential functions (lines) for the three types of grammatical violations in active (Figure 5A) and passive (Figure 5B) sentence frames. Consider first the active ungrammatical strings. Inspection of Figure 5A clearly indicates that the accuracy of detecting thematic role violations (squares) was lower early in processing (interruption times less than 1 s) than either the subcategorization (circles) or category (triangles) violations. Later in processing, as the functions reach asymptote, this difference was eliminated. This presents prima facie evidence that the dynamics for detecting thematic role violations are slower than the dynamics for detecting either subcategorization or category violations. Fits of the exponential quantify this difference.

The best fit for the average data allotted a single asymptotic (λ) parameter and a single intercept (δ) parameter to all three types of violations. The lower performance early in processing for thematic role violations was adequately fit by allotting one rate (β) parameter to the detection of thematic role violations and another rate parameter to the detection of both subcategorization and category violations. This 1-λ, 2-β, and 1-δ model produced a higher R^2 value (.976) than any simpler or more embellished variant. The β parameter for detecting thematic role violations was 34% slower than the estimated β parameter for detecting the other two types of violations (2.04 vs. 3.09). Inspection of Figure 5A shows that there was no evidence that the dynamics or asymptotes for detecting subcategorization violations differed from category violations. The smooth functions in Figure 5A show the best fitting 1-λ, 2-β, and 1-δ model using the parameter estimates presented in Table 2.

The data of 4 of the 6 participants (C.S., M.M., P.K., and S.H.) directly reflected the pattern in the average data, and, accordingly, the 1-λ, 2-β, and 1-δ model produced the highest R^2 values for these participants (.983, .942, .964, and .831, respectively). For these participants, as with the average data, we estimated the β parameter to be substantially slower in detecting thematic role violations than in detecting subcategorization and category violations (ranging from 34% to 48% slower rise to asymptote). One participant (B.K.) was better fit by a more embellished 3-λ, 2-β, and 1-δ model (R^2 value of .829). This participant showed evidence of small but reliable differences in the asymptotes among the three construction types that occasioned an improvement of .01 in the R^2 value for a 3-λ as compared with a 1-λ model. Irrespective of this difference in asymptotes, we estimated B.K.'s β parameter for detecting thematic role violations to be 47% slower than the β parameter for detecting subcategorization and category violations. Finally, 1 participant (J.K.) showed an aberrant pattern. The best fit of J.K.'s data was produced by a 3-λ, 1-β, 1-δ model (R^2 value of .920). As with B.K., there were small but estimable differences in asymptotic performance across the three conditions. However, unlike the other 5 participants, there was no evidence of any dynamics differences among the different violations. (All models that varied δ, β, or both reduced the R^2 value to below .910). J.K. could represent a valid counter example to the claim that the dynamics for detecting thematic role violations are slower than the dynamics for the other two types of violations. However, J.K.'s latencies to respond to the first two interruption cues were extremely long as compared with other participants (over 400 ms as compared with 250–300 ms). Consequently, J.K.'s accuracy was extremely high at the earliest lags (above 1.3 d' units). It is likely that J.K. delayed in responding to early interruption cues and that this delay may have obscured potential differences in dynamics. The parameter estimates for the best fitting exponential model for each participant are shown in Table 2.

Figure 5B presents SAT data for judgments of the passive ungrammatical strings. Inspection of this figure suggests that the slower dynamics for detecting thematic role violations found in judgments of the active strings were also found with the passive strings. The most representative exponential model for the average data was one that allotted a separate asymptotic parameter to each condition, a single rate parameter, and one intercept for judgments of thematic role violations and another for both subcategorization and category violations. The 3-λ, 1-β, and 2-δ model produced the highest R^2 value (.973) for the average data and for 5 (B.K., C.S., M.M., P.K., and S.H.) of the 6 participants (R^2 values ranged from .912 to .947). Three λ parameters were needed to capture the modest differences in asymptote in all 5 participants and the average data. However, as with other conditions, there was no consistent pattern to the differences in asymptote. The largest and most consistent differences were in the intercepts. The estimated intercept for detecting thematic role violations was 291 ms as compared with 155 ms for detecting subcategorization and category violations. All 5 participants who were fit with the 3-λ, 1-β, and 2-δ model showed differences in intercept that ranged from 39 to 238 ms. There was no evidence in any fit of the individual participant data (and hence the average data) that the dynamics differed for detecting subcategorization and category violations. More embellished models, in all cases, reduced the R^2 value and yielded inconsistent parameter estimates. The smooth functions in Figure 5B show the best fitting exponential model for the average data, with the parameter estimates displayed in Table 2.

In the active strings, the slower dynamics for detecting thematic role violations were captured in the rate (β) rather than in the intercept (δ) parameter. This apparent discrepancy does not necessarily reflect a difference in processing. Whether a difference in dynamics is expressed in rate or intercept depends on the size and variability of the underlying difference (see Dosher, 1976, 1981, 1982, 1984; McElree & Dosher, 1989, 1993). Suppose, for example, that the dynamics differences are, as hypothesized, a consequence of one process starting later than another. A large difference in start time coupled with a low variability across trials will produce substantial intercept differences. As the difference in start time is reduced or as variability in start time increases relative to the size of the difference, the underlying dynamics difference will emerge in rate rather than intercept. (In short, this is an artifact of averaging over trials.) Indeed, comparison of Figure 5A and Figure 5B clearly shows that the dynamics differences for passive ungrammatical strings were larger than those observed for active ungrammatical strings. The most parsimonious

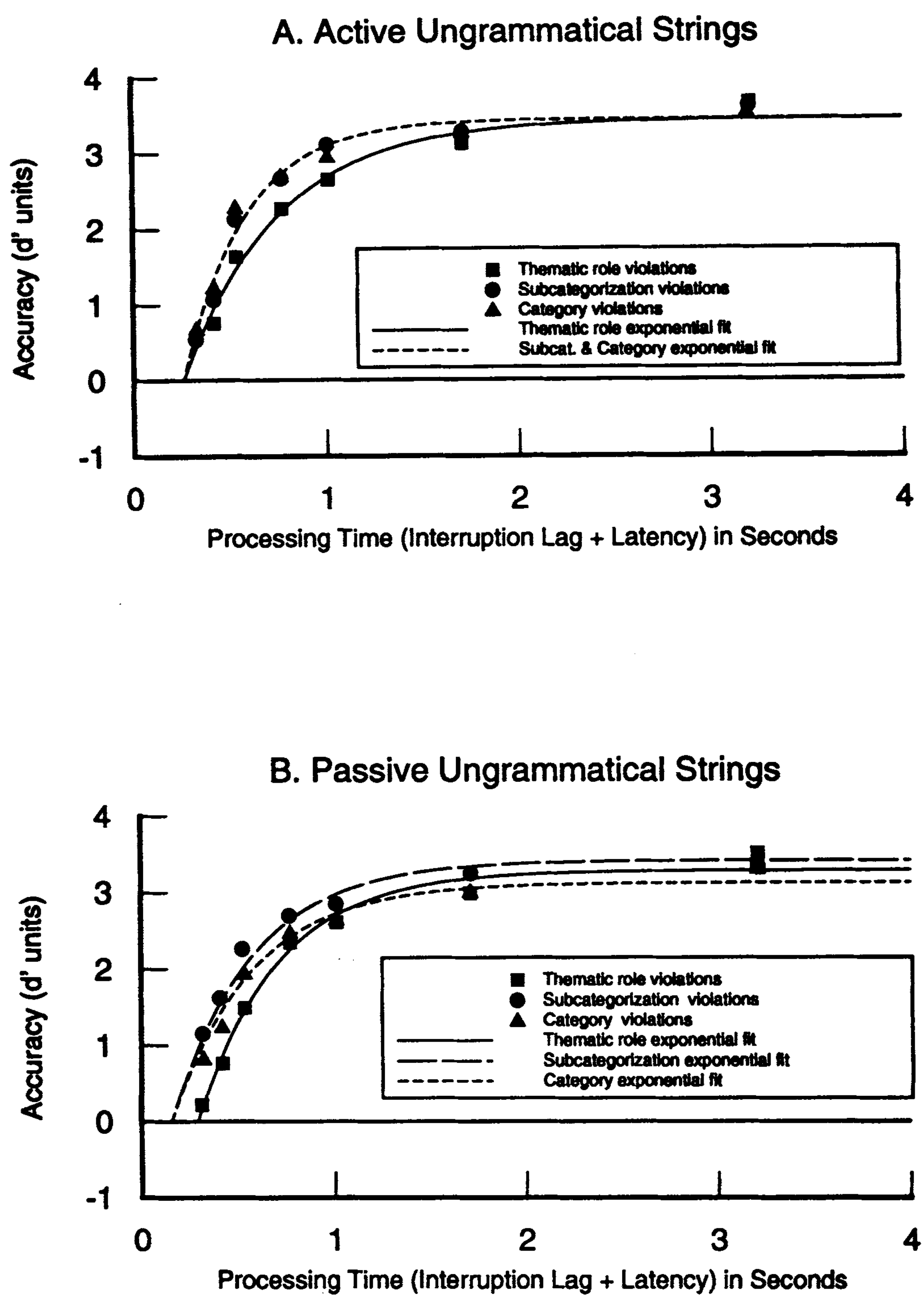


Figure 5. Average (over participants) d' accuracy (symbols) as a function of processing time for judgments of active (A) and passive (B) ungrammatical strings. Smooth curves in each panel show the best fits of Equation 1 with (the average) parameters listed in Table 2. Subcat. = subcategorization.

interpretation is that both differences resulted from the same underlying process(es), but the differences were more extreme in the case of passive strings and hence were expressed in intercept rather than rate.

The only exception to the pattern observed in the average data concerned J.K. J.K.'s passive data were best fit with a 1-λ, 2-β, and 1-δ model, where one β parameter was allotted to thematic role violations and the other to subcategorization and category violations ($R^2 = .675$). J.K. showed nearly identical asymptotic performance across the three types of strings and hence required only one λ parameter. We estimated the rate (β) parameter for detecting thematic role violations to be 40% slower than for detecting subcategorization and category violations. As in the case of judgments of active ungrammatical strings, J.K.'s latency to respond to the early interruption cues was unacceptably long (approximately 400 ms), suggesting that J.K. was delaying in order to respond more accurately. (Again, J.K.'s accuracy at the first 2 interruption cues was well over 1.0 d' units.) Had J.K. responded more quickly to the early interruption cues, we would expect, on the basis of the data from the other participants, that the dynamics difference would have been expressed in intercept rather than in rate.

A two-process model. The fits of the exponential equation to the average and individual participant data served to quantify differences in dynamics (as well as asymptotic) performance. The observed differences in dynamics indicated that thematic role processing was delayed relative to the processing of subcategorization and category information. In this section, we report the application of a two-process model to the data that explicitly assumes that two different types of processing operations—one based on syntactic information, and the other based on thematic role information—contribute to the assessment of the grammatical status of the ungrammatical strings. This two-process account was originally developed for a diffusion or random walk model by Ratcliff (1980) and was extended to exponential form by Dosher (see McElree & Dosher, 1989; see also Dosher, McElree, Hood, & Rosedale, 1989, for application of the two-phase diffusion form).

The model assumes that, during an initial phase of processing, information accrual is described by the exponential approach to a limit in Equation 1 reproduced here as Equation 3a:

$$d'(t) = \lambda_1[1 - e^{-\beta(t-\delta)}], \text{ for } \delta < t < t^*. \tag{3a}$$

In our application, the first phase of processing represents syntactic operations that are initiated at time δ. The parameter λ_1 represents the evidence amassed from syntactic processing during this first phase. β represents the rate at which this evidence grows over processing time. At some point in time $t^* > \delta$, information accrual shifts from the asymptotic level, λ_1, being approached in the first (syntactic) phase to a new level, λ_2, produced by the processing of new information in a second stage of processing. During this second phase (after t^*), Equation 3b holds:

$$d'(t) = [\lambda_2 + (\lambda_1 - \lambda_2)(t^* - \delta)/(t - \delta)] \times [1 - e^{-\beta(t-\delta)}], \quad \text{for } t \geq t^*. \tag{3b}$$

Here the parameter λ_2 represents the evidence derived from the processing of thematic role information. The parameter t^* represents the discrete point in time at which thematic role information begins to be processed.

To facilitate recovery of internally consistent and stable estimates of λ_1, λ_2, β, δ, and t^*, we simultaneously fit five SAT functions for both the active and passive conditions. Figure 6 shows the five functions for the active (Figure 6A) and passive (Figure 6B) conditions. Three of five functions were the same ones that were fit with the simple, descriptive exponential equation, namely the functions for detecting thematic role violations, subcategorization violations, and category violations (filled symbols in Figure 6). We computed two other functions (open symbols in Figure 6) to represent the difference in d' units in detecting (a) thematic role violations as compared with category violations and (b) thematic role violations as compared with subcategorization violations. (Differences between subcategorization and category violations, as evidenced in the previous section, were essentially nonexistent and therefore of no value.) The positive values for the difference scores reflect the higher false-alarm rate (or conversely, the lower correct rejection rate) for detecting thematic role violations relative to the detection rates for either subcategorization or category violations. This is consistent with the notion that, during a first phase of processing, the permissible syntactic representations for the thematic role violations lead to higher false-alarm rates. The difference in false-alarm rates diminishes later in processing as thematic role information begins to be processed. The nonmonotonic form of the difference functions motivates a two-process account and provides a basis on which to separately estimate the contribution of each process.

Three basic asymptotic parameters were used to fit the three ungrammatical conditions and the two difference functions. One asymptotic parameter estimated the evidence for an ungrammatical (no) response that resulted from syntactic processing when the syntactic structure is ill defined. This parameter served as λ_1 for the subcategorization and category violation functions because these constructions had, by definition, an ill-defined syntactic representation. (Here, we assumed that syntactic processes flag an intransitive verb in a $[_S[NP][_{VP}V\ NP]]$ structure as ungrammatical; see the General Discussion section.) A second asymptotic parameter also estimated the evidence for an ungrammatical response resulting from syntactic processing, but for cases in which the syntactic structure was permissible. This parameter served as λ_1 for the thematic role violations. Realistic estimates for the second asymptotic parameter should be substantially lower than for the first and may even take on negative values (reflecting the tendency to false alarm to a well-defined syntactic representation). A third asymptotic parameter estimated the evidence for an ungrammatical response that arose from thematic role processing when thematic role structure was ill defined. This parameter served as λ_2 for the thematic role, subcategorization, and category violations. λ_1 for the difference functions was estimated by the difference between the first and second asymptotic parameters, whereas λ_2 was estimated by the difference between the first and third asymptotic parameters. Jointly fitting the detection and difference functions facilitated recovery of a stable and internally consistent estimate for t^* and the other five parameters.

Table 3 presents the two-process parameter estimates and

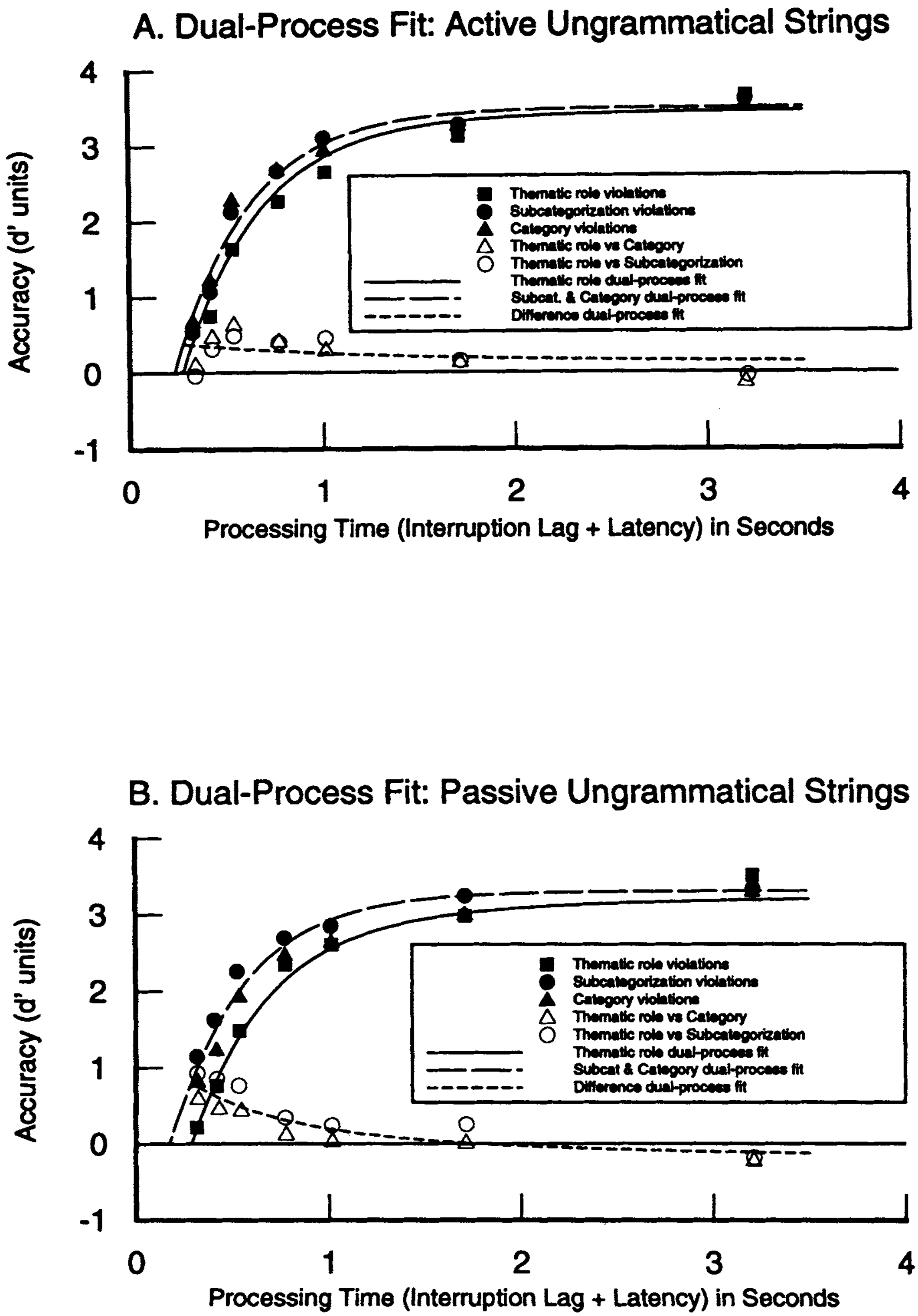


Figure 6. Fits of the dual-process model (smooth curves) to judgments of the active (A) and passive (B) ungrammatical strings. See text for details of the fits. Subcat. = subcategorization.

Table 3
Dual-Process Parameter Estimates

Parameter	Average	Participants					
		B.K.	C.S.	J.K.	M.M.	P.K.	S.H.
Active ungrammatical							
λ for processing ill-defined syntactic structures	3.55	2.26	4.34	3.78	3.70	4.21	3.21
λ for processing well-defined syntactic structures	0.00	−0.35	0.00	0.00	−1.5	0.00	0.00
λ for processing ill-defined thematic structures	3.49	2.63	3.95	3.42	3.38	4.32	3.33
Common β	2.53	1.26	2.69	4.39	2.57	2.24	3.35
Common δ	.232	.147	.294	.296	.229	.185	.277
t^*	.279	.266	.309	.291	.287	.260	.328
R^2	.973	.780	.951	.940	.931	.977	.863
Passive ungrammatical							
λ for processing ill-defined syntactic structures	3.08	1.95	3.88	3.24	3.33	3.39	2.79
λ for processing well-defined syntactic structures	0.00	−0.95	−4.8	−0.81	0.00	0.00	0.00
λ for processing ill-defined thematic structures	3.31	2.25	3.85	3.34	3.44	4.05	3.08
Common β	2.67	1.74	2.56	2.03	2.12	3.58	1.99
Common δ	.173	.100	.219	.172	.100	.149	.243
t^*	.286	.257	.265	.259	.282	.266	.267
R^2	.977	.881	.926	.879	.917	.941	.907

R^2 values for the individual participant and average data for the active and passive conditions. The (solid and dashed) lines in Figure 6 show the model fits for the average data that make use of the parameter estimates in Table 3. In all but one case (J.K. in the active conditions), the estimated time of t^* (the time when thematic role processing begins) was later than δ (the time when syntactic processing begins). For the active conditions, the average estimate of when thematic role processing began was 279 ms as compared with 233 ms for syntactic processing. In the passive conditions, the estimate was 289 ms as compared with 172 ms. It is important to note that t^* was not externally constrained to be longer than δ: The recovered values were solely a consequence of constraints internal to the data sets. This, along with the relatively high R^2 values, provided strong support for the notion that thematic role processing is delayed relative to syntactic processing.

General Discussion

Both the RT and SAT tasks indicated that violations of thematic role structure were detected later in the course of processing a sentence than violations of constituent structure and subcategorization restrictions. The crucial evidence that supports this claim was the observed differences in SAT dynamics. Fits of the descriptive exponential equation found consistently slower estimates for the dynamics parameters (SAT intercept or SAT rate) in judgments of thematic role violations as compared with either category (constituent structure) or subcategorization violations. Fits of a dual-process model indicated that thematic role violations began to be detected 50–100 ms later than either constituent structure or subcategorization violations.

It is important to note that these differences in SAT dynamics were observed in the absence of any systematic effect on asymptotic accuracy when performance levels were low enough to reliably measure such effects. This indicated that the differences in processing the three types of ungrammatical strings were due solely to differences in the time course of processing and not due, in particular, to inherent differences in the strength of evidence implicit in the various ungrammatical strings. We do not claim that all types of syntactic and thematic role violations yield equal levels or degrees of unacceptability. Some other types of syntactic or thematic role violations may well produce systematic differences in (asymptotic) accuracy. (For example, subject–verb agreement violations may be detected with a lower asymptotic accuracy than any of the violations used in this experiment.) The equal level of asymptotic accuracy observed in our experiment indicated only that we selected three types of violations that were roughly equivalent in their evidential status. Our strong and only claim is that our experiments indicate that, independent of differences in the degree of unacceptability, thematic role violations are detected with a slower time course than syntactic (category and subcategorization) violations.

The rapid grammaticality judgment task has been used frequently to measure aspects of sentence comprehension (see McElree, 1993). However, this task requires participants to make an overt judgment that, with the exception of proofreading, is not a typical component of routine sentence processing. It is of course possible that the pattern of results observed with this task may reflect special strategies adopted by the participant rather than standard sentence processing operations. We believe there are a number of aspects of our experiments that speak against this possibility. First, in our instructions to the participants, we emphasized that they could perform the task simply by attempting to recover a coherent meaning or message from the strings. Second, we used a mixed rather than block design in which the various types of ungrammatical (as well as grammatical) strings all occurred randomly within an experimental session. A mixed design prevents participants from adopting specific strategies that are tailored to one type of violation (see Ratcliff, 1978, for a dramatic demonstration of the effects of strategies in a blocked design). Third, language comprehension is an extremely well-developed, highly overlearned skill that is likely to be composed of largely automatic processes. We believe that it is unlikely that participants adopted a novel set of mental procedures to perform the task when standard sentence processing procedures could rapidly and accurately accomplish the task. Indeed, if participants were capable of developing novel procedures, why, for example, could they not have devised a procedure that would detect thematic role violations as quickly as subcategorization and category violations? Ultimately, the question of whether this task taps standard sentence processing operations in a relatively unaltered fashion will be determined by whether our results accord with other experimental procedures. (An event-related brain potential study with similar experimental contrasts may provide one of the least intrusive alternative tasks for assessing convergent validity [e.g., Osterhout & Holcomb, 1992].)

It is our working assumption that the time to detect a particular violation is a function of when the appropriate information is computed in the course of processing a sentence. Hence, we suggest that these experiments indicate that processes early in sentence comprehension seek to compute a syntactic representation for the input string on the basis of knowledge of constituent structure, major category information, and more specific syntactic information such as a verb's subcategorization restrictions. Explicit processing of the thematic role relations occurs later in the course of understanding a sentence. We discuss below two potential reasons why thematic processing may lag behind syntactic processing, casting the discussion in terms of different sentence processing architectures.

Serial Architectures

The delayed processing of thematic role information evidenced in our studies is in part consistent with the two-stage serial architecture of Frazier and colleagues (Frazier, 1987, 1989; Clifton & Ferreira, 1987). With this model, they proposed that processing of thematic role information occurs in a second stage of processing that acts on the output of an initial syntactic stage of processing. The particular model proposed by Frazier and colleagues assumed that the first stage of processing incorporates general knowledge of constituent structure and major category information only. Thus, only specific category information associated with a word but not thematic role or subcategorization information was thought to be relevant to the proposed computations at this stage of processing. Such a view is clearly at odds with the data from our experiments. Although we found evidence for a delay in processing of specific thematic role relations, there was no indication that subcategorization violations were detected later than constituent violations. Consequently, we suggest that a more viable two-stage serial model needs minimally to assume that in the first stage of processing, lexical retrieval routines make available to initial parsing routines subcategorization information (e.g., transitive verb).[5] Trueswell et al. (1993) have also reported results indicating that subcategorization information is used early in the course of processing a sentence.

Perhaps the most controversial component of the two-stage serial model proposed by Frazier and colleagues (Clifton et al., 1991; Frazier, 1978, 1987, 1989; Frazier & Rayner, 1982) is the set of parsing strategies, such as minimal attachment, thought to resolve attachment ambiguities at the first stage of syntactic processing. Because we deliberately sought to contrast the processing of thematic role, subcategorization, and constituent structure information in unambiguous structures, our experiments do not directly bear on the existence of such strategies. Nevertheless, our finding that the processing of subcategorization information is not delayed relative to the processing of constituent structure information does place constraints on the potential utility of these strategies. Frazier (1989) suggested that attachment strategies can freely and generally be applied at the first stage of processing if the computations at this stage are not conditioned on specific lexical properties of the words in a sentence. In the context of a two-stage serial model, our experiments indicate that the first stage of processing is indeed responsive to specific lexical information in the form of subcategorization restrictions. Thus, if attachment strategies are operative early in sentence processing, their application may be more restricted than has been previously assumed.

Parallel Architectures

Although the quantitative analysis of the SAT functions and, in particular, the successful application of a dual-process model to the data are consistent with a serial arrangement of processing stages, they in no sense force a strict serial model. Recently, a number of parallel distributed processing (constraint–satisfaction) architectures have been developed for natural language processing (Bates & MacWhinney, 1989; McClelland et al., 1989; MacDonald, in press; MacDonald et al., 1993; Taraban & McClelland, 1988; Trueswell & Tanenhaus, 1994; Trueswell et al., 1994; see also Reilly & Sharkey, 1992; Sharkey, 1992). The RT and SAT data may be consistent with some, but not all, parallel architectures. Here, we broadly outline the implications that our results have for nonserial architectures.

Consider first strict parallel models that assume that syntactic (category and subcategorization) and thematic role information are processed simultaneously. Two variants of this type of architecture have been proposed. McClelland et al. (1989) argued for a model in which syntactic and other constraints are all instantiated as a single set of connection weights between sentence processing units (role–filler pairs). Clearly, such a model cannot accommodate the time-course differences that

[5] In the context of a two-stage model, our data indicate that both category and subcategorization relations are computed in the first stage of processing. Whether category and subcategorization information are jointly processed by a single set of procedures or by two distinct sets of procedures in this first stage remains an open question.

we observed: Because syntactic and thematic information are given no independent status, this model predicts that both types of information are processed at the same time.

Minimally, our data suggest that syntactic and thematic role information must be independently represented and processed. In contrast to McClelland et al. (1989), constraint-satisfaction models of the type outlined by Bates and MacWhinney (1989) do argue for distinct types of cues or constraints. However, Bates and MacWhinney suggested that all constraints are processed in parallel in an interactive fashion. Parallel models of this type are well understood. A strict parallel model assumes that all processes are initiated at the same time. Once initiated, however, information from each process may accrue at different speeds or rates (see, for example, Murdock, 1971; Townsend & Ashby, 1983). SAT predictions for strict parallel models are straightforward (McElree & Dosher, 1989, 1993): An SAT function that is contingent on the output of one process (e.g., thematic role processing) will display slower rates of rise to asymptote (β in Equation 2) than a function that is contingent on the output of another process (e.g., syntactic processing), just in case the former has a slower rate of information accrual than the latter. Otherwise, the functions will display similar rates of rise. However, even in the case that one process has a relatively slow rate of information accrual, both SAT functions will be associated with a common intercept because, by definition, all processes are initiated at the same time. Hence, the observed intercepts for the SAT functions are the key data that test the applicability of this type of strictly parallel model. We found that the detection of thematic role violations does indeed have a delayed intercept relative to the detection of either category or subcategorization violations (Figure 5B). On this basis, we can reject the class of models in which syntactic and thematic role processing is arranged in a strict parallel fashion such that the output of each process is available at the same time.

Next we consider two parallel architectures that are compatible with our results. McClelland (1979) proposed a hybrid architecture, dubbed *the cascade model,* in which individual processing stages are arranged in a temporal sequence, with one stage starting before another, but with information (activation) continually flowing in a feed-forward fashion. The cascade model posits a limited form of parallelism in that processing in the $n + 1$ stage is initiated as soon as (partial) output is received from the nth stage. This model differs from a strict serial model in which processing in the $n + 1$ stage begins only after the final output is received from the nth stage. McClelland demonstrated that in a cascade arrangement the output from the $n + 1$ stage has a slower dynamics than the output from the nth stage, with a form not unlike the dynamics differences for detecting thematic role violations observed in Experiment 2. The temporal differences for thematic role and syntactic processing can be cast in terms of a cascade rather than a strict serial model with no loss of adequacy in fitting our data.

The delay in the detection of thematic role violations indicates that a thematic representation was established later than a syntactic representation. A cascaded linking of processing stages is one means of capturing this difference. However, we need not assume that thematic role processing was contingent on a first stage of syntactic processing to account for the pattern of data. A second type of parallel architecture that is compatible with our results is one in which thematic role and syntactic structures are independently processed in parallel, but the output from the former is delayed relative to the latter. Indeed, there are some grounds on which to argue that the processing requirements for each type of information may well result in substantial differences in output time. Syntactic parsing is often viewed as a process in which words are mapped onto a finite set of (nonterminal) symbols in a discrete fashion. At the most basic level, for example, a word may be either a noun or a verb, or both if it is lexically ambiguous. The mapping is discrete, nonoverlapping, and admits no degrees. The assignment of words to thematic roles appears to be a more graded, nondiscrete process. For example, consider the strings "Some people amused people/dogs/ants/trees/rocks." Clearly, membership in the experiencer role class for the verb *amuse* admits of degrees that may well depend upon a number of semantic and pragmatic factors (see Dowty, 1991, for arguments that thematic roles are more properly viewed as cluster concepts, or prototypes, than as discrete concepts). The mapping of words onto thematic roles may be slower than onto syntactic categories simply as a result of the more extensive information that needs to be processed to establish a fit.

Finally, we briefly consider how our data relate to constraint-satisfaction models that are currently being developed by Trueswell, Tanenhaus, and colleagues (Trueswell & Tanenhaus, 1994; Trueswell et al., 1993) and MacDonald and colleagues (MacDonald et al., 1993). These approaches have been proposed to account for how local ambiguities, in particular syntactic ambiguities (e.g., Examples 1 and 2) can be quickly resolved by thematic role, semantic–conceptual, and even discourse information. The claim of these models is that thematic role information (as well as other types of information) is used to quickly select among a number of possible syntactic analyses. In the abstract, such a claim is compatible with our data in as much as it suggests that syntactic representations are computed before thematic and other representations. However, we must await further specifications of the computational details of these models to clearly determine whether they are capable of precisely fitting the temporal differences documented in this article.

References

Abney, S. (1989). A computational model of human parsing. *Journal of Psycholinguistic Research, 18,* 129–144.

Adams, B. C., Clifton, C., Jr., & Mitchell, D. C. (1991 November). *Lexical guidance in sentence processing.* Poster Session Presented at the Annual Meeting of the Psychonomic Society, San Francisco, CA.

Bates, E., & MacWhinney, B. (1989). *Functionalism and the competition model.* Cambridge, England: Cambridge University Press.

Bever, T. G. (1970). The cognitive basis of linguistic structures. In J. R. Hayes (Ed.), *Cognition and the development of language* (pp. 279–352). New York: Wiley.

Boland, J., & Tanenhaus, M. K. (1991). The role of lexical representation in sentence processing. In G. B. Simpson (Ed.), *Understanding word and sentence* (pp. 331–366). Amsterdam: North-Holland.

Burgess, C. (1991). *Interaction of semantic, syntactic, and visual factors in syntactic ambiguity resolution.* Unpublished doctoral dissertation, University of Rochester.

Carlson, G. N., & Tanenhaus, M. K. (1988). Thematic roles and language comprehension. In W. Wilkens (Ed.), *Syntax and semantics: Thematic relations* (Vol. 21, pp. 263–300). San Diego, CA: Academic Press.

Chandler, J. P. (1969). Subroutine STEPIT—finds local minimum of a smooth function of several parameters. *Behavioral Science, 14,* 81–82.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

Chomsky, N., & Lasnik, H. (1991). Principles and parameters theory. In J. Jacobs, A. van Stechow, W. Sternefeld, & T. Venneman (Eds.), *Syntax: An international handbook of contemporary research* (pp. 506–509). Berlin: Walter de Gruyter.

Clifton, C., Jr., & Ferreira, F. (1987). Modularity in sentence comprehension. In J. L. Garfield (Ed.), *Modularity in knowledge representation and natural-language understanding* (pp. 277–290). Cambridge, MA: MIT Press.

Clifton, C., Jr., Frazier, L., & Connine, C. (1984). Lexical expectations in sentence processing. *Journal of Verbal Learning and Verbal Behavior, 23,* 696–708.

Clifton, C., Jr., Speer, S., & Abney, S. (1991). Parsing arguments: Phrase structure and argument structure as determinants of initial parsing decisions. *Journal of Memory and Language, 30,* 251–271.

Dosher, B. A. (1976). The retrieval of sentences from memory: A speed–accuracy study. *Cognitive Psychology, 8,* 291–310.

Dosher, B. A. (1979). Empirical approaches to information processing: Speed–accuracy trade-off or reaction time. *Acta Psychologica, 43,* 347–359.

Dosher, B. A. (1981). The effect of delay and interference: A speed–accuracy study. *Cognitive Psychology, 13,* 551–582.

Dosher, B. A. (1982). Effect of sentence size and network distance on retrieval speed. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8,* 173–207.

Dosher, B. A. (1984). Degree of learning and retrieval speed: Study time and multiple exposures. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 541–574.

Dosher, B. A., McElree, B., Hood, R. M., & Rosedale, G. R. (1989). Retrieval dynamics of priming in recognition memory: Bias and discrimination analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 868–886.

Dosher, B. A., & Rosedale, G. (1989). Integrated retrieval cues as a mechanism for priming in retrieval from memory. *Journal of Experimental Psychology: General, 118,* 191–211.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language, 67,* 547–619.

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

Ferreira, F., & Henderson, J. M. (1990). Use of verb information in syntactic parsing: Evidence from eye movements and word-by-word self-paced reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 555–568.

Ferreira, F., & Henderson, J. M. (1991). How is verb information used during syntactic processing? In G. B. Simpson (Ed.), *Understanding word and sentence* (pp. 305–330). Amsterdam: North-Holland.

Ford, M., Bresnan, J., & Kaplan, R. M. (1982). A competence-based theory of syntactic closure. In J. Bresnan (Ed.), *The mental representation of grammatical relations* (pp. 727–796). Cambridge, MA: MIT Press.

Frazier, L. (1978). *On comprehending sentences: Syntactic parsing strategies.* Unpublished doctoral dissertation, University of Connecticut.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention & performance XII: The psychology of reading* (pp. 601–681). Hillsdale, NJ: Erlbaum.

Frazier, L. (1989). Against lexical generation of syntax. In W. D. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 505–528). Cambridge, MA: MIT Press.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14,* 178–210.

Grimshaw, J. (1990). *Argument structure.* Cambridge, MA: MIT Press.

Judd, C. M., & McClelland, G. H. (1989). *Data analysis: A model-comparison approach.* San Diego, CA: Harcourt Brace Jovanovich.

Kaplan, R. M., & Bresnan, J. (1982). Lexical functional grammar: A formal system for grammatical representation. In J. Bresnan (Ed.), *The mental representation of grammatical relations* (pp. 173–281). Cambridge, MA: MIT Press.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization.* New York: Oxford University Press.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide.* Cambridge, England: Cambridge University Press.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review, 86,* 287–330.

McClelland, J. L., St. John, M., & Taraban, R. (1989). Sentence comprehension: A parallel distributed processing approach. *Language and Cognitive Processes, 4,* 287–336.

McDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes, 9,* 157–201.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1993). *The lexical nature of syntactic ambiguity resolution* (Cognitive Science Technical Rep. No. UIUC-BI-CS-93-13). Chicago, IL: University of Illinois, The Beckman Institute.

McElree, B. (1993). The locus of lexical preference effects in sentence comprehension: A time-course analysis. *Journal of Memory and Language, 32,* 536–571.

McElree, B., & Dosher, B. A. (1989). Serial position and set size in short-term memory: The time course of recognition. *Journal of Experimental Psychology: General, 118,* 346–373.

McElree, B., & Dosher, B. A. (1993). Serial retrieval processes in the recovery of order information. *Journal of Experimental Psychology: General, 122,* 291–315.

Meyer, D. E., Irwin, D. E., Osman, A. M., & Kounois, J. (1988). The dynamics of cognition and action: Mental processes inferred from speed–accuracy decomposition. *Psychological Review, 95,* 183–237.

Mitchell, D. C. (1987a). Lexical guidance in human parsing: Locus and processing characteristics. In M. Coltheart (Ed.), *Attention & performance XII: The psychology of reading* (pp. 601–618). Hillsdale, NJ: Erlbaum.

Mitchell, D. C. (1987b). Reading and syntactic analysis. In J. Beech & A. Colley (Eds.), *Cognitive approaches to reading* (pp. 87–112). New York: Wiley.

Mitchell, D. C. (1989). Verb-guidance and other lexical effects in parsing. *Language and Cognitive Processes, 4,* 123–154.

Mitchell, D. C. (1994). Sentence parsing. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics.* New York: Academic Press.

Mitchell, D. C., Corley, M. M., & Garnham, A. (1992). Effects of context in human sentence parsing: Evidence against a discourse-based proposal mechanism. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 69–88.

Murdock, B. B., Jr. (1971). A parallel processing model for scanning. *Perception and Psychophysics, 10,* 289–291.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31,* 785–806.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 59–108.

Ratcliff, R. (1980). A note on modelling accumulation of information when the rate changes over time. *Journal of Mathematical Psychology, 21,* 178–184.

Ratcliff, R. (1988). Continuous versus discrete information processing: Modeling accumulation of partial information. *Psychological Review, 95,* 238–255.

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior, 22,* 358–374.

Rayner, K., Garrod, S., & Perfetti, C. A. (1992). During parsing discourse influences are delayed. *Cognition, 45,* 109–139.

Reed, A. V. (1973). Speed–accuracy trade-off in recognition memory. *Science, 181,* 574–576.

Reed, A. V. (1976). The time course of recognition in human memory. *Memory & Cognition, 4,* 16–30.

Reilly, R. G., & Sharkey, N. (1992). *Connectionist approaches to natural language processing.* Hillsdale, NJ: Erlbaum.

Schmauder, A. R. (1991). Argument structure frames: A lexical complexity metric? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 49–65.

Shapiro, L. P., Nagel, H. N., & Levine, B. A. (1993). Preferences for a verb's complement and their use in sentence processing. *Journal of Memory and Language, 32,* 96–114.

Shapiro, L. P., Zurif, E. B., & Grimshaw, J. (1987). Sentence processing and the mental representation of verbs. *Cognition, 27,* 219–246.

Shapiro, L. P., Zurif, E. B., & Grimshaw, J. (1989). Verb processing during sentence comprehension: Contextual impenetrability. *Journal of Psycholinguistic Research, 18,* 223–243.

Sharkey, N. (1992). *Connectionist natural language processing.* Norwell, MA: Kluwer Academic.

Stowe, L. (1989). Thematic structures and sentence comprehension. In G. N. Carlson & M. K. Tanenhaus (Eds.), *Linguistic structures in language processing* (pp. 318–357). Dordrecht: Kluwer Academic.

Taraban, R., & McClelland, J. (1988). Constituent attachment and thematic role assignments in sentence processing: Influences of content-based expectations. *Journal of Memory and Language, 27,* 1–36.

Townsend, J. T., & Ashby, F. G. (1983). *The stochastic modeling of elementary psychological processes.* Cambridge, England: Cambridge University Press.

Trueswell, J. C., & Tanenhaus, M. K. (1994). Towards a lexicalist framework for constraint-based syntactic ambiguity resolution. In C. Clifton, K. Rayner, & L. Frazier (Eds.), *Perspectives on sentence processing* (pp. 155–179). Hillsdale, NJ: Erlbaum.

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

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 528–553.

Waltz, D. L., & Pollack, J. B. (1984). Massively parallel parsing: A strongly interactive model of natural language interpretation. *Cognitive Science, 9,* 51–74.

Wickelgren, W. (1977). Speed–accuracy tradeoff and information processing dynamics. *Acta Psychologica, 41,* 67–85.

Appendix

Example Materials Used in Experiments 1 and 2

The full set of materials consisted of 10 sets of 448 sentences, with each set made up of 28 sentences from each of the 16 types illustrated in Table 1. The full set of materials is available from Brian McElree. This Appendix lists three additional examples of each of the 16 sentence types.

Active, Thematic Role Violation

Some students infuriate exams.
Some senators offend elections.
Some beggars amuse handouts.

Active, Subcategorization Violation

Some students laugh exams.
Some senators roar elections.
Some beggars chatter handouts.

Active, Constituent Structure Violation

Some students commonly exams.
Some senators repeatedly elections.
Some beggars frequently handouts.

Passive, Thematic Role Violation

Some students were loved by exams.
Some senators were prized by elections.
Some beggars were adored by handouts.

Passive, Subcategorization Violation

Some students were laughed by exams.
Some senators were roared by elections.
Some beggars were chattered by voters.

Passive, Constituent Structure Violation

Some students were commonly at laughed.
Some senators were repeatedly at roared.
Some beggars were frequently at chattered.

Additional Active, Ungrammatical Control String

Some exams love students.
Some elections prize senators.
Some handouts adore beggars.

Additional Passive, Ungrammatical Control String

Some exams were infuriated by students.
Some elections were offended by senators.
Some handouts were amused by beggars.

Active, Grammatical Control String for the Active Thematic Role Violations

Some exams infuriate students.
Some elections offend senators.
Some handouts amuse beggars.

Active, Grammatical Control String for the Active Subcategorization Violations

Some students laugh commonly.
Some senators roar repeatedly.
Some beggars chatter frequently.

Active, Grammatical Control String for the Active Constituent Structure Violations

Some students commonly laugh.
Some senators repeatedly roar.
Some beggars frequently chatter.

Passive, Grammatical Control String for the Passive Thematic Role Violations

Some exams were loved by students.
Some elections were prized by senators.
Some handouts were adored by beggars.

Passive, Grammatical Control String for the Passive Subcategorization Violations

Some students were laughed at commonly.
Some senators were roared at repeatedly.
Some beggars were chattered at frequently.

Passive, Grammatical Control String for the Passive Constituent Structure Violations

Some students were commonly laughed at.
Some senators were repeatedly roared at.
Some beggars were frequently chattered at.

Additional Active, Grammatical Control String

Some students love exams.
Some senators prize elections.
Some beggars adore handouts.

Additional Passive, Grammatical Control String

Some students were infuriated by exams.
Some senators were offended by elections.
Some beggars were amused by handouts.

Received October 4, 1993
Revision received March 31, 1994
Accepted April 1, 1994 ■