The Time Course of Grammaticality Judgment

Arshavir Blackwell   Elizabeth Bates   Dan Fisher

University of California, San Diego

Language and Cognitive Process, 11(4), August, 1996, 337-406

ABSTRACT. Three experiments investigating the time course of grammaticality judgment are presented, using sentences that vary in error type (agreement, transposition, omission of function words), part of speech (auxiliaries vs. determiners) and location (early vs. late error placement). Experiment 1 is a word-by-word cloze experiment in which subjects are presented with successively longer fragments of a sentence and instructed to complete the sentence grammatically, if possible. Experiment 2 is a self-paced, word-by-word grammaticality judgment experiment. Results of both these experiments are quite similar, showing that some error types elicit a broad and variable "decision region" instead of a more punctate "decision point." To explore the implications of this finding, Experiment 3 looks at on-line judgments of the same stimuli in an RSVP paradigm, with a single response (and reaction time). Correlations amongst the three experiments are extremely high and all significant, suggesting that the incremental tasks are tapping into the same decision-making process as is found on-line. Implications of these findings for the error types that do and do not appear in aphasia are discussed.

INTRODUCTION

Halfway through the twentieth century, linguistics underwent a major methodological shift, from distributional analysis of native-speaker speech (Bloomfield, 1961), to the analysis of native-speaker intuitions about legal sentence types (Chomsky, 1957; for reviews, see Newmeyer, 1980; Sells, Shieber & Wasow, 1991). In most cases, the native speakers who furnish these intuitions have been linguists trained to detect subtle structural facts that may not be obvious to laymen confronted with the same sentence stimuli. As a result, the conclusions reached by linguists do not always match the conclusions that one might draw if analyses were based on grammaticality judgments by naive listeners (for a detailed discussion of this point, see Levelt, 1974). This is a perfectly legitimate reason for linguists to keep their judgments in-house. However, it

This research was supported by an award from NIH/NIDCD 2 R01 DC00216-10 to Elizabeth Bates, "Cross-linguistic studies in aphasia." We are grateful to Jeff Elman, Judith Goodman, Mark St. John, and Marta Kutas for comments on an earlier draft of this paper. Please send correspondence to: Arshavir Blackwell, Center for Research in Language, University of California, San Diego, La Jolla, CA 92093-0526. E-mail: arshavir@andrew.cmu.edu

is not a good reason for psycholinguists to avoid the study of grammaticality judgment as a processing domain.

Because judgments of well-formedness lie at the heart of one of the most important movements in modern cognitive science, it would be useful if we could learn more about the nature and time course of this psychological process. What psycholinguistic phenomena may influence such metalinguistic judgments? (e.g., Levelt, 1972 1974, 1977). This is sufficient rationale for explorations of grammaticality judgment as a psychological process (in naive as well as expert subjects), although there are other reasons why this performance domain should be studied in more detail. For example:

Aphasia: Grammaticality judgments have played an increasingly important role in research on language breakdown in aphasia (Caplan, 1981; Caramazza & Berndt, 1985; Caramazza & Zurif, 1976), where one continuing puzzle has been that if these patients suffer from a deficit in the on-line activation of grammar, why are they able to make reasonably good judgments of grammaticality in on-line studies (for details, see Linebarger, Schwartz & Saffran, 1983; Shankweiler, Crain, Gorrell & Tuller, 1989; Wulfeck & Bates, 1991; Wulfeck, 1987; Tyler, 1992)? To answer this question we need more information about normal on-line grammaticality judgment.

ERP studies: Studies of event related brain potentials (ERP) of subjects exposed to linguistic stimuli have been used to draw a variety of conclusions about the language processor; e.g., that semantic processes and syntactic processes have at least partially separate biological components (Hagoort, Brown & Groothusen, 1993; Neville, Nicol, Barss, Forster & Garrett, 1991; Osterhout & Holcomb, 1993; Brown, Hagoort, & Vonk, 1995). On the whole, these types of studies have assumed a punctate point at which the sentence became ungrammatical, and thus compared ERPs at that only one point between the ungrammatical and grammatical control sentences.[1] Because these studies often use a word-by-word grammaticality judgment paradigm similar to those we use (i.e., subjects read a sentence one word at a time while their ERPs are recorded), knowing more about the nature of this psychological process may offer new insights into what is happening in these experiments, and thus perhaps provide alternative interpretations of the results.

The Experiments

Experiment 1 ascertains what sorts of grammatical completions subjects entertain as the sentence unfolds. Subjects are asked to provide a possible grammatical completion of the sentence at each word. This *cloze experiment* should yield valuable information about the number, range and strength of the alternative completions that subjects may have in mind at each point across the course of the sentence.

Experiment 2 is a self-paced, word-by-word reading task where, after each word appears, subjects press one of three buttons (“grammatical”, “ungrammatical”, “not sure”), indicating their judgment of the grammaticality of the sentence to that point. We expect some sentence stimuli to yield a sharp boundary after which most subjects agree that the sentence cannot be salvaged (i.e., there is no well-formed way for it to continue). We term this a “*decision point*.” However, other sentence stimuli may yield a decision-making region that spans several words. We term this a “*decision region*.” Furthermore, subjects may show marked individual differences in the size of this decision region, and the speed with which decisions are made at each point within that region. The elicitation of word-by-word grammaticality judgments bears a clear relationship to other word-by-word techniques in the visual modality (e.g., Just & Carpenter, 1980; Rayner, Carlson & Frazier, 1983; see also Boland, Tanenhaus, Carlson & Garnsey, 1989; Boland, Tanenhaus & Garnsey, 1990; Mauner, 1992).

As we shall demonstrate below, our technique will elicit some of the classic effects reported by authors using these three paradigms. Finally, our technique is related to recent studies of sentence processing (including sentences with grammatical violations) using event-related brain potentials as the primary dependent variable (Kutas & Kluender, 1991; Hagoort et al., 1993; Neville et al., 1991; Osterhout & Holcomb, 1993). However, our paradigm requires conscious judgments of grammaticality at every time point, whereas the ERP technique can be used to detect response to violations with no explicit task other than reading or listening. In Experiment 3, the same sentence stimuli are used in a simple reaction time study, where subjects are asked to push the button once for each sentence as soon as they know whether that sentence is grammatical or not. As we shall see, any conclusions that can be drawn about the time course of grammaticality judgment will depend crucially on the point that is used to define the onset of the error, a finding that presents an interesting challenge to research programs that assume a single violation point.

GENERAL METHOD

Grammaticality Judgment Stimuli for All 3 Experiments

Ungrammatical targets. Stimuli were 168 sentences: 84 ungrammatical target sentences, 40 grammatical control sentences matched for length and grammatical structure, and 44 distractors (see below). Experimental design focused on the ungrammatical targets, which varied in: a) part of speech of the error (auxiliary vs. determiner); b) the position of the error (early or late in the sentence), and, most importantly, c) type of violation (i.e., errors of omission, agreement and transposition). Thus, the ungrammatical target sentences formed a 2 × 2 × 3 design, with part of speech, location, and error type as within-subject variables.

Grammatical controls. Each of the twelve cells in the design had seven ungrammatical sentences. For each of these ungrammatical sentences, there was a grammatical control sentence matched for length and grammatical structure. To keep the experiment reasonably short, some grammatical sentences were used as controls for more than one particular ungrammatical sentence. There were also 44 distractor sentences (22 grammatical and 22 ungrammatical) from 3 to 17 words long, and of various structures. Distractors were to prevent subjects from detecting regularities in the length and nature of the target sentences (see Appendix I for all stimuli).

Creation of ungrammatical targets. The 84 ungrammatical targets and 40 grammatical controls come from a pool of grammatical sentences from 8 to 12 words long. This pool of sentences represents a range of seven structural types, varying in presence and location of prepositional phrases, presence or absence of relative clause or subordinate clauses, and the number of adjectives modifying the subject and object (see Appendix II). Approximately twenty different sentence tokens were constructed for each of these seven structural types, and randomly assigned to the appropriate ungrammatical target cell or grammatical control condition. Half of the sentences in this pool had at least one auxiliary verb to be the target of an auxiliary violation, while the other half of sentences had at least one determiner (including numerals and demonstrative adjectives) to be the target of a determiner violation:

> *Auxiliary verb sentences:* On half of these items, the auxiliary was located early in the sentence (e.g., "They **were** reading several large maps while waiting for the next train."), while on the other half, the auxiliary was located near the end of the sentence (e.g., "In a big, old, red boat, two girls **were** rowing slowly.").
>
> *Determiner sentences:* On half of these items, the target determiner was located early in the sentence (e.g., "**The** girl was eating some dark chocolate ice cream."), while on the other half, the

target determiner was located near the end of the sentence (e.g., “My new blue and green silk ball gown was costing **a** fortune.”).

Location of error. Early errors occurred within the first 1200 msec (milliseconds) of the sentence (in the RSVP task), while late errors occurred after this point. The licensing word and the error were always adjacent (i.e., all local errors). Thus, we used errors such as “The girl were * going,” or “A girls * were going” (where the error was caused by the wrong juxtaposition of two directly adjacent words) but not “A large black-and-white dogs were going” (where the mismatch is between “A” at the beginning of the sentence and “dogs” several words downstream). Because omission, agreement and transposition errors were created from the same basic sentence types, it can be argued that these stimuli represent a set of minimal contrasts. Nevertheless, even within a well-controlled stimulus set, there are complicating factors affecting our interpretation, to which we now turn.

Rationale for the Stimulus Materials

In designing stimuli for grammaticality judgment, the experimenter has two choices: create grammatical deformations which cleave along the lines of some linguistically motivated theory (usually but not necessarily Generative Grammar; e.g., Kluender, 1992; Linebarger et al., 1983) or create sentences whose ill formedness is agnostic as to particular linguistic theory, and is motivated by an empirical demonstration of some psycholinguistic differences between the error types (e.g., Wulfeck & Bates, 1991; Wulfeck, Bates & Capasso, 1991; Wulfeck, 1987). We have opted for the second strategy. Our choice of materials for these studies is motivated (at least in part) by recent research on grammatical breakdown in aphasia. In particular, we know that some error types (i.e., omission and/or substitution of functors) are very common in speech production by aphasic patients. Other error types (i.e., word order violations like “dog the” or morpheme order violations like “ing-kiss”) are exceedingly rare (Bates, Wulfeck & MacWhinney, 1991). One possible explanation for this sharp difference in the probability of error types might lie in the monitoring mechanism that normals and aphasics use to detect errors in their own speech and/or to weed out errors before they are produced. If normal listeners are particularly sensitive to word order errors, but less sensitive to errors of agreement and omission, then we may conclude that the monitoring device is less sensitive to errors of agreement and omission under pathological conditions. To test this hypothesis, we are building on an earlier grammaticality judgment study by Wulfeck & Bates (1991). Using auditory stimuli, these authors showed that normal English listeners are faster at detecting errors produced by moving a function word downstream from its normal position (e.g., “She is selling books…” * “She selling is books…”), compared with errors produced by substituting an incorrect form of the same function word within its usual position in the sentence (e.g., "She is selling books…" * "She are selling books…."). In the present study, we have expanded the set of violations used by Wulfeck et al. to include omission errors (e.g., "She is selling books…." * "She selling books…"). We have also moved to the visual modality (removing any cues to ungrammaticality that might be due to intonation and/or coarticulation), and added the cloze and incremental grammaticality judgment (GJ) experiments.

Rules for creating the three error types

Omission errors: remove the relevant word (auxiliary or determiner) from the sentence (see Table 1). The asterisk (never visible to subjects) refers to the aforementioned divergence point. Thus, for omission errors the divergence point is just after the word following the point where the omitted element should go.

An additional complication comes from the contrast between early and late omission errors. Because all of our sentences are marked with normal English punctuation, late omission errors often involve a double cue. For example, given a late auxiliary omission error such as, "While sitting on the red sofa, her older friend eating* some cake," the subject actually has two cues to help him decide whether an error has occurred. First, after reading the word "eating," the subject knows that a verb that should have been proceeded by an auxiliary was not. Second, because the word "eating" is soon followed by a period (visible at the end of every sentence stimulus), the subject may conclude that no further items will come along to salvage the sentence (e.g. the sentence will not turn into something such as, "While sitting on the red sofa, her older friend eating some cake was watching TV."). Hence we might argue that the above examples each provide the subject with two distinct error cues, illustrated as follows: "While sitting on the red sofa, her older friend eating * some cake. *"

Agreement errors: replace the target word with an item that doesn't agree in number. Note that violations of determiner agreement within a subject noun phrase provide two cues to the agreement violation. Cue one is the mismatch in number between determiner and noun (i.e., "a girls *"); cue two is the mismatch between the auxiliary verb and the determiner (the auxiliary verb can only agree with one of the two elements within the subject noun phrase, either locally with the preceding noun, or globally with the determiner). This situation can be symbolized as, "A girls * were * working quietly near the small, red house." The divergence point is just after the noun (for determiner errors) or verb (for auxiliary errors) which licenses the element that is in error.

Transposition errors: move the relevant word one word downstream from where it belongs. The divergence point is just after the word following where the moved element should go and before where the moved element actually is. This matches the divergence point for omissions and is the first point at which the subject might notice that a potential element is missing (although see the note above about this). This suspicion will, of course, be confirmed when the subject encounters the displaced element. Hence transposition errors constitute another instance in which there are really two cues to the existence of an error, one at the first point at which a subject might notice that there is a hole (similar to omission errors) and another at the point further downstream where the displaced element occurs.

Late errors. There is one further difference between late omission errors and the other two late error types: On transposition errors, the moved element means that the sentence will necessarily last one word longer after the divergence point than it does with errors of omission or agreement. Because the three error types share the same divergence point (i.e., they start to deviate at exactly the same point in the sentence), this need not constitute a problem. However, if subjects cannot make up their minds at the divergence point and want to wait for more information before they decide, then we are faced with an artifact: Subjects are forced to make up their minds at the divergence point on many late agreement and omission errors, because the sentence is already over (as indicated by a period—see Appendix I); by contrast, they are able to delay their decisions for a while on the transposition errors. Hence any differences that we may observe in the size of the decision region for late errors may be a by-product of unavoidable structural differences among the three late-violation types. For this reason, all analyses of timing and decision points will be conducted separately for early vs. late errors.

Variability within types. This design has violation points for what is putatively the "same" error not necessarily always at the same structural point, as shown in sentences 1.1 and 1.2 above. This leaves us open to a potential criticism, that we are creating our effects by artificially choosing some arbitrary point (the divergence point) where subjects "should" detect the error, and then demonstrating that they do not necessarily detect the error at that point. We must again stress that ***the divergence point is not necessarily where subjects will first detect an error***, though experimental subjects will certainly never detect an error (correctly) earlier than the divergence point. The divergence point is **a point structurally common** across the various error types and items (to the extent possible), as well as being the "origination point" of all of the error deformations (as

Table 1 shows). It is an empirical question *where* subjects will make their decisions, and, of course, that is part of what we are investigating. The diversity of sentences with the "same" error type is deliberate, and a strength of these stimuli, as they directly map onto the error types that we are investigating. These are error types which, as stated above, do appear to have some kind of psychological reality. Thus, for example, Wulfeck (1987) reported differential sensitivity to transposition and agreement errors, not to, e.g., transposition errors of only one certain type. Certainly that leaves open whether errors of a certain type are hewn from one homogeneous kind. However, we would argue that the proper first step is to examine more complete, albeit variegated, sets of each of the various error types, as that is what we know—at this point—to have psychological reality, rather than to cleave off and examine only sub-types of these various errors.

Stimulus design considerations

For a particular error type (e.g., transposition errors on determiners early in the sentence) the ungrammaticality does not necessarily begin at the same structural point (e.g., directly after the auxiliary verb or determiner), yet it is this structural point that the sentences have in common. For example, both sentence 1.1 and 1.2 are early determiner transposition errors, yet we have found that for sentences such as 1.1, subjects tend to indicate that the ungrammaticality occurs just after the transposed determiner "the", while they judge 1.2 as ungrammatical after the first word *after* the transposed determiner (in this case "are"), entertaining completions such as "Women three hundred years ago were the subject of oppression."

1.1 Women the * + are walking to the store

1.2 Women three * are + walking to the store.

Our approach to this issue is to let the subjects decide where the error begins (i.e., this is an empirical question), locking all sentences within a particular class to a common divergence point, defined operationally as the point at which ungrammatical and grammatical sentences of a particular type differ due to the violations that we have imposed (see Table 1).

Reading-span test: One technique we used to attempt to account for individual differences is the "reading span" test (Carpenter & Just, 1989; Daneman & Carpenter, 1980; Just & Carpenter, 1992). However, the test had little to tell us about the results in these experiments, and for the sake of brevity it is not reported on here.

Table 1. Grammaticality judgment stimuli

auxiliary verbs	omission	Joan [was] making * several big and tasty ice cream drinks.
	agreement	Joan were * making several big and tasty ice cream drinks.
	transposition	Joan [ ] making * was several big and tasty ice cream drinks.
determiners	omission	[A] Boy * is driving a large van that the artist has painted.
	agreement	Those boy * is driving a large van that the artist has painted.
	transposition	[ ] Boy * a is driving a large van that the artist has painted.

EXPERIMENT 1: Cloze

Method

Subjects. Ten college students (all right-handed) participated in the experiment for course credit and payment. All subjects were native English speakers, with little if any facility in any other languages.

Stimuli. See “General Method”.

Equipment. Each sentence was presented one word at a time, using an IBM-PC/XT with a GoldStar 1210A amber screen monitor. Subjects’ spoken response was recorded on a Marantz PMD201 tape recorder, using a Beyer-Dynamic Soundstar MK-II microphone. Subjects also responded using a Carnegie-Mellon button box. Subjects responded with one of two button presses: “good,” (meaning that they completed the sentence grammatically), or “bad” (meaning that they could not complete the sentence grammatically).

Procedure. A trial began with a “READY” cue appearing near the bottom center of the screen. The subject pressed the middle button to bring the first word of the sentence to the screen. Subjects were instructed to use the index finger of their dominant hand.

The sentence was centered vertically and started at the left side of the screen. Each button press brought the next word onto the screen, until the entire sentence was visible. After the last word appeared the button press caused the next “READY” cue to appear.

The experimenter instructed subjects to try to complete, aloud, the sentence as read so far, and to press the “good” button if they did so. They were told that “any grammatical sentence is acceptable as a completion.” Subjects were instructed that if there was “*no way* to finish the sentence grammatically,” they were to say “can’t complete” and press the “bad” button. Subjects were instructed to read the entire sentence aloud, rather than merely their completion. The experimenter told subjects that once they believed that the sentence could not be completed grammatically, they should continue with the "can't complete" response if they continued to believe that the sentence could no longer be completed—even if the remainder of that sentence seemed well-formed. They were instructed to complete the sentence only if they could generate a complete, grammatically correct sentence. During the instruction phase, some subjects asked the experimenter whether a particular practice item was correct or incorrect. When this occurred, subjects were again told to base their responses on what they themselves considered to be correct grammar. When the entire sentence was on the screen (including the period), subjects were instructed to read it aloud if they believed it to be grammatical, and press the "good" button, or, if they thought it not grammatical, to again say "can't complete" and press the "bad" button.

The actual experiment consisted of 168 trials using the sentence stimuli described in the "General Methods" section. Each subject received the sentences in a different random order, determined by the computer program. Subjects were told that they would receive a break at the mid-point of the experiment (this was after trial 84). At this point, instead of the "READY" cue, the subject received a "PLEASE WAIT" cue.

Subjects were given ten to twenty practice sentences of similar kind before the actual experiment, depending upon how clearly they understood the task.

Scoring: Both the point at which subjects first said "can't complete" and the sorts of grammatical completions subjects gave up until that point were transcribed. Our primary dependent measure is the mean number of words past the divergence point that subjects first said they could not complete the sentence grammatically.

Results

Overall performance for non-filler sentences

The statistics we report on are only for the "core stimuli" or non-filler sentences. Almost all of the stimulus sentences were correctly judged by the end of the sentence. Subjects had a mean hit rate to ungrammaticals of 96.8%, with only 2.8% false alarms, which is an A' of 98.5 (A' is a non-parametric statistic used to correct for response bias (Grier, 1971; Pollack & Norman, 1964). No individual subject A' was below 97.8.

By-item analyses

94.1% of the ungrammatical experimental stimuli were responded to correctly by the end of the sentence by at least 90% of all subjects, with all but one of the remaining items responded to correctly by at least 70% of all subjects. The one item which had a 40% correct rate (sentence #8.11) is dropped from further analysis.

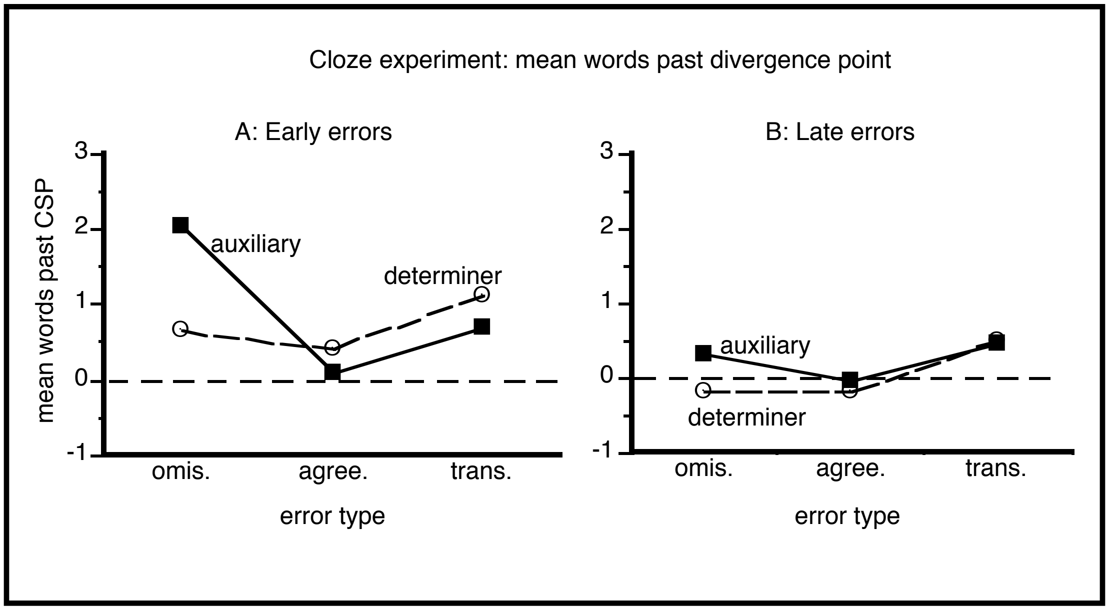

Figure 1. Cloze experiment: Mean number of words past the divergence point. omis. = omission; agree. = agreement; trans. = transposition

Analysis of variance

Subject responses were converted to a score indicating mean number of words past the divergence point at which subjects first gave a "cannot complete" response (that is, at which subjects could no longer generate a grammatical completion to the sentence as read so far). The data were submitted to two analyses of variance, one for early errors, the other for late errors. The within-subject factors were part of speech (auxiliary vs. determiner) and error type (omission, agreement, transposition), with subjects as the random factor. A parallel analysis by items, with the between-subjects factors of part of speech and error type, and with sentences as the random factor, is also presented.

Early errors: For early errors, both type (F1(2,18) = 5.69, $p < 0.0122$; F2(2,36) = 8.13; $p < 0.05$) and part of speech × type (F1(2,18) = 7.03, $p < 0.0055$; F2(2,36) = 6.62; $p < 0.05$) were significant. Agreement errors had a mean score of 0.24, transposition 0.91, and omission 1.36; a Newman-Keuls analysis showed agreement and omission to be significantly different from each other (by items, agreement < transposition = omission). A breakdown of the interaction, by part of speech, showed that for auxiliary errors, omission errors (2.06) were significantly higher than either transposition (0.71) or agreement (0.07), using Newman-Keuls. For determiner errors, transpositions (1.11) were significantly higher than omission (0.66) or agreement errors (0.40). See Figure 1A.

Table 2. Cloze Experiment: Percent sentences judged ungrammatical at divergence point

		early	late
auxiliary	omission	69.6	77.6
	agreement	97.1	100.0
	transposition	60.9	51.5
determiner	omission	56.5	100.0
	agreement	66.2	100.0
	transposition	11.4	30.8

Late errors: For late errors, type (F1(2,18) = 8.26, $p < 0.0029$; F2(2,35) = 9.66; $p < 0.05$) was significant, and part of speech × type (F1(2,18) = 3.80, $p < 0.05$; F2(2,35) = n.s.) was marginally significant. Agreement errors had a mean score of -0.11, omission 0.09, and transposition 0.49; a Newman-Keuls analysis showed transposition significantly different from the other two conditions (also by items). A breakdown of the interaction, by part of speech, showed that for auxiliary errors, agreement errors (-0.04) were significantly lower than either transposition (0.47) or omission (0.33), using Newman-Keuls. For determiner errors, transposition errors (0.50) were significantly higher than omissions (-0.16) or agreement errors -0.18). See Figure 1B.

What sorts of responses are subjects making?

The cloze experiment, besides allowing us to see at what point subjects can no longer generate a grammatical completion of a sentence, also permits us to ask what sorts of completions subjects are making at each point, when they still believe the sentence can be saved. Overall, 67.7% of correctly-responded-to ungrammaticals were deemed ungrammatical by the divergence point. Some responses fell into a miscellaneous category including sentences where subjects briefly (for a few words) changed their choice; e.g., giving a grammatical completion for several gates, saying at the next word “can't complete”, then continuing the grammatical completion. This occurred in roughly 3% of ungrammatical sentence responses, and is ignored in this analysis.

Table 2 shows the by-cell percentage of sentences judged ungrammatical at the divergence point. Here is a breakdown of the sorts of grammatical completions subjects provided when they continue to give a response after the divergence point; see Figure 2 for a graphical representation for the major categories. (we recognize that some completions may fall into more than one category; however, each sentence was only placed in one).

Early errors:

Auxiliary errors: For omissions, 69.6% were deemed ungrammatical ("can't complete") by the divergence point. The grammatical completions at or after that point were either:

- present-participial verb-phrase completion (e.g., "The boy taking [sentence fragment seen by subject]… *a black car is a criminal* [subject's completion]," 90.5%) or
- gerund + "that" clause completion ("Tom's mother forgetting…*that he had already packed his lunch began to pack his lunch*," 9.5%).

For agreement errors, 97.1% were deemed ungrammatical by the divergence point. The grammatical completions were all corrections of the existing grammatical error. For transposition errors, 60.9% were deemed ungrammatical by the divergence point. The grammatical completions were:

- present-participial verb-phrase completion (88.8%),
- gerund + "that" completion (7.4%), and
- use of noun as adjective ("Students writing… *is put in the offices of some elementary schools*," 3.7%; note that many of these types of completions involved subjects mistakenly using a plural noun as a possessive; recall that the stimuli are *visual*.)

Determiner errors: For omissions, 56.5% were deemed ungrammatical by the divergence point. The grammatical completions were:

- use of noun with copula ("Boy… is a term that is used with a condescending air," 30.0%),
- use of noun as title or proper noun ("Boy… *George is a very strange person*," 26.7%),
- use of noun as adjective ("Woman… *doctors are better than man doctors*," 20.0%),
- use of noun in a general sense, or to stand in for a group (Man… *is said to be God's greatest creation*," 13.3%), and
- noun as interjection ("Boy …*do I have a sore finger*," 10.0%).

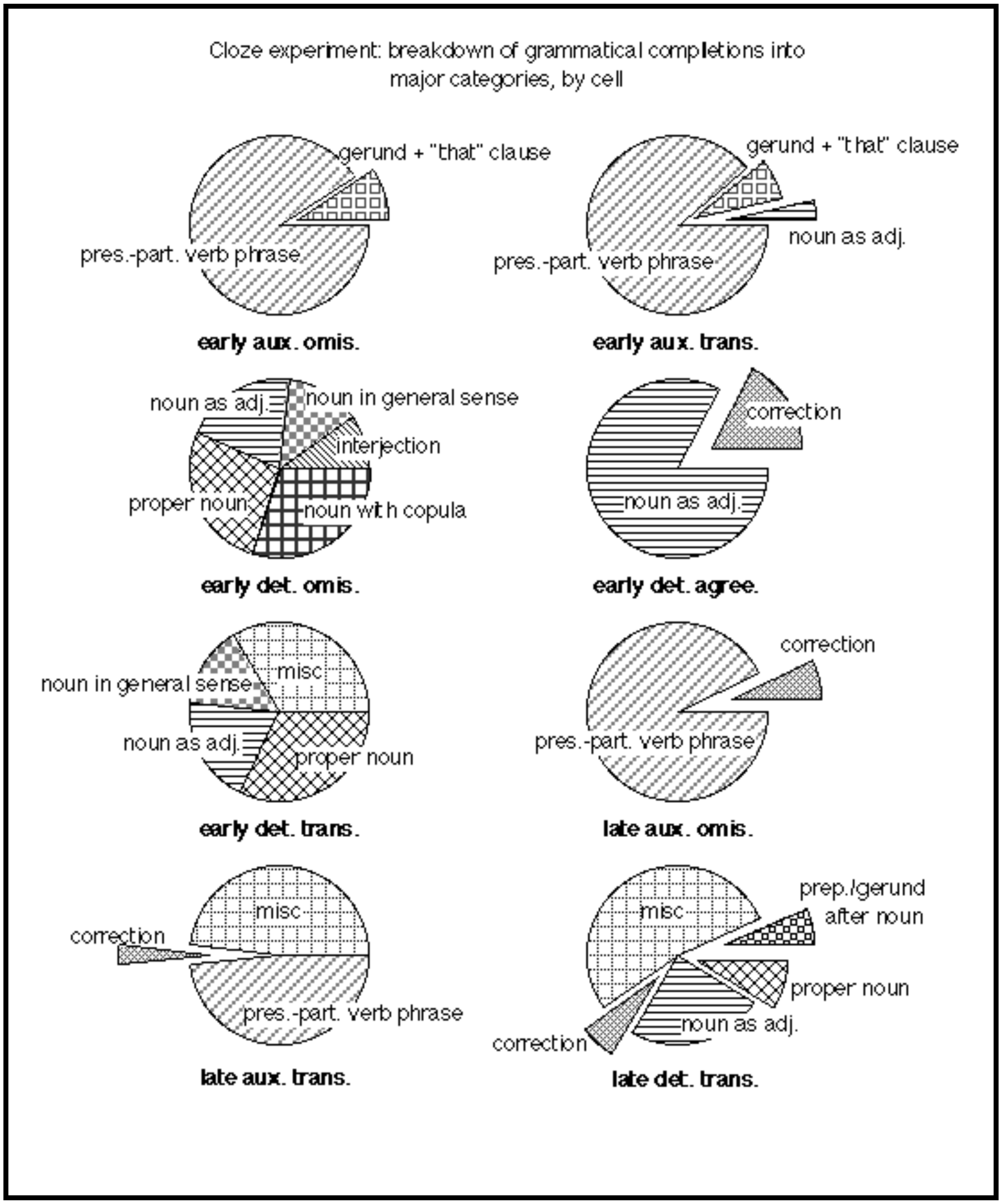

Figure 2. Cloze experiment: Breakdown of grammatical completions into major categories by cell.

For agreement errors, 66.2% were deemed ungrammatical by the divergence point. The grammatical completions were:

- use of noun as adjective ("Several sailor… *uniforms were in my bag,*" "A boys… *life is very simple*," 82.6%), and
- correction on the existing grammatical error (17.4%).

For transposition errors, 11.4% were deemed ungrammatical by the divergence point. The grammatical completions were:

- use of noun as title or proper noun ("Guest… *number three entered through the door*", "Announcer… *Chuck Hern is a very funny guy*," 32.3%),
- use of noun as adjective (19.4%),
- use of noun in a general sense (14.5%),
- use of displaced element as adjective following noun ("Women three… *decades ago did not have the same rights as they do today*," 8.1%),
- use of noun as interjection (3.2%),
- use of noun with copula (3.2%), and
- other grammatical completion following unmodified noun.

Late errors:

Auxiliary errors: For omissions, 77.6% were deemed ungrammatical by the divergence point. The grammatical completions were:

- present-participial verb-phrase completion (86.6%),
- corrections on the existing grammatical error (6.7%), and
- use of verb gerund as adjective ("The young, new president of John's college speaking… *school is an idiot*," 6.7%).

For agreement errors, 100% were deemed ungrammatical by the divergence point.

For transposition errors, 51.5% were deemed ungrammatical by the divergence point. The grammatical completions were:

- present-participial verb-phrase completion (42.4%),
- correction on the existing grammatical error (3.0%), and
- other grammatical completion following verb ("Those pilots were saying that several clouds covered… *the entire sky*," 54.5%).

Determiner errors: For late determiner errors, the divergence point for both omission and agreement errors was also the last word of the sentence; thus, 100% of the correct responses in this cell were by the divergence point, by necessity. For transposition errors (where there is one more element—the displaced determiner—*after* the divergence point) 30.8% were deemed ungrammatical by the divergence point. The grammatical completions were:

- use of noun as adjective (24.4%),
- use of noun as title or proper noun (8.9%),
- correction on the existing grammatical error (6.7%),
- Prepositional or gerundive phrase following unmodified noun ("The magazine reporter was donating one hundred dollars to hospitals… *treating AIDS*," 6.7%),
- reduced relative clause (2.2%), and
- other grammatical completion following unmodified noun ("George's remaining dinner guests were drinking wine… *and eating rolls*," 51.1%).

Summary of results for Experiment 1

Native speakers offer a range of alternative completions for the 12 error types employed in these experiments at or after the divergence point (i.e., the point at which the stimuli deviate from each other and from grammatical controls). These include many grammatical or (in some cases) semi-grammatical completions.

Early auxiliary errors. Subjects provided grammatical completions to early auxiliary omissions and transpositions at the divergence point an average of 35% of the time, less than for the corresponding early determiner errors (see below), but more than for early auxiliary agreement errors, for which subjects provided a grammatical completion at the divergence point only 3% of the time. For both early auxiliary omissions and transpositions, about 90% of all grammatical completions were present-participial verb-phrase completions such as, "Mrs. Brown[,] working at the library…"

Early determiner errors. Subjects provided grammatical completions to early determiner omissions (44%) and transpositions (88%) at the divergence point an average of 66% of the time, suggesting that to some extent they believed the sentence to be grammatical to that point in many cases, but that there was also some doubt. Subjects provided a variety of completions at this point for both error types, including use of the bare noun as proper noun or title (e.g., "President… *Clinton was briefed by his advisors.*"), use of noun in the general sense (e.g., "Man… *is a fragile creature.*"), and use of noun as adjective (e.g., "Woman… *doctors…*"). Subjects provided grammatical completions to early determiner agreement errors at the divergence point an average of 34% of the time, suggesting that fewer believed the sentence to be grammatical at that point compared to the other two early determiner error types. 83% of these early determiner agreement error completions involved the use of the bare noun as an adjective (e.g., "Several sailor… *uniforms were in my bag,*" "A boy[']s… *life is very simple,*"), including many completions where subjects mistakenly used a plural noun as a possessive.

Late errors. As mentioned above, the divergence point for both late determiner omission and agreement errors was also the last word of the sentence; thus, 100% of the correct responses in this cell had to be before or at the divergence point. Subjects provided grammatical completions to late determiner transpositions at the divergence point an average of 69% of the time, providing a variety of completions such as use of noun as adjective, use of noun as title or proper noun, correction of the grammatical error, and prepositional or gerundive phrase following unmodified noun. Subjects never provided grammatical completions to late auxiliary agreement errors at the divergence point—i.e., if a subject indicated that the sentence was ungrammatical on the last button press, they had indicated it by the divergence point. Subjects provided grammatical completions for the other two auxiliary error types an average of 35% of the time, with a large number of those corrections being present-participial verb-phrase completions.

To summarize, subjects were more likely to provide grammatical completions at the divergence point for errors appearing early in the sentence than for those appearing late, for omission and transposition errors than for agreement errors, and for early determiner errors than for early auxiliary errors.

EXPERIMENT 2: Incremental Grammaticality Judgment

Experiment 2 is a self-paced, word-by-word reading task where, after each word appears, subjects press one of three buttons (“grammatical”, “ungrammatical”, “not sure”), indicating their judgment of the grammaticality of the sentence to that point. We expect subjects’ judgment of grammaticality in this task to be quite consistent with the number and range of completions offered at each word in the cloze experiment.

Method

Subjects. Subjects were thirty-five college students (five left-handed; twenty-two female and thirteen male) who participated in the experiment for course credit, or for a payment of $7.00. All subjects stated that they were native speakers of English.

Stimuli. The stimuli were identical to those of Experiment 1.

Equipment. Each sentence was presented one word at a time, using an IBM-PC/XT with a GoldStar 1210A amber screen monitor. Subjects responded using a Carnegie-Mellon button box, accurate to one millisecond. Subjects responded with one of three button presses: “good,” “bad,” or “not sure.”

Procedure. A trial began with a “READY” cue appearing near the bottom center of the screen. The subject pressed the middle button, corresponding to “not sure,” to bring the first word of the sentence to the screen. Subjects were instructed to use the index finger of their dominant hand, and to keep the finger at a home spot beneath the middle key between button presses.

The sentence was centered vertically and started at the left side of the screen. Each button press brought the next word onto the screen, until the entire sentence was visible. After the last word appeared the button press caused the next "READY" cue to appear.

The experimenter instructed subjects to decide, after each word appeared upon the screen, whether the sentence up to that point was "grammatically correct." We did not elaborate upon what "grammatically correct" meant, and if subjects asked, we simply re-iterated that we wanted them to decide whether the sentence was grammatically correct or incorrect. The experimenter told subjects that, once they believed that the sentence had gone bad, they should continue pressing the "bad" button if they continued to believe that the sentence could no longer be saved—even if the remainder of that sentence seemed well formed. They were instructed to press the "good" button again only if they had changed their mind about the error. During the instruction phase, some subjects asked the experimenter whether a particular practice item was correct or incorrect. When this occurred, subjects were again told to base their responses on what they themselves considered to be correct grammar.

The actual experiment consisted of 168 trials of the same sentence stimuli as Experiment 1. Each subject received the sentences in a different random order, determined by the controlling computer program. Subjects were told that they would receive a break at the mid-point of the experiment (after trial 84). At this point, instead of the "READY" cue, the subject received a "PLEASE WAIT" cue.

Subjects were given twenty practice sentences before the actual experiment.

Both button presses and reaction time were collected. Reaction time was measured from the onset of the current word to the button press.

Scoring. A button press was recorded for every word of every sentence. Reaction time to each word was also recorded. The following dependent variables were derived from these data:

1. Final button press (a measure of overall accuracy);
2. Normalized word-by-word button press (explained below), to determine the shape of the decision function for each item type;
3. Normalized word-by-word reaction time, a complementary measure of the shape of the decision function. This included only reaction times for button presses before an "ungrammatical" response was made—i.e.,

only button presses where subjects were still making a decision about ungrammaticality.

Because individual sentences varied in length, and we wished to compare several different points across different sentences, word-by-word data were temporally normalized (or aligned) in the following way (see Figure 3): The first button press of each sentence was synchronized at “first,” the last button press at “last.” The divergence point is labeled “zero” on the graphs. In those cases where a sentence either began or ended on the divergence point, this point was synchronized at zero and not at first or last. Words in between the first point and the divergence point, and between the divergence point and the last button press, were binned and averaged within the bins. For early errors, there was one bin between the first word and the divergence point, corresponding to all words between (and not including) the first button press and the divergence point (in fact, this bin existed for early auxiliary but not early determiner errors, because there were no words between the first word and the divergence point for early determiner errors). After the divergence point, data were binned into a “20%” interval (corresponding to the first 0-20% of the sentence past the divergence point), a “40%” interval (corresponding to the first 20-40% of the sentence past the divergence point), and so on. Because each sentence was from eight to twelve words in length, each bin roughly corresponds to one word. The scheme for late errors was similar: The first bin corresponds to the first button press, followed by the “—80%” interval (the first 100-80% of the sentence before the divergence point, excluding the first word), the “—60%” interval and so on.

Final button press refers to the judgments obtained on the last button pressed for ungrammatical sentences. The final button press was evaluated using A’ to grammaticals and ungrammaticals combined. As we noted above, A’ is a non-parametric statistic used to correct for response bias (Grier, 1971; Pollack & Norman, 1964). As such, it is similar to d’. Raw percent correct scores for grammatical and ungrammatical stimuli do not account for the possibility of subject response bias. For example, a score of 100 for ungrammatical stimuli (all ungrammatical stimuli correctly identified) could mean that the subject is perfect at the task—or simply that the subject has an overwhelming tendency to guess that a sentence is ungrammatical. This cannot be determined without looking at both hits and false alarms. The above subject might also have a false-alarm rate of 100, indicating that in fact they are incapable of differentiating grammatical from ungrammatical sentences and judge everything as ungrammatical. Conversely, a false-alarm rate of 0.00 (with a hit rate of 100) would constitute perfect performance. A’ is a unified statistic that corresponds to the underlying percent correct in a two-option forced

early errors

Mrs.	Brown	working *	quietly	in	the	church		kitchen.
She		signing *	was	her	newest	and	biggest story	collection.
		Girl *	was	eating	some	dark	chocolate ice	cream.
first	>0%	zero	<20%	<40%	<60%	<80%	<100%	last

late errors

A	small	and	harmless	black	dog	chasing*	was	chickens.
The	magazine	reporter was	donating one	hundred	dollars to	hospitals*		those.
first	>80%	>60%	>40%	>20%	>0%	zero	<0%	last

Figure 3. Incremental grammaticality judgment bins

choice task, *correcting* for bias. It can range from 50.0 (chance performance) to 100.0 (perfect discrimination). Of course, normal subjects should not show such strong biases, but A' still permits discrimination of subject accuracy differences that are more subtle; for example, which subject is more accurate, one with 90% hit rate and a 1% false-alarm rate, or one with a 99% hit rate and a 10% false-alarm rate? (for details, see Pollack & Norman, 1964).

For A', *all* subject responses for target stimuli were used (i.e., ungrammaticals and their grammatical controls). For each of the twelve cell conditions and for each subject, a signal-detection matrix was generated, and "hits" (correct judgments of ungrammaticality) and "false alarms" (incorrect decisions that a sentence is ungrammatical) were calculated.

To examine the on-line course of grammaticality judgments, the judgments made at each word in the sentence ("bad," "not sure," and "good") were translated into values of 0, 50, and 100, respectively. These judgments were then averaged over subjects and over sentences within a cell.

Results and Discussion

Accuracy on grammatical and ungrammatical targets

Overall accuracy levels, defined by the subject's decision on the last button press, were very high in this experiment, averaging 94.6% (see Table 3). For ungrammatical sentences, subjects used the "not sure" option at some point in their choice only 17% of the time. For ungrammatical sentences, subjects gave non-monotonic responses only 3.8% of the time. The average A' over subjects was a high 97, with no subject outliers (defined as any subject with an A' more than 2.5 standard deviations from the mean). Nor was any subject's mean "by word" reaction time more than 2.5 standard deviations from the grand experimental mean of 935 msec (this relatively high value was due in part to last button press, as we shall see later). For the bin-by-bin reaction times reported below, individual reaction time points greater than 2.5 standard deviations from the mean (more than 4100 msec) were eliminated (this constituted removing less than 2% of the data set). The A' analysis was conducted over subjects only, since the logic of A' is difficult to apply in an analysis by items.

Table 3. Incremental Grammaticality Judgment Experiment: Final button press on target sentences

	"good"	"not sure"	"bad"
gram.	94.50%	2.36	3.14
ungram.	4.01	1.29	94.69

Performance on individual sentences was examined to determine whether any of the sentences were outliers (defined as any sentence with a response more than 2.5 standard deviations from the mean, by item analysis). Only one sentence (#8.11 in Appendix I.C) met this criterion, classified as ungrammatical by only 29% of subjects. This sentence was dropped from all further analyses, and A' scores were calculated with this sentence removed. A 2 × 2 × 3 ANOVA with subjects as the random factor was conducted on these A' scores.

The following effects were significant (in all cases in this report, significant means $p < 0.05$): part of speech ($F1(1,34) = 7.27$, $p<0.0108$; $F2(1,71) = 13.10$, $p < 0.0006$) and part of speech × type ($F1(2,68) = 9.70$, $p<0.0002$; $F2(2,71) = 7.07$ $p < 0.0016$); location × part of speech was also significant by items only, $F2(1,71) = 4.78$, $p < 0.04$).[2]

Post-hoc tests were used to explore the various significant effects; unless otherwise stated, all post-hoc tests reported are Newman-Keuls. The pattern of accuracy for error types was different for the two parts of speech: for auxiliary errors, omissions > transpositions; for determiner errors, transpositions = agreement errors > omissions. Subjects showed an effect of part of speech only for omissions errors (auxiliary omissions > determiner omissions; also by items) not for agreement or transposition errors (see Figure 4).

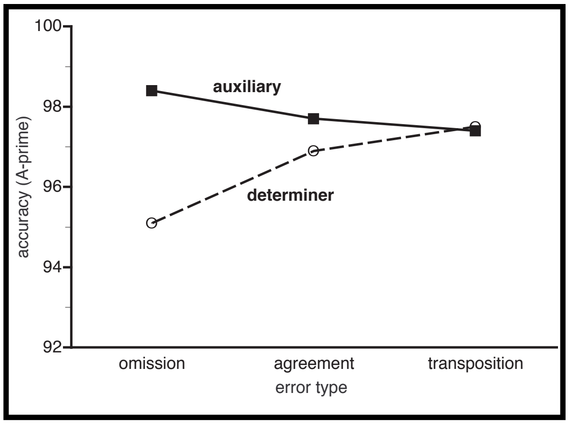

Figure 4. Incremental grammaticality judgment. A' by error type and part of speech

To summarize results for this analysis, overall accuracy was very high across all categories. For all subjects, there was a small disadvantage for violations involving determiners (especially determiner omissions).

Normalized bin-by-bin judgments and reaction times

In this section, we will begin with a global comparison of changes in judgment and reaction time for ungrammatical sentences and their grammatical controls. Early vs. late errors will be handled separately, but all other factors are collapsed at this level of description. We do not present significance tests at this point, but simply present the overall shape of the data. Then we will present detailed results for all twelve error types, evaluated in four separate analyses of variance over subjects (judgment and reaction times on early errors; judgment and reaction times on late errors). In each of these analyses, part of speech, error type and sentence position (or "bin") serve as within-subject factors. To maintain our focus on patterns of change over time (and to avoid redundancy with the previous section), we will restrict our discussion to main effects and interactions involving the factor bin.

Figure 5A illustrates the normalized bin-by-bin judgments observed on all ungrammatical targets with an early violation (collapsed over part of speech and violation type), compared with their grammatical controls. The vertical axis represents the mean rejection rate (i.e., mean percent judged ungrammatical) at each point in the sentence, from 0% (always judged grammatical) to 100% (always judged ungrammatical). The horizontal axis represents percentage of the sentence read so far, normalized for sentence length, with zero being the divergence point (see Methods section). The divergence point, which, recall, is the same structural point for omission, agreement and transposition errors, indicates the point at which we might expect a divergence between comparable grammatical and ungrammatical forms. This is, in fact, exactly what we observe. Notice, however, that the decision function for grammatical controls is not flat. Instead, there is a slight rise in the false-alarm rate that is most visible on the last word in the sentence.

Figure 5B compares the bin-by-bin judgments observed on all ungrammatical targets with a *late* violation, compared with their grammatical controls. Once again, we see the predicted divergence between grammatical and ungrammatical sentences at the divergence point. And we also see a slight rise in false alarms for grammatical controls toward the end of the sentence (averaging 4%).

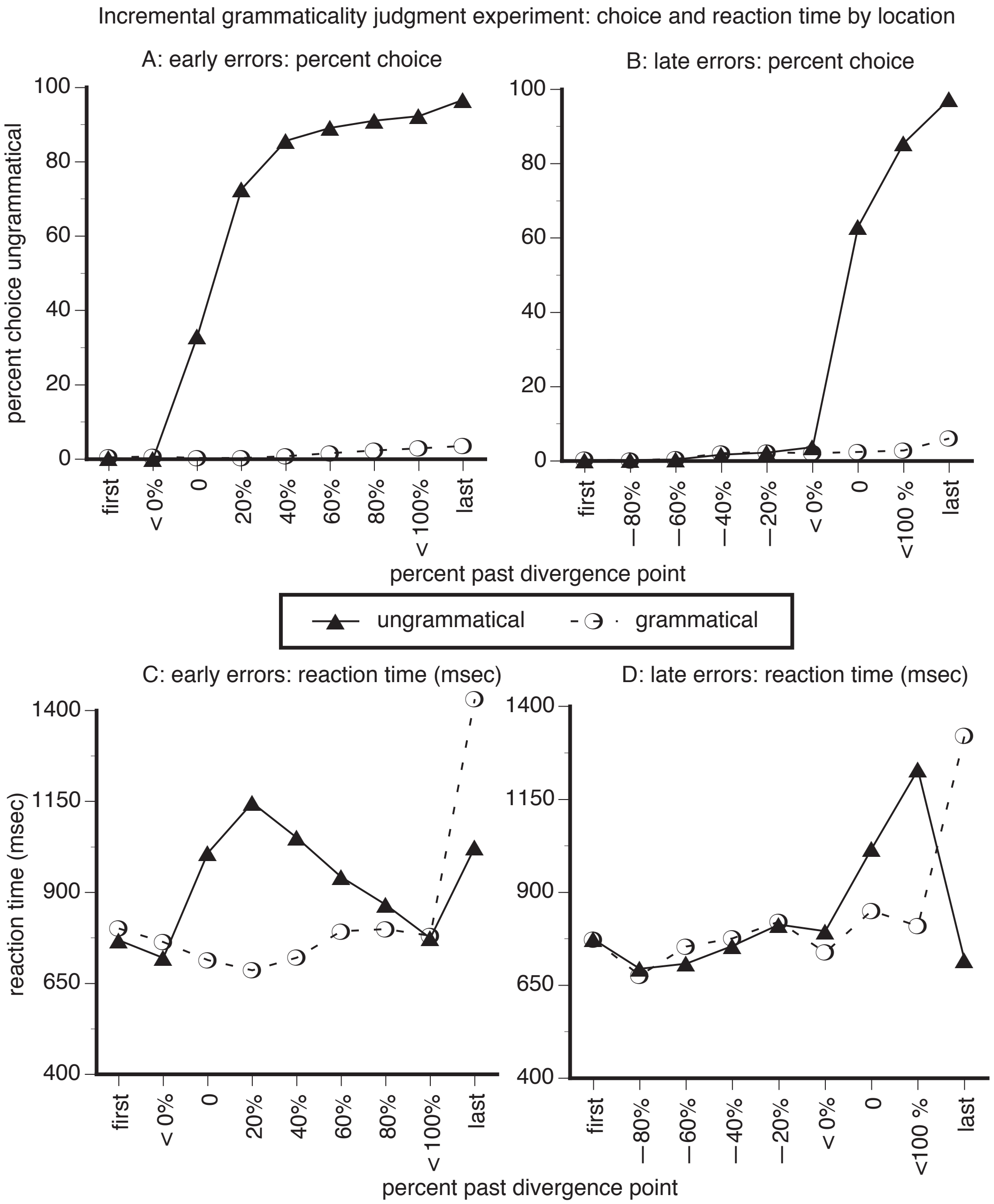

Figure 5. Incremental grammaticality judgment: Choice and reaction time by location

Figure 5C compares the bin-by-bin reaction times observed for ungrammatical targets with an early violation with their grammatical controls (collapsed over part-of-speech and violation type; recall that these are only reaction times for button presses before an "ungrammatical" response was made, and that individual data points more than 2.5 standard deviations from the overall mean were eliminated). Here (as in Figure 5A for bin-by-bin judgments), we would expect a marked divergence in reaction times beginning around the divergence point, with subjects slowing down as soon as they detect a potential error. This is, in fact, what we observe. Notice, however, that the reaction time function for grammatical controls is not flat. We might have expected roughly equivalent reading times at every point across the course of the sentence. Instead, we observe a slight slowing of reaction times toward the end of the sentence (compared to the middle) for grammatical controls, including a great increase in reaction time at the last button press. The gradual deceleration in grammatical sentences before the last press may reflect increased processing load and/or increased caution as information accumulates and the end of the sentence nears, while the particularly sharp increase at the final word may be due to subjects making a final check for potentially missed errors. For ungrammatical sentences, the predicted increase in reaction times after the divergence point is followed by a gradual drop back to the pre-error baseline. Ungrammatical sentences also show this effect, but in this case the effect appears to be restricted to the last word in the sentence. Figure 5D presents bin-by-bin reaction times for ungrammatical sentences with a late violation, compared with grammatical controls. Like Figure 5B for judgments, this figure also shows a marked divergence in reaction times around the divergence point. However, there is much less divergence between late violations and their grammatical controls, due to a confound between error detection (which slows reaction times toward the end of the sentence for items with a late violation) and the last-press elevated-reaction time effect described earlier (which slows reaction times at the end of the sentence for grammatical controls).

The twelve error types analyzed separately: Finally, let us turn to the patterns of change associated with each of the 12 error types. As noted, there were four separate analyses of variance: early judgments, late judgments, early reaction times and late reaction times; for reaction times, missing observations were replaced with cell means. We will restrict ourselves here to effects involving the factor "bin". All four analyses yielded a very large main effect of bin ($p < 0.0001$ in every case, also by items), a significant two-way interaction between bin and error type ($p < 0.0001$ in every case, also by items except for early reaction times, $p < 0.03$), and a significant two-way interaction between bin and part of speech ($p < 0.0001$ in every case; $p< 0.015$ by items). Most interesting for our purposes here are the three-way interactions of bin, part of speech and error type. This three-way interaction reached significance in the analysis of early decisions

		bin								
		early errors						late errors		
	factor	0%	20%	40%	60%	< 100%	last	0%	50%	last
"not sure"	type	■	■	■	■	■			■	
	part of speech	■		■	■	■				
	type X part of speech	■	■	■	■	■				
"ungrammatical"	type	■	■	■	■	■			■	■
	part of speech	■		■	■	■				
	type X part of speech		■	■	■	■	■			

Figure 6. Incremental grammaticality judgment: Analyses of Variance

(F1(12,408) =7.13, p < 0.0001; F2(12,216) = 2.75, p < 0.0017), the analysis of late decisions (F1(12,408) = 22.44, p < 0.0001; F2(12,210) = 3.93, p < 0.0001), the analysis of early reaction times (F1(12,408) = 8.76, p < 0.0001; F2(12,216) = 3.00, p<0.0007), and the analysis of late reaction times (F1(12,408) = 3.56, p < 0.0001; F2(12,210) = n.s).

Figure 6 presents a more detailed, by-bin breakdown of significant effects for early and late errors separately.

Figure 7A to Figure 7C present the bin-by-bin judgments observed for early auxiliary errors (as analyzed just above) for omissions, agreement errors, and transpositions. Both "not sure" responses (striped) and "bad" responses (light gray) are shown. Agreement errors are resolved fairly quickly, and all three error types have significantly different rejection rates at the zero point (using Newman-Keuls): agreement (67%) > omission (44%) > transposition (34%). Agreement errors reach 92% by the next interval (i.e., the "20%" interval), where they are still significantly higher than the other two error types.

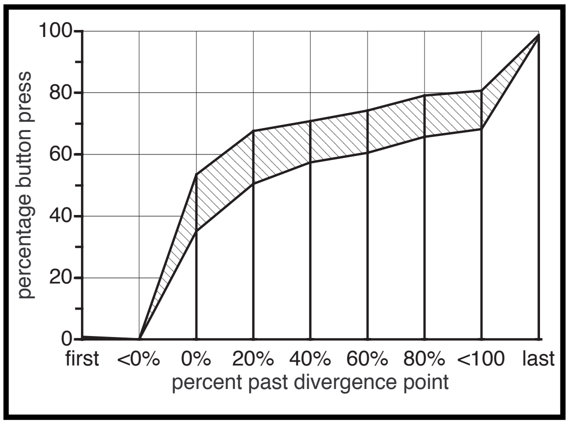

Figure 7A. Early auxiliary omission, Choice

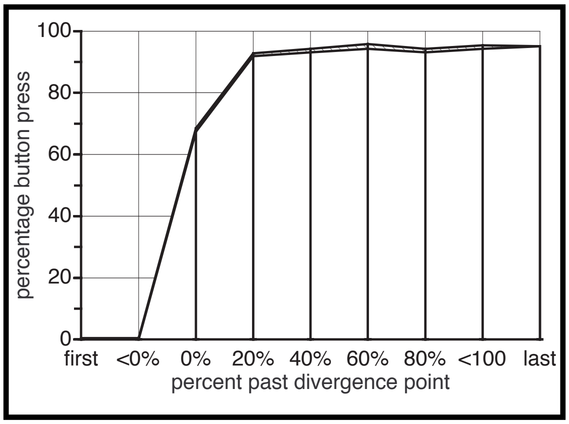

Figure 7B. Early auxiliary agreement, Choice

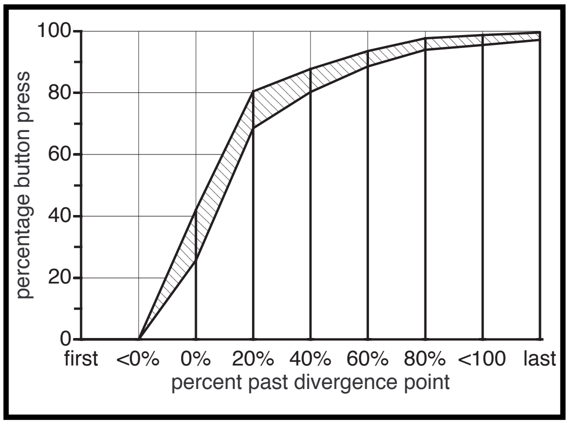

Figure 7C. Early auxiliary transposition, Choice

Omission and transposition errors switch rank order at the "20%" interval (which corresponds to the displaced element on transposition errors), with transpositions at 75% and omissions significantly less at 59%. This suggests that the second piece of information (a displaced auxiliary verb) is sufficient to quell doubts for most of our subjects.

By contrast, omission errors show a more protracted rise in rejection rates across the rest of the sentence, with many subjects refusing to make up their minds before the final button press. The pattern agreement > transposition > omission obtains up to and including the "40%" interval. After that point, transpositions rise to meet agreement, and the two are no longer significantly different, while omissions remain significantly less than these two until all three converge at the last button press.

Figure 7D presents complementary data for bin-by-bin *reaction times* on early auxiliary errors, with all three error types on one graph.

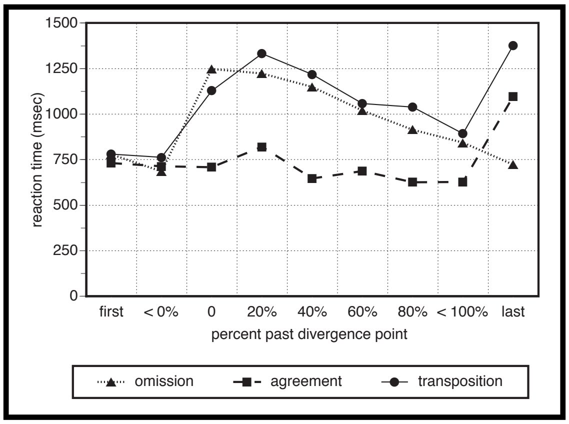

Figure 7D. Early auxiliary errors, Reaction time

Recall that these data are only for responses where the subject has *not yet* indicated "ungrammatical" to the sentence—i.e., where the subject is still deciding. Reaction times first show a significant difference between types at the divergence point: agreement (709 msec) < transposition (1,129 msec) = omission (1,247 msec). Both omissions and transpositions show a significant reaction time jump from the interval just before the divergence point to the divergence point (in this case using paired t-tests[3]). At the next interval, the "20%" interval, omissions (1,224 msec) and transpositions (1,332 msec) switch order from the interval before, though again the difference is not significant; both are still significantly slower than agreement errors (820 msec). Reaction times for omissions and transpositions show a slow drop over the rest of the sentence, until just before the last word. The pattern of agreement errors being significantly faster than the other two error types obtains up to and including the "60%" interval; after that, the differences between error types do not reach significance. Note that on the last word, the reaction time for omissions continues to drop, while that for both agreement and transposition rises, though statistical analyses including reaction time from the last word could not be performed due to too many missing observations (recall that by this point most subjects have pressed the "ungrammatical" button, and thus almost all reaction times for this interval have been culled).

To summarize so far, grammaticality judgments on these three early auxiliary error types result in markedly different bin-by-bin judgments and reaction times, corresponding to the amount of ambiguity and/or the number of cues to violation associated with each violation type, based upon our experience from Experiment 1. For judgments, responses diverge at the divergence point with the profile agreement > omission > transposition; at the next interval, agreement has nearly reached asymptote, and transposition has risen to overtake omission. Both transposition and omission continue to rise over the rest of the sentence, with transposition reaching asymptote by the "60%" interval and omission not reaching it until the last word. For reaction times, agreement errors are essentially constant over the course of the sentence (at least until the last word), while both transposition and omission errors rise significantly at the divergence point. This jump appears to continue until the moved element (the "20%" interval) for transpositions and then slowly decrease, while for omissions the decrease begins immediately after the divergence point. The correspondences between judgment and reaction time for all three error types are also interesting: The quickly-resolved agreement errors' relatively constant reaction time suggests that the perceived divergence point is rather punctate; that is, subjects either catch the error (and they usually do), or they miss it entirely. By contrast, the elevated reaction times at and after the divergence point for omissions and transpositions make sense in the context of the more extended decision region that these two error types evince, as they suggest a "zone of uncertainty" in which subjects, at any one interval, are hesitant to commit to an "ungrammatical" button press (hence the slower rise of the functions) but are also uncertain about the possibility of

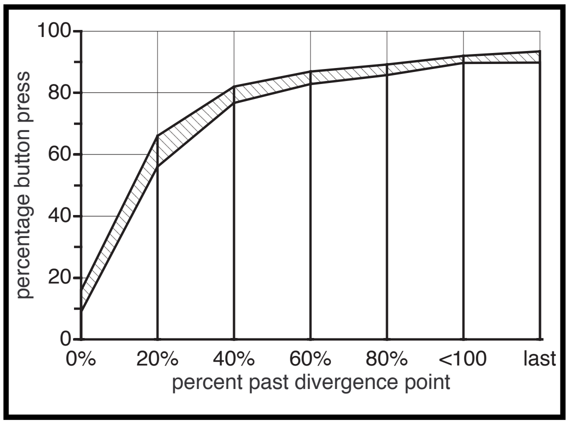

Figure 8A. Early determiner omission, Choice

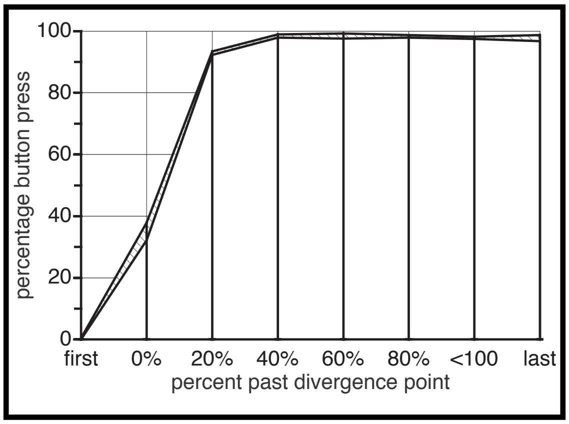

Figure 8B. Early determiner agreement, Choice

the sentence's continuing grammatically (hence the concomitant rise in reaction times). These results suggest that the bin-by-bin reaction times can be interpreted as a kind of confidence rating for each judgment.

Figure 8A to Figure 8C display the bin-by-bin judgments observed with early determiner errors, while Figure 8D presents the bin-by-bin reaction times for the same items.

Recall that for both omissions and transpositions the divergence point is always the first word of the sentence. The same error type profile at the divergence point as for early auxiliaries is seen, although the overall mean is lower, and all three error types have significantly different rejection rates: agreement (35%) > omission (12%) > transposition (4%). Agreement errors reach 93% by the next interval, the "20%" interval, where they are still significantly higher than the other two error types (both 61%). By the next interval, the "40%" interval (the interval *after* the moved element for transpositions), transpositions (93%) have risen to the point where agreement = transposition > omission (79%). Omissions slowly rise up over the course of the sentence, but are still significantly less than the other two error types up to and including the "80%" interval.

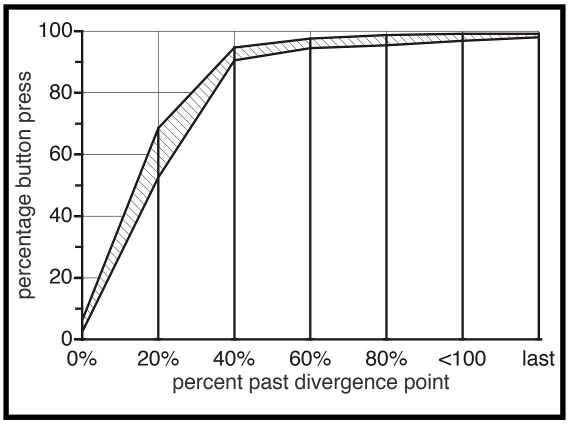

Figure 8C. Early determiner transposition, Choice

In general, these results parallel our findings for early auxiliary errors: Agreement errors are resolved rather quickly, while omissions take much longer to resolve, and transposition errors fall somewhere in between. However, a comparison of Figure 7 (early auxiliary errors) to Figure 8 (early determiner errors) suggests some subtle but interesting differences between auxiliaries and determiners. First, on auxiliary agreement errors, subjects tended to make their decision directly at the verb (e.g., "The boy were * + walking," with a 67% rejection rate by the divergence point), while they tended to wait for one more word after the determiner agreement error occurred (e.g., "A girls * were + walking," with a 35% rejection rate by the divergence point and a 97% rejection rate at the next interval). This delay may be due to competition from an alternative completion for early determiner agreement errors (e.g., "A boy[']s life…" even though the punctuation provided on the screen is not compatible with a completion of that kind). A similar conclusion applies to the other two early determiner types. On early determiner omissions (e.g., "Boy * was +…") and early determiner transpositions (e.g., "Boy * the + was…"), subjects do not start rejecting the sentence until *after* the divergence point has passed—until at least one more word (the "20%" interval). In other words, they do not judge the error at the divergence point. In fact, these subjects are right: As we saw in the Experiment 1, many of the sentences with early determiner omissions or transpositions can be salvaged at the divergence point by completions in which the bare noun phrase is actually working as a modifier (e.g. "Boy… *George is a very strange person*,"). However, the next cue is apparently not sufficient to remove all doubts about ungrammaticality. Shortly after the "20%" interval, the rejection rates for early determiner transpositions move toward asymptote (at about the "40%" interval). Omissions take considerably longer, and do not reach the 95% rejection level until the end of the sentence in many cases.

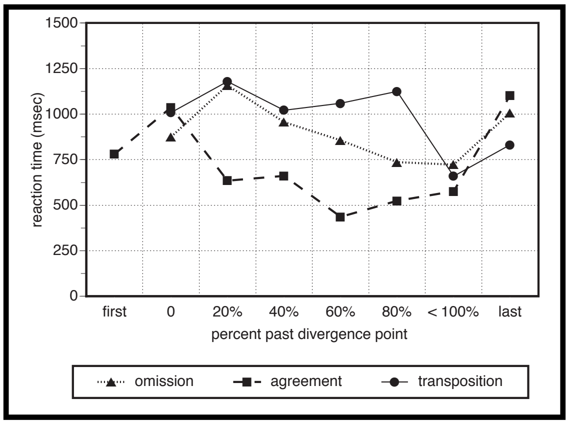

Figure 8D. Early determiner errors, Reaction time

Figure 8D presents the bin-by-bin reaction times observed for early determiner errors, with all three error types on one graph. This profile matches that for early auxiliary error reaction times in its gross outlines, with an elevated reaction time early in the sentence, a gradual decline in reaction time over the rest of the sentence (although transpositions rise again at the "60%" interval), and with agreement errors notably faster than the other two types. However, type only reaches significance at the "20%" interval, with agreement significantly faster

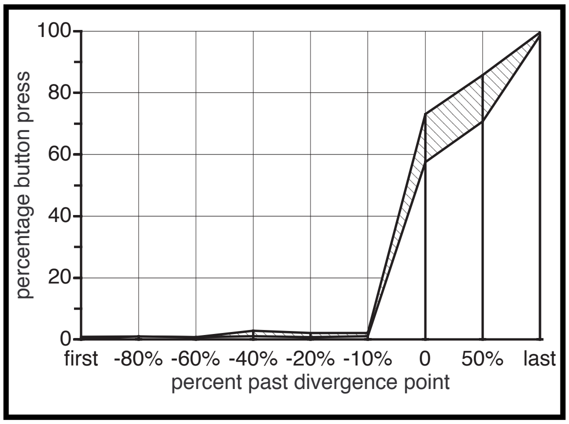

Figure 9A. Late auxiliary omission, Choice

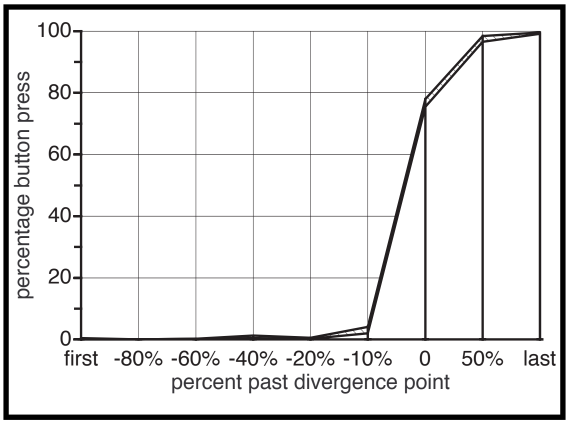

Figure 9B. Late auxiliary agreement, Choice

than the other two error types. For agreement errors, reaction time at the divergence point is significantly slower than at the intervals on either side (using t-tests); for omission errors, reaction time at the divergence point (for this error type, always the first word) is significantly *faster* than at the next interval. There were too few data points at the last word for a meaningful statistical analysis.

To summarize for early errors, subjects reject sentences at (for auxiliary errors) or just after (for determiner errors) the divergence point for agreement errors. Transposition error rejection rates rise more slowly, with most rejections not occurring until the moved element has appeared. Omissions have the slowest rejection rise time, taking most or all of the sentence to rise to asymptote. Reaction times tended to reflect judgment patterns. The two error types with the more protracted decision regions, omissions and transpositions, showed a reaction time jump at or just after the divergence point, followed by a slow reaction time drop over the rest of the sentence. Agreement errors, with a more punctate decision point, were faster overall, with either no reaction time change at the divergence point (for auxiliary errors) or with a reaction time jump directly at the divergence point that fell again at the next interval (for determiners).

Let us turn now to the six late-error types, displayed in Figure 9 and Figure 10. Figure 9A to Figure 9C present bin-by-bin judgments associated with late auxiliary errors, while Figure 9D presents complementary information on bin-by-bin reaction times for these items.

In contrast with the early auxiliaries (Figure 7), the late auxiliary errors are all resolved fairly quickly compared to some early errors, peaking at or shortly after the divergence point. To some extent this finding was inevitable, because these errors are

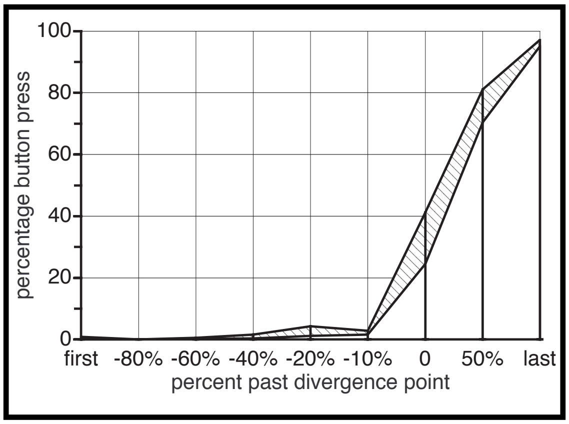

Figure 9C. Late auxiliary transposition, Choice

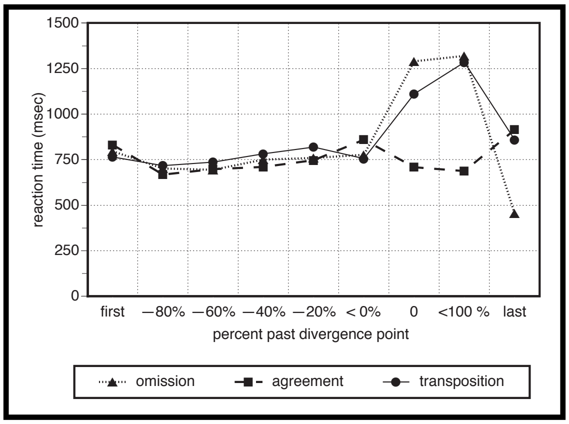

Figure 9D. Late auxiliary errors, Reaction time

located at the end of the sentence, where subjects are forced to make a quick decision.

The three error types diverge at the divergence point, with all three significantly different in the order agreement (77%) > omissions (65%) > transpositions (33%). At the next interval (which comprises all words after the divergence point except for the last word of the sentence), agreement has risen to 97%, significantly higher than either transposition (76%) or omission (78%). At the last word, all three error types have risen to between 96% (transpositions) and 99% (omissions and agreement errors); the difference is significant.

Reaction times diverge at the divergence point as well; omissions jump significantly (from 777 msec at "<0%" to 1290 msec at the divergence point, using a t-test) as do transpositions (from 755 msec to 1111 msec). Agreement errors dropped significantly (from 861 msec to 709 msec). Agreement errors are significantly faster than the other two error types both at the divergence point and at the next interval. The further increase in reaction time from the divergence point to the next interval is significant for transpositions only. Again, there were too few data points at the last word for a meaningful statistical analysis.

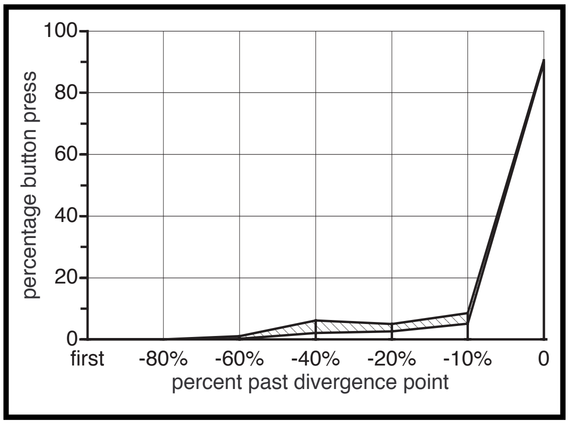

Figure 10A. Late determiner omission, Choice

Finally, Figure 10A to Figure 10C present the bin-by-bin judgments observed on late determiner errors, while comple-

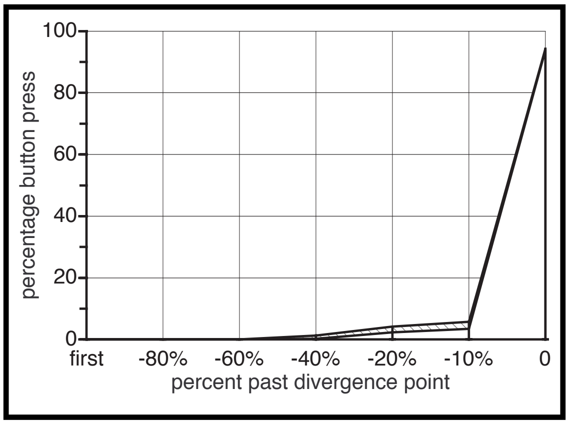

Figure 10B. Late determiner agreement, Choice

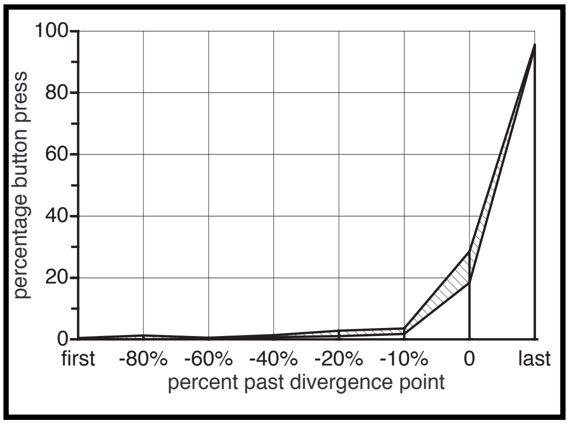

Figure 10C. Late determiner transposition, Choice

mentary information on bin-by-bin reaction times is presented in Figure 10D.

By necessity (because the divergence point so often corresponds to the last word in the sentence), all of these errors are resolved fairly quickly, compared to some early errors. At the divergence point, transpositions (which still have one more word to go, by necessity—the moved element) are significantly lower (23%) than the other two error types (90% or better). The increased reaction time at the divergence point for transpositions is significant, using a t-test, (from 724 msec to 1,064 msec) as is the decreased reaction time for agreement errors (857 msec to 643 msec).

In summary, late errors by necessity show fewer between-type differences due to the small interval between the error and the end of the sentence. Nevertheless, some interesting generalizations can be made: Again, agreement errors stand out in that they are resolved sooner than the other two error types (although this effect may not be seen very clearly in the late determiner case due to the confound of the end-of-sentence effect).

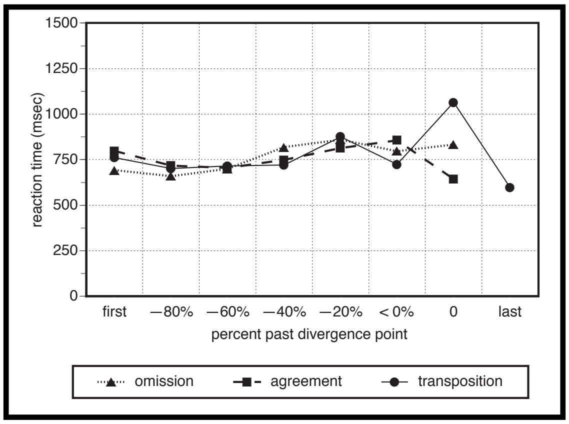

Figure 10D. Late determiner errors, Reaction time

Transpositions and omission errors again show a jump at the divergence point (though again, for omissions this effect may not be seen clearly due to the confound of the end-of-sentence effect), with agreement errors being faster than the other two error types at and after the divergence point.

The "Protracted Decision Region" of Early Auxiliary Omissions—an Artifact?

One possible interpretation of the "protracted decision region" especially seen in early auxiliary omission errors is that it is not a function of individual subjects making up their minds throughout that region, but rather that it is an artifact of averaging over items, or over subjects, or both.

Subject variance: assume that one subset of subjects, "early responders", usually indicates that the sentence is ungrammatical at or just after the divergence point, while the other subset, "late responders", entertain potential grammatical (or semigrammatical) completions until nearly the end of the sentence. Averaging over cells might then produce an effect that could be wrongly interpreted as a protracted decision region for individual subjects (note, however, that the slow rise of the decision function over the course of the sentence, rather than a sudden jump at the divergence point, a plateau, and then another jump near the end of the sentence disproves at least the strong version of this argument).

Item variance: Similarly, subjects (some or all) might be inclined to always indicate an ungrammaticality immediately at the divergence point for some items, and near the end of the sentence for other items, again producing the spurious appearance of what appeared to be an across-item decision region.

To test the first possibility we examined only early auxiliary omission errors, by items, and calculated the "mean words past divergence point" for each subject; i.e., the average number of words past the divergence point for each sentence that subjects first pressed the "ungrammatical" button. If subjects fall into two classes, "early" and "late" responders, then individual subjects should have relatively little variability on this measure regardless of exactly where in the sentence they tend to respond. However, the average standard deviation over subjects in this cell was 2.48, indicating that even individual subjects varied on where they believed the sentence had become ungrammatical, for this cell of the design.

To test the second possibility, we performed a mean split on the subjects and proceeded with those subjects whose standard deviation on this measure was above the group mean. By eliminating low-variance subjects, we reduced the chance of creating the appearance of variance on particular items by pooling early responders and late responders together. We then examined each item's variance, over subjects. Six of the seven items ranged in standard deviation from 2.4 to 3.4 (sentence 1 had a standard deviation of 1.0; when *all* subjects were included standard deviation ranged from 2.0, again for sentence 1, to 3.3). Thus, there is a fair degree of within-item variability (with the possible exception of sentence 1), and therefore the variability of the “protracted decision region” is *not* an artifact of subject variability or item variability.

Although it seems clear that the apparent “decision zone” is not an artifact of collapsing over items or subjects, a third issue remains: how are reaction times and decisions related within this decision zone, for individual subjects struggling with an individual item? If subjects are entering into a protracted phase of indecision, then we might expect to find that reaction times increase well before the point at which a decision is finally made. To investigate this possibility, we began by locating the point at which individual subjects switched from “good” to either “not sure” or “bad” for the first time, for each item. We will call this the “zero point.” If all decisions are really punctate (and the protracted decision zone is an artifact of averaging), then the space between elevation of reaction time and the decision to reject a sentence should be very small, and it should be the same for all item types. Two patterns are possible under the artifact interpretation:

1. reaction times do not go up until the point where decisions change, or
2. reaction times go up at the button press just before the point where decisions change (as subjects get ready to move their fingers from one button to another), but not before.

Furthermore, this pattern should be the same for all the major violation types.

If, on the other hand, the major violation types differ in the relation between reaction time increase and button press at the level of individual items and individual subjects, then we can conclude with some confidence that these variations in the size of the decision region are not an artifact of averaging. To ask this question, we began by locating the point at which individual subjects switched from “good” to either “not sure” or “bad” for the first time, for each item. We will call this the zero point.

We then examined the reaction time for each of the five words prior to that zero point, which we will call:

- 1 (the word prior to the button press shift),
- 2 (the word before -1), -3, -4 and -5, respectively.

This analysis was conducted on early items only (both auxiliaries and determiners), and because of variations in sentence length, the number of items contributing to each cell is necessarily smaller the farther back we go (e.g. sentences on which the zero point occurred on the second word can only furnish a single -1 point, and no reaction time measurements from -2 through -5).

If, for example, auxiliary omissions yield a longer decision region than auxiliary transpositions, then the reaction times from points -2 to -5 should be larger for auxiliary omissions.

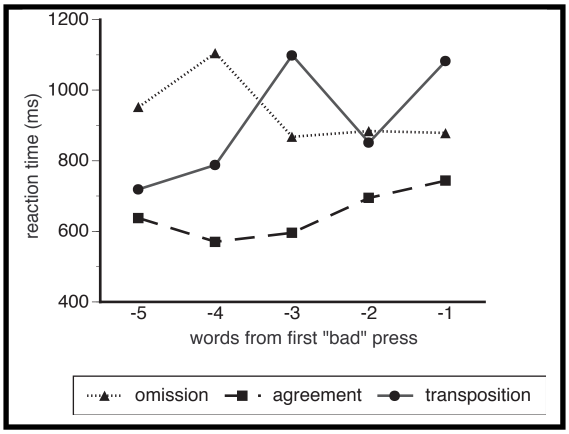

Figure 11A. Incremental grammaticality judgement: Reaction times before first “bad” press for early auxiliary errors

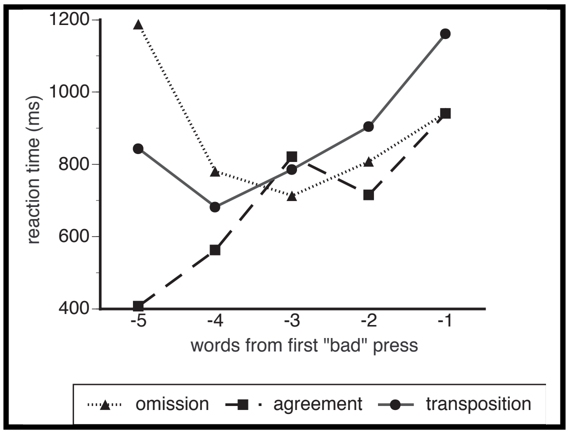

Figure 11B. Incremental grammaticality judgement: Reaction times before first “bad” press for early determiner errors

Two separate analyses of variance were conducted, one for auxiliaries and another for determiners. Each of these involved a 3 (omission vs. transposition vs. substitution) by 5 (position -1 to position -5) within-subjects design.

For early auxiliaries, the analysis yielded no main effect of word position, but a significant main effect of type (F(2,68)=4.70, p < 0.02) and a significant interaction between type and word position (F(8,188)=3.80, p < 0.00004).

The interaction is illustrated in Figure 11A. Post-hoc analyses at each position showed that the error type difference was significant at position -4 (F(2,36)=6.70, p < 0.004, where omission > transposition = agreement), at position -3 (F(2,58)=6.83, p < 0.003, where omissions = transposition > agreement), at position -2 (F(2,68)=4.47, p < 0.02, where omission = transposition > agreement), and at position -1 (F(2,68) = 15.64, p < 0.0001, where transposition > omission > agreement). The difference was in the predicted direction for position -5 as well, but this analysis was not reliable (due perhaps to the small number of items contributing to this cell for each subject).

For early determiners, the only significant effect of the analysis of variance was a main effect of word position (F(4,87)=3.42, p < 0.02). However, results were in the same general direction reported above, with larger differences between item types at the earlier positions (see Figure 11B).

This analysis of the relationship between reaction time and judgment demonstrates that some subjects are sensitive to some errors well before they are willing to register their decision, certainly more than just a word or two. Furthermore, it shows that the distance between this sensitivity (manifest in the point in the reaction time data where reaction time begins to increase) and the eventual decision (manifest in the button press) is different for different error types. This is what we mean by a "protracted decision region".

Comparison of Experiments 1 and 2

In Experiment 2, early determiner omissions and transpositions were still often acceptable to our subjects at the divergence point, suggesting that they still have a range of possible completions in mind. However, the reaction time data suggest that subjects are already doubtful about the grammatical status of these items. Experiment 1 is consistent with this. For these two error types, subjects provided a grammatical completion past the divergence point an average of 66% of the time. Subjects provided a variety of completions at this point, including use of the bare noun as proper noun or title (e.g., "President… *Clinton was briefed by his advisors*."), use of noun in the general sense (e.g., "Man… *is a fragile creature*."), and use of noun as adjective (e.g., "Woman… *doctors*…").

In addition, in Experiment 2 early determiner *agreement* errors (e.g., "A girls *…") appeared to be resolved approximately one word later than early *auxiliary* agreement errors (e.g., "John are * …"), again consistent with Experiment 1. Subjects in the cloze experiment provided completions at the divergence point for 34% of the early determiner agreement error sentences, compared with only 3% for the corresponding auxiliary errors. 82.6% of these completions involved the use of the bare noun as an adjective, such as, "Several sailor… *uniforms were in my bag*," including many completions where subjects mistakenly used a plural noun as a possessive, such as, "A boy[']s… *life is very simple*."

Finally, in Experiment 2 early auxiliary omission errors started to be perceived as ill formed at the divergence point, but many subjects were still unwilling to make up their minds about these error types until the very last word in the sentence. In between, there was a long and monotonic drop in acceptability (i.e., a true "decision region"), with substantial variability over individual subjects and items. Experiment 1 was also consistent with this, with subjects delaying their decision on early auxiliary omissions as they considered a participial interpretation such as, "Mrs. Brown[,] working at the library[,] is…" (even though punctuation did not support this interpretation, and most of the item types within this cell involve unique referents—proper nouns, pronouns or other unique individuals—that are unlikely candidates for such a participial interpretation). Subjects provided completions at the divergence point for 30% of the sentences of this type. All of these completions involved either present-participial verb phrase completions or gerund + "that" clause completions.

The incremental grammaticality judgment (GJ) and cloze procedures yield very similar results. To quantify this observation, we calculated for the data of Experiment 2 the “mean words past divergence point” measure for each item, as done in Experiment 1. The experiments correlated significantly on this by items measure, with a Pearson correlation coefficient of 0.83 (p < 0.0001). Interestingly, in Experiment 2 the grand mean for “mean words past divergence point” was higher—that is, subjects also tended to wait longer to give an “ungrammatical” response in Experiment 2, a point to which we shall return.

This significant 0.83 correlation lends support to the notion that both experiments are tapping into essentially the same underlying, ongoing structure-building process. In addition, the profile of means in both experiments were almost identical (although the exact patterns of significance revealed by post-hoc tests were somewhat different). In both experiments, for early auxiliary errors, the order of means was omission > transposition > agreement; for late auxiliary, transposition > omission > agreement. In both experiments, for early determiner errors, agreement errors had the lower mean, while for late determiner errors, transposition errors always had the higher mean (for structural reasons already discussed).

What is going on in this “decision region” from the subject’s point of view? Are they conscious of the error at the point where reaction times start to increase? Are they postponing a decision until more information is available, and all possible completions have been eliminated? Or have they made their judgment (i.e. they already “know” that the sentence is bad), but have decided for some reason to postpone a final button press (like an engaged couple who are not quite ready to announce their plans to the family)? The strong correlations that we have observed between performance in the Cloze experiment and performance on the judgment tasks suggests that the subjects are still weighing alternatives. However, the experiments presented here do not permit us to draw strong inferences about the phenomenology of grammaticality judgment, i.e. we do not know what is going on in our subjects’ minds, before, during or after the proposed “decision region”. It is possible that the distinction between sensitivity to error (a perceptual event) and the decision to push a button (a form of motor planning) could be disentangled with another methodology (e.g. event-related brain potentials). For present purposes, however, we can conclude with some confidence that grammatical violations differ in the amount of time required to register a decision.

Summary of results for Experiment 2

Experiment 2 has yielded a great deal of information about the time course of grammaticality judgment, much of it consistent with Experiment 1, and summarized briefly as follows:

1. **Accuracy.** Overall, end-of-sentence ac-

curacy was very high in this experiment, averaging around 95% correct rejections for ungrammatical sentences and 95% correct acceptances for their grammatical controls. An analysis of variance A' scores (corrected for response bias) yielded very few differences among the various error types, although performance was slightly worse overall for determiner omissions.

2. **Relationship between judgments and reaction times**. There were striking parallels between the decision and reaction time data, suggesting that the word-by-word reaction times obtained with this paradigm can be viewed as an indirect index of the degree of confidence associated with grammaticality judgments at each point in the sentence, as well as, perhaps, a decision process in which subjects attempt to generate alternatives. In general, both sources of information (word-by-word decisions and reaction times) offer useful and complementary information about the time course of grammaticality judgment.

3. **Size and shape of the decision function**. The twelve relatively simple error types that we have manipulated here are associated with markedly different decision functions, with a significant correlation between the cloze and incremental GJ techniques. For some error types, it seems fair to conclude that there is a single decision point, located at or close to our predetermined divergence point. This is true for early auxiliary agreement errors, and it is true for most errors located late in the sentence—although the latter finding is probably due to the uninteresting fact that subjects are forced to make up their minds by the presence of a period signaling the end of the sentence. For all the remaining violation types, we have to abandon the punctate view in favor of something that is best described as a "decision region." This conclusion is forced by the following facts:

a. Early determiner agreement errors (e.g., "A girls *…") appear to be resolved approximately one word later than early auxiliary agreement errors (e.g., "John are * …"). To explain this difference, we noted that in Experiment 1 subjects provided completions such as, "A boy[']s life is very simple"—despite the fact that such completions should be ruled out by the absence of an apostrophe to signal a possessive reading.

b. Early auxiliary omission errors start to be perceived as ill formed at the divergence point, but many subjects are still unwilling to make up their minds about these error types until the very last word in the sentence. In between, there is a long and monotonic drop in acceptability (i.e., a true "decision region"), with substantial variability over individual subjects and items. This is also consistent with Experiment 1, where subjects delayed their decision on early auxiliary omissions because they were still considering a participial interpretation such as, "Mrs. Brown[,] working at the library…" However, most of the item types within this cell involve unique referents (proper nouns, pronouns or other unique individuals) that are unlikely candidates for such a participial interpretation (see Appendix I). Such interpretations would be possible with a different form of punctuation (e.g., a non-restrictive clause such as "Mrs. Brown, working at the library, called home to say…"). But no such punctuation was provided in this experiment. Perhaps our subjects delay their decisions on early auxiliary omission items because of their *partial* overlap with or resemblance to participial constructions. In addition, these issues could be resolved by further studies systematically varying the number, frequency and degree of plausibility of competing sentence completions—a point to which we shall return later.

c. Early auxiliary transposition errors are resolved in at least two steps: Rejection rates start to go up at the divergence point (where omissions and transpositions are still equivalent), with a sharp increase at the next word (the displaced auxiliary, which serves as a second cue). Still, these errors do not reach asymptote until about 60% past the divergence point (i.e., roughly six words after the divergence point), suggesting that many subjects are unwilling to make up their minds until the end of the sentence. A similar second-cue effect is observed on late auxiliary errors, although these items are then forced to asymptote by punctuation signaling the end of the sentence.

d. Early determiner omissions and transpositions are still acceptable to our subjects at the divergence point, consistent with the range of completions subjects provided in Experiment 1 (e.g., "Boy George…"). However, the reaction time data suggest that subjects are already doubtful about the grammatical status of these items. For approximately half the subjects (and/or half the items), this suspicion is confirmed by the next word. Nevertheless, judgments and reaction times associated with these early determiner items do not reach asymptote until 40% past the divergence point.

It seems fair to conclude that grammaticality judgment is a matter of degree, a protracted and variable process. To what extent is this result an artifact of the word-by-word judgment task itself? To answer this question, we proceed to our third and final experiment.

EXPERIMENT 3: Rapid Serial Visual Presentation

In order to prove that our results are not an artifact of incremental presentation and judgments (which certainly are further from the processing that occurs in real life), the third experiment tests subjects with the same stimuli using the rapid serial visual presentation (RSVP) paradigm.[5]

Method

Subjects. Subjects were thirty-two UCSD students who completed the experiment either for course credit or for a $7 payment. One subject was dropped from subsequent analyses for having A' scores more than 2.5 standard deviations from the mean (see below). Of the thirty-one remaining subjects, twenty-five were male and two were left-handed. All subjects were native speakers of English.

Stimuli. The materials were the same as those used in Experiments 1 and 2.

Equipment. The experiment was conducted on an IBM PC-XT, using a GoldStar 1210A amber screen monitor and a Carnegie-Mellon University button box, accurate to one millisecond. Stimuli appeared at the center of the screen, one word at a time.

Procedure. Subjects first practiced using the button box for twenty trials. First, the word "READY" appeared at the bottom center of the screen. Then, the single word "Correct" or "Wrong" appeared in the center of the screen, for 350 msec. Subjects were told to push the corresponding button as fast as possible. In contrast, with Experiment 2, only two buttons were used in this experiment (i.e., no "not sure" option was provided). Reaction times were recorded. This task provided practice with the button box, together with a baseline reaction time for each subject.

After practice with the button box, subjects were given an opportunity to practice the judgment procedure. During the sentence practice session, subjects received twenty trials. The practice sentences were comparable in length, structure, and error type to the actual data set, but did not overlap with this data set.

Both the button pressed ("GOOD" or "BAD") and the reaction time (in msec) were recorded at each trial. For ungrammatical sentences, reaction time was measured from the same divergence point as Experiments 1 and 2. For grammatical sentences, reaction time was measured from sentence onset.

A trial consisted of the following:

1. The screen was clear for 500 msec.
2. The word "READY" appeared near the bottom center of the screen, for 1000 msec.
3. The screen cleared, and a 2000-msec pause followed.
4. The sentence appeared in the middle center of the screen, one word at a time. Each word appeared for 350 msec, without a pause between

words.

5. As soon as subjects had made the grammaticality judgment—even if the sentence was still running—they were to press the appropriate button.
6. At the end of the sentence, the screen was blank for 3000 msec, during which time the program would still accept a button press.
7. This constituted the end of a trial. The following trial then began, with another 500 msec pause and "READY" cue.

The experimenter instructed subjects to read the sentences carefully as they appeared on the screen, and to press the button as quickly as possible after making their decision, even if the sentence was still running. Subjects were instructed to focus on what they considered proper grammar, and not on ideal style, punctuation, or spelling, which were always correct.

The actual experiment consisted of 168 trials of the sentence stimuli described above. Each subject received the sentences in a different random order, determined by the controlling computer program. Subjects were told that they would receive a break at the mid-point of the experiment (after trial 84). At this point, instead of the "READY" cue, the subject received a "PLEASE WAIT" cue.

Data analyses. Two dependent measures were used: A' (see above), and reaction time. Reaction times were used only for correctly answered ungrammatical core stimuli. These reaction times were measured from the divergence point. As described in Experiments 1 and 2, the divergence point corresponds to the first point at which there was a divergence between ungrammatical sentences and their grammatical controls. Although as the cloze experiment has shown us there are still a variety of ways that some of the sentence types might be saved beyond this point (particularly true if the subjects are willing to ignore punctuation), this is the first point at which the error types manipulated in this experiment could conceivably be detected. Omission, transposition and agreement errors all share the same divergence point (i.e., they all diverge from grammatical controls on the same word).

Results and Discussion

Overall accuracy for grammatical and ungrammatical sentences

All of the analyses which follow are based upon the thirty-one subjects who remained after the one outlier subject was dropped. Performance on individual sentences was examined to determine whether any of the sentences were outliers (defined as an accuracy level more than 2.5 standard deviations below the mean). As with Experiments 1 and 2, sentence #8.11 (in Appendix I.C) met this criterion, classified as ungrammatical by only 21% of subjects. So did sentence #5.1, classified as ungrammatical by only 50% of subjects. These two sentences are dropped from all further analyses, and the following A' scores were calculated with these sentences removed.

Accuracy levels were roughly similar to those of Experiment 2, suggesting that the added pressure to respond quickly did not result in an increased level of error. Subjects were at 90.8% on grammatical sentences and 93.0% on ungrammatical sentences, which corresponds to an A' of 95.6.

An analysis of variance was conducted on the A' scores, treating subjects as a random variable within a 2 × 2 × 3 within-subjects design. Location of error (early vs. late), part of speech (auxiliary vs. determiner) and violation type (omission, agreement, transposition) were the factors. This analysis yielded two significant main effects, and a significant interaction: a main effect of error type (F(2,60) = 4.93, p < 0.011; by items, F(2,110) = 3.69, p < 0.003) and a main effect of part of speech (F(1,30) = 5.76, p< 0.0228; by items, F(1,110) = 7.99, p < 0.056). The interaction was part of speech × type (F(2,60) = 3.21, p < 0.047; by items F(2,110) = n.s). In addition, location approached significance (F(1,30) = 3.88, p < 0.0581; by items, F(1,110) = 16.85, p < 0.0001).

The main effect of part of speech was due to subjects being more accurate with auxiliary errors (A'=96.5) compared to determiners (A'=95.0). The main effect of error type was explored using standard planned comparisons. Because of our *a priori* predictions about "error type" differences, the less conservative planned comparisons were used to investigate the main effect of error type over subjects. All other post-ANOVA analyses use the more conservative Newman-Keuls test at p < 0.05.

Subjects were significantly more accurate at detecting transposition errors (mean A' = 96.6) than they were at detecting errors of omission (mean A' = 94.8), with agreement in between (mean A' = 95.9) and not significantly different from either (this effect also held by items, using Newman-Keuls). This result can be summarized as transposition > omission (though note our comparisons of error type at each of the two levels of part of speech, below).

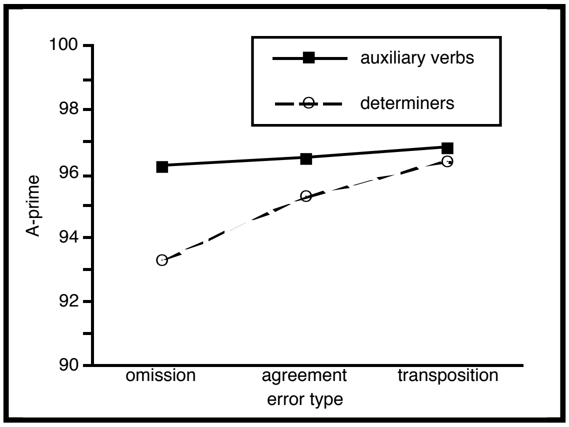

Figure 12. Visual grammaticality judgement experiment: A' by error type and part of speech

Post-hoc tests were used to explore the significant part of speech × type interaction (see Figure 12). Analyzing by part of speech revealed no significant effects of type for auxiliaries, with determiner omissions (mean A' = 93.3) significantly less accurate than determiner transpositions (mean A' = 96.4), over subjects only. Determiner agreement errors were in between (mean A' = 95.3) and not significantly different from either. A post-hoc analysis by type of error showed the part of speech difference to be significant only for omission errors, with auxiliary omission errors (mean A' = 96.2) significantly more accurate than determiner omission errors (mean A' = 93.3).

Reaction times

Two analyses of variance were conducted to evaluate reaction times from the divergence point: an analysis over subjects, and an analysis by items. In addition, analyses of early and late errors were carried out separately, as in Experiments 1 and 2, for a total of four analyses. The subject analyses followed the same 2 × 3 design, with part of speech (auxiliary vs. determiner) and violation type (omission, agreement, transposition) serving as within-subject variables. The item analyses followed a 2 × 3 design, with part of speech and violation type both between-subjects variables.

For early errors, the analysis over subjects yielded one significant main effect, for violation type (F1(2,60) = 58.47, p < 0.0001; F2(2,35) = 29.45, p < 0.0001). Planned comparisons of the main effect of type (see Figure 13) showed that this was due to the pattern omissions > (1777 msec) > transpositions (1612 msec) > agreement errors (1110 msec). This profile was identically significant over items. There was also one significant interaction: violation type × part of speech (F1(2,60) = 3.71, p < 0.031; F2(2,35) = n.s.).

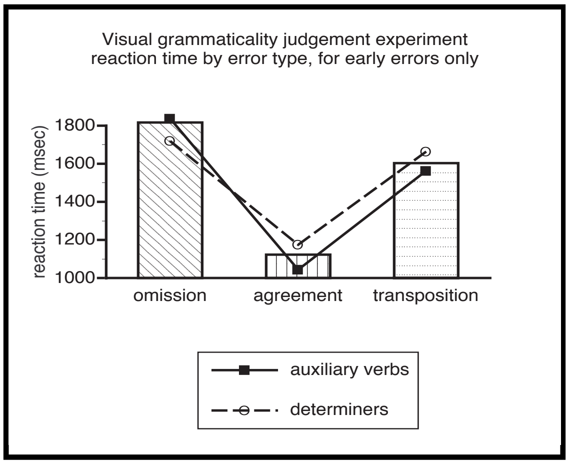

Figure 13. Visual grammaticality judgement experiment: reaction time by error type for early errors only

Post-hoc tests were used to explore the significant interaction (see Figure 13). For the violation type by part-of-speech interaction, for auxiliary verb errors, omissions (1836 msec) were slower than transpositions (1562 msec), which were slower than agreement errors (1044 msec). For determiner errors, there was no significant difference between omission (1719 msec) and transposition (1663 msec) errors, which were both significantly slower than agreement errors (1175 msec). Comparing auxiliary and determiner errors by type of error, the difference between the two was significant only for agreement errors, with auxiliary agreement errors (1044 msec) faster then determiner agreement errors (1175 msec).

Turning now to an analysis of the late errors, there were two main effects, one for part of speech (F1(1,30) = 13.19, p < 0.001, F2(2,35) = 5.44, p < 0.03; auxiliary verbs were at 1052 msec and determiners at 937 msec), and one for violation type (F1(2,60) = 21.45, p < 0.0001; F2(2,35) = 12.44 p < 0.0001). Planned comparisons of the type effect showed the effect transpositions (1160 msec) > omissions (911 msec) = agreement errors (915 msec). This profile was identically significant over items.

Reaction times are relatively fast for all late violations, although post-hoc tests indicate that late transposition errors are still significantly slower than the other two late-error types, which do not differ significantly from one another. As noted with the earlier experiments, these differences probably reflect that sentences with a late transposition error tend to be one word longer after the divergence point than the other late error types, giving the subjects just a little longer to make up their minds if they are so inclined. The pattern of responses differs for early and late errors: for early errors the reaction time trend can be summarized as omission > transposition > agreement. For late errors the trend can be summarized as transposition > omission = agreement.

It is hopefully clear by now that there are marked differences in the pattern of results obtained for accuracy vs. reaction time in this experiment. To examine the nature of the relationship between speed and accuracy in more detail, we calculated the Pearson correlation coefficient between A’ scores and average reaction time across all subjects. This analysis yielded a non-significant correlation of +0.06. We may conclude that there is little evidence for a speed/accuracy trade-off over subjects (i.e., it is not the case that some subjects were sloppier than others, rushing through the experiment). Next we calculated the speed/accuracy correlation across all 82 ungrammatical targets (excluding the two outliers). On this analysis, accuracy was defined as percent correct rejection (recall that A' is not a property of individual items). The resulting correlation was positive and significant at +0.42 (p < 0.001). In other words, there is a speed/accuracy trade-off at the individual-item level. Some items take a longer time to resolve because subjects are being particularly careful; other items are resolved quickly, but they also result in more false negatives (i.e., they are incorrectly accepted as grammatical).

Correlations and partial correlations amongst the three experiments

Despite the complexity of these findings, one conclusion is very clear: **In almost all respects, the reaction time results obtained with this RSVP technique parallel results from Experiments 1 and 2 on the size of the decision region that is observed with word-by-word judgments of grammaticality.** To quantify this intuition, we calculated two Pearson correlation coefficients for all 82 ungrammatical targets (with the two outliers removed), comparing the **mean reaction time** obtained in Experiment 3 with Experiment 1 and Experiment 2's **mean words past divergence point**—the average number of words past the divergence point for each sentence that subjects first produced an "ungrammatical" response. All outlying items were removed before the correlation was run (defined as any item with a score more than 2.5 standard deviations from the mean). For Experiments 1 and 3, this analysis yielded a correlation of 0.75 ($p < 0.0001$). For Experiments 2 and 3, this analysis yielded a correlation of +0.91 ($p < 0.0001$; not surprising, since, as reported above, Experiments 1 and 2 also correlated significantly), which confirms that these two techniques yield very similar results when they are applied to the same sentence stimuli. When the same correlation was run separately for early and late errors, to discount some of the variance caused by late errors having the advantage of end-of-sentence cues, the correlation was still high; for early errors, r = +0.87 ($p < 0.0001$); for late errors, r = +0.83 ($p < 0.0001$).[6]

In addition, we also performed partial correlations to determine whether the two word-by-word methods make independent predictions of the reaction times observed in Experiment 3. When variance from the cloze experiment is removed on Step 1, the partial correlation between Experiment 2 (incremental grammaticality judgment (GJ)) and Experiment 3 (RSVP) on Step 2 is 0.77, indicating that even after all of the predictive information offered by the cloze experiment is accounted for, the incremental GJ experiment still offers additional information about the RSVP experiment. When the contribution of incremental GJ is removed on Step 1, the partial correlation between the Experiment 1 (cloze) and Experiment 3 on Step 2 is 0.02, indicating that after all of the predictive information offered by the incremental GJ experiment is accounted for, the cloze experiment offers (essentially) no additional information about the RSVP experiment. Thus, both the cloze and incremental GJ experiments predict reaction time in the RSVP experiment; however, the incremental GJ experiment is a better predictor, and completely overlaps the predictive information offered by the cloze experiment. Although we cannot be certain why incremental GJ provides a better fit to "one shot", on-line judgments, we offer two possible reasons. First, Experiment 1 has heavy task demands. Subjects tend to provide a "can't complete" response in the cloze experiment sooner than they provide an "ungrammatical" response in the incremental GJ experiment, perhaps because they find it tiring to generate a grammatical completion on each word, and want to get each stimulus (and the experiment as a whole) over with as fast they can without violating the rules of the game. Second, in the cloze task subjects can provide only one response at each word. Therefore, for any one subject sensitivity to, e.g., the number of potential completions still possible is lost.

These results also suggest that our choice of the divergence point for each sentence type is crucial in determining the outcome that is observed with either procedure. At the end of Experiment 2, we concluded that many error types have no identifiable "decision point". Instead, they are resolved across an extended "decision region", marked by ample variation over subjects and items. This was also seen in Experiment 1. And yet, by definition the RSVP technique requires us to assign a single point in time from which all reaction times are measured, a point to which we return in the final discussion.

Summary of results for Experiment 3

The results of Experiment 3 are complementary in many respects to the results observed in Experiments 1 and 2:

1. **Accuracy**. Overall accuracy levels were very high on Experiment 3, averaging around 93% correct rejections for ungrammatical stimuli and 91% correct acceptances for grammatical controls. An analysis of variance on A' scores (which corrects for response bias) suggests that accuracy levels are higher overall for transposition errors. The type × part of speech interaction suggests that the most vulnerable items (i.e., the violations that are most often missed) are those that involve determiner omissions. The apparent disadvantage for determiner omissions was also found in Experiment 2, in the final button press measure. Hence the relative vulnerability of determiner omissions appears to be a robust finding.

2. **Reaction times.** This analysis yielded an array of complex interactions involving location, part of speech and error type. In general, the fastest reaction times come from early violations of agreement and late violations of omission. The slowest reaction times and the largest decision regions come from early auxiliary omissions. Despite their apparent complexity, these reaction time results are quite compatible with results from Experiment 2 on the size and shape of the decision region for each item type. Indeed, these two indices were significantly correlated (+0.91), suggesting that the reaction time results obtained in Experiment 3 are a direct reflection of the size of the decision region for each item type.

CONCLUSION

The purpose of this study was to investigate the time course of grammaticality judgment as a performance domain, applying three different techniques to obtain convergent data: cloze completion, incremental GJ, and judgments of well-formedness. Our results include the finding that some error types are associated with a clear-cut "decision point," while others are best described in terms of a protracted "decision region" with ample variability by items and subjects.

The traditional use of reaction time techniques in cognitive psychology and psycholinguistics has been to ascertain the type and number of putative processes involved in some cognitive operation. For diachronic stimuli, this overlooks an important additional and potentially confounding source of variance: the point at which relevant information becomes available. For example, if active declarative sentences are processed faster than passive negative sentences, measured by reaction time on some task using the sentence, it could be because the active sentence requires fewer transformations between deep and surface structure or it could be because the information needed to successfully complete the experimental task is available sooner in the case of active sentences.

By their very nature, then, reaction time techniques require us to impose two points on what appears to be a continuous landscape: the point from which reaction times are measured, and the point at which the behavior in question takes place. This methodological fact has serious theoretical consequences. Obviously, the pattern of reaction times that we observe is entirely determined by the point at which we decide to start the clock. But, at least for some error types, we have seen that where we start the clock is an uncertain thing. If, for example, we were to use the results of Experiment 2 to design empirical "divergence points," what threshold should we use to indicate where the "point" is—the 50% rejection threshold, the 75% rejection threshold, the 90% threshold? This is further complicated by the variability in the size of the decision region between error types. In other words, the problem is not only that there are differences in *absolute* reaction times depending upon the threshold point used, but also that different choices of threshold will change the *rank order* amongst the different error types. For example, measuring reaction time in Experiment 3 from the divergence point (an early threshold) resulted in the pattern omission > agreement. However, measuring reaction time from a late threshold, such as the 75% threshold, would provide the exact opposite profile, agreement > omission, for agreement errors, with their quick resolution in Experiment 2, would have a 75% threshold at essentially the same point as the divergence point, while omission errors, with their protracted decision region, would have a 75% threshold, and a corresponding point from which reaction times were measured, much later.

What are we to do with this insight? One alternative might be to abandon punctate reaction time techniques altogether, in favor of methods that provide a more faithful picture of the continuous and probabilistic process that underlies judgments of well-formedness. This might include self-paced word-by-word reading, eye movement monitoring, event-related brain potentials, and/or the incremental GJ paradigm used here. Unfortunately, most of these left-to-right methods are costly, and all of them are very time-consuming, generating a large number of data points that are difficult to explore within a standard experimental design. However, one strength of the techniques used in the first two experiments of this paper is that the guesses or word completions that subjects provide at each word provide useful information about the number and range of alternatives that these subjects still have in mind. These completions provide a relatively faithful reflection of the competing alternatives from the subject's point of view.

Our results have shown that for some error types the divergence point is not necessarily at the same place that the experimenter believes it to be—indeed, sometimes there is no one *point* at all. Our results have implications for a promising new research area in psycholinguistics; i.e., the use of event-related potentials (ERP) as an index of sensitivity to semantic or syntactic violations (Hagoort, Brown & Groothusen, 1993; Neville, Nicol, Barss, Forster & Garrett, 1991; Osterhout & Holcomb, 1993; Brown, Hagoort, & Vonk, 1995). At the very least, our results suggest that all future ERP studies (1) pre-test materials using one or a number of the techniques we have uses here in order to empirically determine the divergence point (if there is one); (2) investigate ERPs over the entire course of the sentences, rather than just at the divergence point (e.g., King & Kutas, 1995).

For example, Neville et al. (1991) showed different (though not completely orthogonal) waveforms for semantic anomalies and three different types of syntactic anomalies, using this finding as support for the biological reality of semantic vs. syntactic processes as well as for the three different (Government-and-Binding-theory-motivated) syntactic violation types. In some conditions, the divergence points were quite punctate, like our agreement errors (e.g., phrase-structure violations such as, "The scientist criticized Max's of proof the theorem,") while others had divergence points that were less certain, like our omission and transposition errors (e.g., subjacency violations such as, "What was a proof of criticized by the scientist?"). The data reported ERPs only at the particular word that the experimenters considered to be the divergence point; it is possible (and suggested by our research) that (1) subjects are showing effects of the formedness manipulations at points other than the divergence point and that (2) the waveform differences may be due to the same factors—expectancy and potential alternate completions at the particular point—which we believe to be affecting the dependent variables in our experiments. For example, in the Neville experiment subjects' overt judgment of grammaticality was not uniform across conditions, ranging from 72% correct detection of WH-movement violations to 98% correct detection of phrase-structure violations (although they do not provide significance levels) As our experiments have shown, differences in potential alternate completions *absent* any theory-specific difference between sentences can have effects both on the size of the decision region as well as on the subject's final decision of grammaticality. The differences in judgment of grammaticality in the Neville experiment suggest that this may indeed be at least part of what is happening, and therefore that at least part of the difference in waveforms may be attributable to the effects of expectancy and potential alternate completions.

An ERP study by Hagoort, Brown, and Groothusen (Hagoort et al., 1993) was even more suggestive that a punctate divergence point can not always be assumed in such studies. The authors used both procedures that we have suggested here, pre-testing their materials (though in a serial visual presentation task and not in an incremental GJ task) and measuring and reporting on ERPs throughout the sentence. The pre-testing revealed effects similar to those of our experiment: For some error types subjects responded mostly at the divergence point, while for others responses were more frequent after the divergence point. Waveform differences between ungrammatical and control grammatical sentences revealed significant differences at the divergence point, after the divergence point, (including at the sentence-final position, reminiscent of our own sentence-final elevated-reaction time effects) and in some cases before the divergence point, supporting our contention that one cannot assume, without empirical support, that subjects invariably perceive an ungrammaticality at a particular point.

Implications for aphasia. We chose to study this particular set of violations for two reasons: (1) to determine whether the pattern of errors observed in speech production by aphasic patients can be explained by variations in the degree of sensitivity displayed by normal listeners exposed to the same error types, and (2) to start our on-line investigations of error detection in normals with a well-defined set of minimal contrasts over materials that are comparable in every other respect. With regard to the first rationale, we have uncovered new information about the processing characteristics that may make some errors more vulnerable (i.e., harder to detect) than others, which may in turn help to explain why aphasic patients are more prone to produce those error types. With regard to the second rationale, it is now clear (as we suspected from the outset) that these supposed "minimal contrasts" are really not minimal at all, because these error types (i.e., agreement, omission and transposition) differ markedly in the range of alternatives that are kept open at various points from the divergence point to the end of the sentence.

Starting with the first rationale, it has been known for some time that aphasic patients tend to produce errors involving grammatical function words—although the nature of those errors may vary across different types of aphasia (i.e., more errors of function word omission in non-fluent patients; more errors of agreement, coupled with a tendency toward overuse of function words in the "empty speech" of some fluent patients—for a review, see Bates & Wulfeck, 1989a). Furthermore, some function word errors are very frequent (i.e., agreement errors and omissions), while other are relatively rare (i.e., transposition errors). Building on earlier work in the auditory modality by Wulfeck and her colleagues (Wulfeck & Bates, 1991; Wulfeck et al., 1991; Wulfeck, 1987), we hypothesized that a similar gradient of sensitivity to error types may be found in grammaticality judgments by normal subjects (i.e., less sensitivity to errors of omission and/or agreement; more sensitivity to transposition errors). If this proved to be the case, it would provide support for the idea that aphasic patients suffer from deficits that affect or interact with the process by which normal subjects monitor for errors in their own speech and the speech of others. Experiment 3 provided some support for this view. Although accuracy levels were very high overall, they were generally higher for errors of transposition and lower for omission errors, with agreement errors in between. Hence errors that are rare in aphasia seem to be easy for normals to detect, and errors that are common in aphasia tend to be harder to detect.

There are a number of possible explanations for these error type differences. First, normals and aphasic patients may display greater sensitivity to transposition errors because these errors always involve at least two cues, the "hole" (i.e., the point at which subjects realize that an omission may have occurred) and the displaced element (i.e., the moved element is encountered at an unexpected point).

Second, the advantage of transpositions over omissions might be explained by the number of bigrams violated in each error type. If 2.1 is a grammatical string, 2.2 is a transposition error on that string and 2.3 an omission error. The transposition error has *three* illegal bigrams, "AC", "CB", and "BD", while the omission error has only one, "AC".

2.1 A B C D E

2.2 A C B D E

2.3 A C D E

Third, we have seen that omissions and transposition errors both yield a "decision region" that varies in length depending on a number of factors, while agreement errors are usually resolved within a short time (often corresponding to a single point). It is possible that the lengthy decision process reflected in resolution of omission and transposition errors is experienced subjectively (albeit unconsciously) as a long period of perturbation. By contrast, the period of uncertainty associated with agreement errors tends to be relatively short (assuming that the subject detects this error in the first place). If grammaticality judgment is (as we have proposed) a close relative of the monitoring processes used in language production and comprehension, then we may speculate that long periods of uncertainty are more likely to bring the error above thresholds of attention. That is, "big perturbations" (accompanied by a larger array of alternative completions) may result in better error detection than "small perturbations". This result is consistent with the reaction time differences between error types at the divergence point in Experiment 2, where transpositions and omissions showed a reaction time jump while agreement errors remained faster and relatively constant across the course of the sentence. Thus, aphasic patients (like normal controls) may be vulnerable to agreement errors because the perturbations produced by such errors are harder to detect. In regard to the greater vulnerability of omission errors as compared to transpositions, Elman (personal communication) reports that simple recurrent nets (SRNs) trained to anticipate temporally ordered stimuli also tend to be more sensitive to transposition errors than to omission errors (though he cautions that these findings may not be intrinsic to SRNs but may be dependent upon the particular tasks that he has trained them upon.) When such networks have learned a simple grammar, they are more able to continue successfully with the prediction task (i.e., recover) when the error is an omission rather than a transposition error. This profile, as Elman points out, indicates a sensitivity to *relative* rather than absolute order. The omission error of 2.2 has three elements in the wrong absolute position, "C", "D", and "E", while the transposition error of 2.3 has only two elements in the wrong position, "C" and "B". If these networks (and, by extension, our subjects) were sensitive to absolute order, one would expect the opposite profile of sensitivity, with omissions better detected than transpositions. Should we continue to find that humans and networks show similar task profiles on these sorts of well-formedness judgments, this is good evidence that such models organize and process information in a manner analogous to humans.

Given these three possible explanations, why, then, are omission errors relatively common in aphasia (especially nonfluent aphasia)? One possibility is that the two common error types (agreement and omission) have a different causal base. Agreement errors are "real misses", observed in most often in fluent patients because these patients suffer from what can be characterized as a "speed/accuracy tradeoff" (Bates, Appelbaum & Allard, 1991; Bates & Wulfeck, 1989b; Kolk, 1985; Kolk & Heeschen, 1985; Haarmann & Kolk, 1992). By contrast, omission errors may occur more often in patients who are all too aware of their limitations, patients who produce an omission error (often with complete awareness) in order to get around painful output limitations. This is, as noted, pure speculation at this point—but it is a possibility worth pursuing.

Aside from their implications for neurolinguistic research, our materials were chosen to reflect a set of minimal contrasts. It is now clear that these three error types yield markedly different performance profiles despite their superficial similarities. Agreement (or substitution), omission and transposition errors are often compared and analyzed together in aphasia research because they do appear to form a natural contrast set (e.g., Miceli, Silveri, Romani & Caramazza, 1989). However, in all three of our experiments we have found striking differences in the size and shape of the decision and reaction time functions associated with these error types (omission, agreement, transposition), and with variations in part of speech (auxiliary vs. determiner) and location (early vs. late). From the alternative completions that subjects offered in Experiment 1 (the cloze procedure), we may infer that the critical differences among these stimuli lie in the number and range of alternative completions that the subjects are still willing to entertain. In addition to the well-formed completions that naive subjects provided in the cloze task, we also found completions that ought to be ruled out if subjects were following the rules of their language in a strict fashion (e.g., a restrictive relative clause interpretation should not be possible after a proper noun; a non-restrictive relative clause must be set off by punctuation). In other words, some subjects appear to hesitate in classifying a sentence as ungrammatical because of a partial overlap between ungrammatical stimuli and legal alternatives in the language. Competing alternatives may die away slowly; they are not necessarily eliminated in a stepwise fashion, and they may hang around to cause trouble even though they do not provide a discrete "yes/no" fit to the rules of the language (see also MacDonald, Pearlmutter & Seidenberg, 1994).

This brings us back to the methodological recommendations raised earlier. In particular, we think it would be important to design stimuli that vary consistently in the number and strength of the alternative interpretations that subjects have in mind at each point across the course of the sentence. This approach has been used in studies of sentence comprehension (see, for example, the large literature on "minimal attachment" and other strategies associated with the processing of sentence ambiguities—Taraban & McClelland, 1988; MacDonald et al., 1994; Trueswell & Tanenhaus, 1994; Trueswell, Tanenhaus & Garnsey, 1994; MacDonald, 1994). It seems likely that this approach will be equally useful in the study of grammaticality judgment. The cloze method used in our Experiment 1 may be particularly useful in this regard, to make sure that subjects perceive the same ambiguities that we have in mind in designing our materials.

With this recommendation, we return to the original motivation for on-line studies of grammaticality judgment. For close to fifty years, grammaticality judgments by trained native speakers have been the method of choice for linguists working within the generative tradition. And yet we still know very little about the cognitive processes that underlie such judgments, and thus the factors that may affect them. Our own work has focused on the judgments produced by naive listeners, with sentence materials that are in some sense "pretheoretic" (i.e., they were not designed to discriminate among current theories of syntactic structure). However, we believe that the on-line methods investigated here could provide a useful adjunct to current linguistic research, applied to a richer set of linguistic materials as they are processed by "expert" listeners. Sentences may appear to be more or less grammatical in orthodox linguistic research not because of variations in the number of rules violated (as proposed, for example, by Chomsky, 1965), but rather because of variations in the number, frequency and nature of the possible completions and partially overlapping alternatives that native speakers entertain while each sentence is evaluated.

References

- Bates, E., & Wulfeck, B. (1989a). Crosslinguistic studies of aphasia. In B. MacWhinney & E. Bates (Eds.), *The crosslinguistic study of sentence processing,* pp. 328-371. Cambridge: Cambridge Univ. Press.
- Bates, E., & Wulfeck, B. (1989b). Reply: comparing approaches to comparative aphasiology. *Aphasiology*, *3*(2), 161-168.
- Bates, E., Appelbaum, M., & Allard, L. (1991). Statistical constraints on the use of single cases in neuropsychological research. *Brain and Language*, *40*, 295-329.
- Bates, E., Wulfeck, B., & MacWhinney, B. (1991). Cross-linguistic research in aphasia: An overview. *Brain and Language*, *41*(2), 123-148.
- Bloomfield, L. (1961). *Language*: New York: Holt, Rinehart and Winston.
- Boland, J. E., Tanenhaus, M. K., Carlson, G., & Garnsey, S. E. (1989). Lexical projection and the interaction of syntax and semantics in parsing. *Journal of Psycholinguistic Research*, *18*(6), 563-576.
- Boland, J. E., Tanenhaus, M. K., & Garnsey, S. M. (1990). Evidence for the immediate use of verb control information in sentence processing. *Journal of Memory and Language*, *29*, 413-432.
- Brown, C., Hagoort, P., & Vonk, W. (1995). On-line sentence processing: Parsing preferences revealed by brain responses. *Eighth Annual CUNY Conference on Human Sentence Processing*, Tucson, AZ, March.
- Caplan, D. (1981). On the cerebral organization of linguistic functions: Logical and empirical issues surrounding deficit analysis and functional localization. *Brain and Language*, *14*, 120-137.
- Caramazza, A., & Berndt, R. (1985). A multicomponent deficit view of Broca's aphasia. In M. L. Kean (Ed.), *Agrammatism*. Orlando: Academic.
- Caramazza, A., & Zurif, E. B. (1976). Dissociation of algorithmic and heuristic processes in language comprehension: Evidence from aphasia. *Brain and Language*, *3*, 572-582.
- Carpenter, P. A., & Just, M. A. (1989). The role of working memory in language comprehension. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: the impact of Herbert A. Simon*, pp. 31-6). Hillside, NJ: Lawrence Erlbaum.
- Chomsky, N. (1957). *Syntactic structures*. the Hague: Mouton and Co.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450-466.
- Grier, J. B. (1971). Non-parametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin*, *75*(6), 424-429.
- Haarmann, H., & Kolk, H. (1992). The production of grammatical morphology in Broca's and Wernicke's aphasics: Speed and accuracy factors. *Cortex*, *28*, 97-112.
- Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, *8*(4), 439-483.
- Just, M., & Carpenter, P. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*, 329-353.
- Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*(1), 122-149.
- King, J. & Kutas, K. (1995) Who did What and When… *Journal of Cognitive Neuroscience, 7(3),* 376-395.
- Kluender, R. (1992). *Cognitive constraints on variables in syntax*. Unpublished doctoral dissertation, UCSD.
- Kolk, H. (March, 1985). *Telegraphic speech and ellipsis,* Presented at the Royaumont Conference Centre, Paris, France, March.
- Kolk, H., & Heeschen, C. (1985). *Agrammatism versus paragrammatism: A shift of behavioral control*, Paper presented at the Academy of Aphasia 23rd Annual Meeting, Pittsburgh, PA.
- Kutas, M., & Kluender, R. (1991). What is who violating? A reconsideration of linguistic violations in light of event-related potentials. *Center for Research in Language Newsletter*, *6*(1).
- Levelt, W. J. M. (1972). Some psychological aspects of linguistic data. *Linguistische Berichte*, *17*, 18-30.

- Levelt, W. J. M. (1974). *Formal grammars in linguistics and psycholinguistics. Vol. 3: Psycholinguistic applications*. The Hague: Mouton.
- Levelt, W. J. M. (1977). Grammaticality, paraphrase, and imagery. In S. Greenbaum (Ed.), *Acceptability in language*. The Hague: Mouton.
- Linebarger, M., Schwartz, M., & Saffran, E. (1983). Sensitivity to grammatical structure in so-called agrammatic aphasics. *Cognition*, *13*, 361-392.
- MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*, *9*(2), 157-201.
- MacDonald, M., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review.*, *101*, 676-703.
- Mauner, G. (1992). *Syntactic control and the interpretation of VP anaphors*. Paper presented at the NELS 22.
- Miceli, C., Silveri, M. C., Romani, C., & Caramazza, A. (1989). Variation in the pattern of omissions and substitutions of grammatical morphemes in the spontaneous speech of so-called agrammatic patients. *Brain and Language*, *36*, 447-492.
- Neville, H., Nicol, J. L., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, *3*, 151-165.
- Newmeyer, F. (1980). *Linguistic theory in America*. New York: Academic Press.
- Osterhout, L., & Holcomb, P. J. (1993). Event-related potentials and syntactic anomaly: Evidence of anomaly detection during the perception of continuous speech. *Language & Cognitive Processes*, *8*(4), 413-437.
- Pollack, I., & Norman, D. A. (1964). A non-parametric analysis of signal detection experiments. *Psychonomic Science*, *1*, 125-126.
- 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.
- Sells, P., Shieber, S., & Wasow, T. (1991). *Foundational issues in natural language processing*. Cambridge: MIT Press.
- Shankweiler, D., Crain, S., Gorrell, P., & Tuller, B. (1989). Reception of language in Broca's aphasia. *Language and Cognitive Processes*, *4*(1), 1-33.
- Taraban, R., & McClelland, J. L. (1988). Constituent attachment and thematic role assignment in sentence processing: Influences of content-based expectations. *Journal of Memory and Language*, *27*, 597-632.
- Trueswell, J. C., & Tanenhaus, M. K. (1994). Towards a lexicalist framework of constraint-based syntactic ambiguity resolution. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing*, pp. 155-179. Hillside, NJ: Lawrence Erlbaum.
- Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic disambiguation. *Journal of Memory and Language*, *33*, 285-318.
- Tyler, L. K. (1992). *Spoken language comprehension: An experimental approach to disordered and normal processing*. Cambridge, MA: MIT Press.
- Wulfeck, B., & Bates, E. (1991). Differential sensitivity to errors of agreement and word order in Broca's aphasia. *Journal of Cognitive Neuroscience*, *3*, 258-272.
- Wulfeck, B., Bates, E., & Capasso, R. (1991). A cross-linguistic study of grammaticality judgments in Broca's aphasia. *Brain and Language*, *41*(2), 311-336.
- Wulfeck, B. B. (1987). *Sensitivity to grammaticality in agrammatic aphasia: processing of word order and agreement violations*. Unpublished Doctoral dissertation, University of California, San Diego.

Endnotes

1. Although in some cases comparisons were also made at sentence end and in some cases one word before or after as well.

2. Because the A' score is an index of accuracy designed to correct for response bias, on psychological grounds the A' score can only be analyzed over subjects (i.e. treating subjects as a random variable and items as a fixed effect), and not over items (i.e. treating items as a random variable and subjects as a fixed effect). Thus, all item analyses for accuracy are for percent correct to ungrammatical.

3. Used because we were only comparing two points in each case and because t-tests are less conservative and provide more power than Newman-Keuls.

4. For late auxiliary error reaction times at the "> 20%" interval, just before the divergence point, there was a small but significant difference between agreement errors (861 ms) and the other two error types (omission = 777 ms, transposition = 755). For late determiner errors, judgments showed a slight yet significant increase in omissions (4%, compared to 1% or less) at the ">40%" interval, and reaction times showed for omissions (819 ms) were significantly slower (agreement = 749 ms, transposition = 721 ms); these effects are most likely spurious, as sentences to this point have no structural differences and the variance at this intervals is quite low.

5. Note that we are assuming in what follows that the visual RSVP task is at least in some respects comparable with auditory processing.

6. For Experiments 1 and 3, the correlation for early errors was r = +0.70 ($p \leq 0.0001$), and for late errors r = +0.69 ($p \leq 0.0001$). For Experiments 1 and 2, the correlation for early errors was r = +0.80 ($p \leq 0.0001$), and for late errors r = +0.76 ($p \leq 0.0001$).

Appendix I.a. The 12 types of ungrammatical sentence

part of speech	type of error	location of error	
		early	late
auxiliary	omission	Mrs. Brown working * quietly in the church kitchen.	She is reading that mystery novel that her mother writing. *
	agreement	The writer were * holding a very big party.	While sitting on the couch, Mr. Lane's daughters was * watching a movie.
	transposition	Miss Hope sending * was several green dresses that Lisa had ordered.	While talking to Jane, Joseph knitting * was a sweater.
determiner	omission	Girl * was working quietly near the small, red house.	The small, thin green vine was sprouting flower. *
	agreement	A boys * are driving a large van that the artist has painted.	Larry is saying that his mother was planting that bushes. *
	transposition	Helicopter * a was hovering loudly over the army base.	The girls were watching the stars while camping in desert * that.

Appendix I.b

Examples of error types in the ungrammatical core stimuli. The asterisk indicates the location of the error. The number indicates number of words past the logical error point. The order of error types is omissions, agreement errors, transposition errors in all of the following cells.

Early auxiliary errors

Mrs. -2	Brown -1	working * 0	quietly 1	in 2	the 3	church 4	kitchen. 5			
The -2	writer -1	were * 0	holding 1	a 2	very 3	big 4	party. 5			
Miss -2	Hope -1	sending * 0	was 1	several 2	green 3	dresses 4	that 5	Lisa 6	had 7	ordered. 8

Early determiner errors

Girl *	was	working	quietly	near	the	small,	red	house.			
0	1	2	3	4	5	6	7	8			
A -1	boys * 0	are 1	driving 2	a 3	large 4	van 5	that 6	the 7	artist 8	has 9	painted. 10
Helicopter * 0	a 1	was 2	hovering 3	loudly 4	over 5	the 6	army 7	base.	8		

Late agreement errors

She	is	reading	that	mystery	novel	that	her	mother	written. *		
-9 While	-8 sitting	-7 on	-6 the	-5 couch,	-4 Mr.	-3 Lane's	-2 daughters	-1 was *	0 watching	a	movie.
-8 While	-7 talking	-6 to	-5 Jane,	-4 Joseph	-3 knitting *	-2 was	-1 a	0 sweater.	1	2	3
-5	-4	-3	-2	-1	0	1	2	3			

Late determiner errors

The	small,	thin	green	vine	was	sprouting	flower. *				
-7	-6	-5	-4	-3	-2	-1	0				
Larry -9	is -8	saying -7	that -6	his -5	mother -4	has -3	planted -2	that -1	bushes. * 0		
Those -10	girls -9	were -8	watching -7	the -6	bright -5	lightning -4	while -3	camping -2	in -1	desert * 0	that. 1

Appendix I.c: Core stimuli, organized by cell

Underlined sentences are grammatical controls; bolded sentences are grammatical repeats.

1. Early Auxiliary Omission

1.1. They examining * several expensive old paintings while walking through the art museum.

1.2. They were reading several large maps while waiting for the next train.

1.3. My mother visiting * an expensive and famous plastic surgeon.

1.4. The man was playing both old and modern piano pieces.

1.5. Joan making * several big and tasty ice cream drinks.

1.6. Julie was eating a large, creamy, chocolate and coconut pie.

1.7. My cousin drawing * three small pictures of his mother’s new cats.

1.8. Her mother was reading some old articles on famous Hollywood movie actors.

1.9. The boy taking * a black feather that the pigeon had dropped.

1.10. The doctor is reading the medical report that the nurse has written.

1.11. Mrs. Brown working * quietly in the church kitchen.

1.12. A small boy was walking slowly down the beach.

1.13. Tom’s mother forgetting * that he had taken his new car.

1.14. Several people were saying that fishermen had killed those blue dolphins.

2. Late Auxiliary Omission

2.1. While sitting on the red sofa, her older friend eating * some cake.

2.2. While babysitting for their neighbors, Mrs. Johnson’s daughters were eating some candy.

2.3. Her older brother’s first guest drinking * a beer.

2.4. My young cousin’s very first dinner party guest was making some drinks.

2.5. The two very famous Italian chefs making * a salad.

2.6. The two famous New York chefs were making a cake.

2.7. In the very big and shady front yard, Bill’s mother picking * flowers.

2.8. Near the big, old summer house, several animals were drinking water.

2.9. She is reading that mystery novel that her mother written. *

2.10. They are eating the candy bars that Mrs. Morton has brought.

2.11. The young, new president of John’s college speaking * briefly.

2.12. That very old friend of my father's was walking slowly.

2.13. John’s boss is upset that his new secretary stolen * a typewriter.

2.14. Sam’s friend is saying that his two sisters have made some cookies.

3. Early Determiner Omission

3.1. Boy * was entering a contest while staying at the hotel.

3.2. The girls were eating some fries while waiting for their friends.

3.3. Girl * was eating some dark chocolate ice cream.

3.4. The woman was having a very big dinner party.

3.5. Clerk * was reading several very old and important letters.

3.6. The woman was painting several very large, colorful pictures.

3.7. Woman * was watching some orange butterflies in the small back garden.

3.8. **Her mother was reading some old articles on famous Hollywood movie actors.**

3.9. Woman * is visiting the old dairy farm that her father has bought.

3.10. The clerk is sending several cotton shirts that Dorothy’s mother has ordered.

3.11. Girl * was working quietly near the small, red house.

3.12. The balloon was floating slowly through the air.

3.13. Woman * was saying that her husband had bought several big tomatoes.

3.14. The man was reading that many people had protested those new taxes.

4. Late Determiner Omission

4.1. The boy was finding many big sea shells while playing on beach. *

4.2. **They were reading several large maps while waiting for the next train.**

4.3. My new blue and green silk ball gown was costing fortune. *

4.4. The large and pale gray cruise ship was hitting an iceberg.

4.5. The small, thin green vine was sprouting flower. *

4.6. Her two favorite great aunts were making some pie.

4.7. Alice was calling her old college friend at hotel. *

4.8. Martha was bringing several old dance records to the party.

4.9. The maid whom Sally has hired is cleaning bathroom. *

4.10. The woman whom Anne's father has hired is cleaning the windows.

4.11. Two very famous art critics were speaking briefly at museum. *

4.12. A plane was flying slowly over the old landing strip.

4.13. The woman was writing that her two daughters had bought car. *

4.14. The train conductor was saying that some trash had blocked the tracks.

5. Early Auxiliary Agreement

5.1. The women was * drinking some wine while talking about the movie.

5.2. **The girls were eating some fries while waiting for their friends.**

5.3. The writer were * holding a very big party.

5.4. **The man was playing both old and modern piano pieces.**

5.5. The vine were * growing a few red and yellow flowers.

5.6. **Julie was eating a large, creamy, chocolate and coconut pie.**

5.7. The men was * reading those papers on the train.

5.8. **Martha was bringing several old dance records to the party.**

5.9. She were * seeing the place where her two older sisters had worked.

5.10. They were visiting the house where Nancy's parents and grandparents had lived.

5.11. Soap bubbles was * floating slowly into the summer sky.

5.12. Honey bees were flying loudly around a large, old oak tree.

5.13. Mike's parents was * hoping that he had passed the final exam.

5.14. **Several people were saying that fishermen had killed those blue dolphins.**

6. Late Auxiliary Agreement

6.1. While sitting on the couch, Mr. Lane's daughters was * watching a movie.

6.2. **While babysitting for their neighbors, Mrs. Johnson's daughters were eating some candy.**

6.3. Some famous old Hollywood actor were * having a party.

6.4. Several very young children were watching a play.

6.5. The old, red brick houses was * blocking the view.

6.6. **Her two favorite great aunts were making some pie.**

6.7. In Mrs. Hart's small rose garden, the gardener were * planting bushes.

6.8. **Near the big, old summer house, several animals were drinking water.**

6.9. John is eating the pizza that his mother have * made.

6.10. **They are eating the candy bars that Mrs. Morton has brought.**

6.11. In the bank's very large lobby, the men was * talking quickly.

6.12. In a big, old, red boat, two girls were rowing slowly.

6.13. Susan is saying that she have * cleaned it.

6.14. Chris is saying that his mother has bought a house.

7. Early Determiner Agreement

7.1. Those girl * was visiting Jack while driving through the town.

7.2. The boy was reading a comic book while standing on the corner.

7.3. A women * were watching the Fourth of July fireworks.

7.4. **The woman was having a very big dinner party.**

7.5. Two woman * was selling several expensive imported gowns.

7.6. Those models were wearing that new wave hairstyle.

7.7. A boys * were feeding the small, brown bird in the yard.

7.8. Those girls were petting the small, brown cat in the yard.

7.9. A boys * are driving a large van that the artist has painted.

7.10. **The clerk is sending several cotton shirts that Dorothy's mother has ordered.**

7.11. Those house * was selling quickly, for very little money.

7.12. **The balloon was floating slowly through the air.**

7.13. Several sailor * was saying that the man had predicted a storm.

7.14. **The man was reading that many people had protested those new taxes.**

8. Late Determiner Agreement

8.1. Jim's sisters were watching the ocean waves while sitting on that rocks. *

8.2. Mrs. Taylor was eating a turkey sandwich while talking on the phone.

8.3. The very famous rock singer was performing several song. *

8.4. **My young cousin's very first dinner party guest was making some drinks.**

8.5. Mr. Hall's entire class was watching several cartoon. *

8.6. **The two famous New York chefs were making a cake.**

8.7. Arthur's daughters were driving that red sports car over those mountain. *

8.8. **Those girls were petting the small, brown cat in the yard.**

8.9. Several workers whom Mr. Stevens has hired are painting those fountain. *

8.10. **The woman whom Anne's father has hired is cleaning the windows.**

8.11. The young man was speaking loudly with two salesman. *

8.12. **A small boy was walking slowly down the beach.**

8.13. Larry is saying that his mother has planted that bushes. *

8.14. **Chris is saying that his mother has bought a house.**

9. Early Auxiliary Transposition

9.1. Jane's friends watching * were some fireworks while standing on the hill.

9.2. **Mrs. Taylor was eating a turkey sandwich while talking on the phone.**

9.3. Those girls seeing *were some old and famous silent movies.

9.4. The artists were selling several small but expensive watercolor paintings.

9.5. She signing * was her newest and biggest story collection.

9.6. **The woman was painting several very large, colorful pictures.**

9.7. Students writing * are several math problems on the blackboard.

9.8. Jane's mother is renting a small apartment in New York.

9.9. Miss Hope sending * was several green dresses that Lisa had ordered.

9.10. Jan's hairdresser was learning a new look that Jan had wanted.

9.11. The boy walking * was quickly to the store.

9.12. **The balloon was floating slowly through the air.**

9.13. That woman saying * is that her two friends have stolen several things.

9.14. **Sam's friend is saying that his two sisters have made some cookies.**

10. Late Auxiliary Transposition

10.1. While talking to Jane, Joseph knitting * was a sweater.

10.2. **While babysitting for their neighbors, Mrs. Johnson’s daughters were eating some candy.**

10.3. A small and harmless black dog chasing * was chickens.

10.4. **The large and pale gray cruise ship was hitting an iceberg.**

10.5. My old junior high school friend's favorite little cousin watching * was cartoons.

10.6. My old army friend's beautiful, bright red sports car was burning oil.

10.7. In music class, two students singing * were songs.

10.8. **Near the big, old summer house, several animals were drinking water.**

10.9. Horses are eating the sugar cubes that Martin brought * has.

10.10. **They are eating the candy bars that Mrs. Morton has brought.**

10.11. In a large, old, silver car, several boys driving * were recklessly.

10.12. **In a big, old, red boat, two girls were rowing slowly.**

10.13. Those pilots were saying that several clouds covered * had the sky.

10.14. **The train conductor was saying that some trash had blocked the tracks.**

11. Early Determiner Transposition

11.1. Man * that was reading some books while staying at the hotel.

11.2. **The boy was reading a comic book while standing on the corner.**

11.3. Guest * the was eating a cheese and sausage pizza.

11.4. **The artists were selling several small but expensive watercolor paintings.**

11.5. Students * several were buying some cheap French cheese.

11.6. **Those models were wearing that new wave hairstyle.**

11.7. Women * three are opening a small shop in the city.

11.8. **Jane’s mother is renting a small apartment in New York.**

11.9. President * the was reading the report that his advisor had written.

11.10. **The doctor is reading the medical report that her nurse has written.**

11.11. Helicopter * a was hovering loudly over the army base.

11.12. **A plane was flying slowly over the old landing strip.**

11.13. Announcer * the is saying that a big accident has blocked one lane.

11.14. **Sam's friend is saying that his two sisters have made some cookies.**

12. Late Determiner Transposition

12.1. Those girls were watching the bright lightning while camping in desert * that.

12.2. **Mrs. Taylor was eating a turkey sandwich while talking on the phone.**

12.3. The art museum's owner was buying paintings * several.

12.4. **Several very young children were watching a play.**

12.5. George's two remaining dinner guests were drinking wine * some.

12.6. **Her two favorite great aunts were making some pie.**

12.7. The magazine reporter was donating one hundred dollars to hospitals * those.

12.8. <u>The police officer was giving a speeding ticket to that guy.</u>

12.9. The man whom Jack's sister has dated is cleaning car * the.

12.10. **The woman whom Anne's father has hired is cleaning the windows.**

12.11. Some drunk men were dancing wildly in streets * the.

12.12. **A small boy was walking slowly down the beach.**

12.13. Jerry is hoping that his friends have visited doctor * a.

12.14. **Chris is saying that his mother has bought a house.**

Appendix I.d: Fillers

Grammatical fillers

1. She instructed her secretary to hold all calls.
2. A jeep, the local beach guard noticed, was driving down to the water.
3. Those teachers were reading.
4. Driving down the road, he passed a huge grove of pecan trees.
5. Sherry was eating a pie.
6. Steve said that he was promoted quickly because he had worked so hard.
7. Sally believed that she had a detailed knowledge of car engines.
8. The most recent of the conferences differed from those others on several important points.
9. Mr Harrison, the first successful publisher and editor of the Times, would seem to be one of these entrepreneurs.
10. The weather a week ago Saturday, rain, and lots more rain to come, was depressing.
11. They have talked to John.
12. The man displayed a fuzzy toy that delighted the young child.
13. Once again, Rob planned his vacation late.
14. They were watching some movies.
15. While the economy appears sluggish, certain parts are improving.
16. What began worrying people in town was the opening of a third huge and sprawling shopping mall.
17. We saw, while visiting the dairy farm, a Holstein cow.
18. By the time Mrs. London was through, the restaurant had become one of the most popular spots in town.
19. Don spoke to her and laughed.
20. She was trying to fix up the car.
21. Jim's cousin was on Jack's mind.
22. Joy noticed several blue dolphins were playing in the water.

Ungrammatical fillers

1. Ellen read, while traveling on the train, several large and complicated company technical reporters.
2. Sam appeared to be thinking hardly.
3. I have remembering that particular watercolor painting because of its sharp and vivid blues and greens.

4. Holds in the ship is so big that you could store a house.
5. Jack was fixing car a.
6. Last weeks, Mary and her two brothers saw a bald eagle flying over the Foothills Fashion Mall.
7. Jane have walked.
8. Will talked has to her.
9. Walk to that houses.
10. A horse were running.
11. Mrs Jones was claiming that by the age of two her daughter Carol walking and talking was in full sentences.
12. Three cats drinking.
13. One in my friends is often working quite late.
14. Those film director was protesting the destruction of the Amazon rain forest, as are many well-known artists and writers.
15. One of Jane's dogs are often playing in the yard.
16. John seemed to be thinking as he walked aloud.
17. Other administration officials calls the Green Berets hostages.
18. She went at that direction, passing one car as she walked.
19. Several in the books, said the librarian, were unsuitable for young children.
20. Her report was so well written that she receiving a promotion.
21. As a resulting of her flight delay, Sam's mother was staying in New York an extra night.
22. Three thousand dollars were the minimum bid set by the art gallery.

Appendix II

Sentences were drawn at random from a pool of seven different sentence types. Each cell of the design received one of each of the following sentence types. A sentence demonstrating each sentence type is also given.

1. While clause: “John was eating some cake while talking to Mary.”
2. SVO with heavy object: “Her husband was picking a few small, white and yellow daisies.”
3. SVO with heavy subject: “My little six-year-old cousin was watching cartoons.”
4. SVO with prepositional phrase: “My friend was reading the paper on the express bus.”
5. Relative clause: “Meg was reading the book that her mother had written.”
6. SV-prepositional phrase with adverb: “A balloon was floating slowly to the ground.”
7. Subordinate clause: “Jack was saying that the teacher has graded the tests.”