Journal of Psycholinguistic Research, Vol. 25, No. 1, 1996

997

Tasks and Timing in the Perception of Linguistic Anomaly

Janet Dean Fodor,[1,2] Weijia Ni,[2,5] Stephen Crain,[2,3] and Donald Shankweiler[2,4]

Three experiments were conducted to investigate the relative timing of syntactic and pragmatic anomaly detection during sentence processing. Experiment 1 was an eye movement study. Experiment 2 employed a dual-task paradigm with compressed speech input, to put the processing routines under time pressure. Experiment 3 used compressed speech input in an anomaly monitoring task. The outcomes of these experiments suggest that there is little or no delay in pragmatic processing relative to syntactic processing in the comprehension of unambiguous sentences. This narrows the possible explanations for any delays that are observed in the use of pragmatic information for ambiguity resolution.

INTRODUCTION

Suppose there were experimental results which showed, without any shadow of a doubt, that during ambiguity resolution there is a brief interval in which

The research reported in this paper was supported by Program Project Grant HD-01994 to Haskins Laboratories from the National Institutes of Child and Human Development, and in part by a grant to Stephen Crain and Donald Shankweiler from the University of Connecticut Research Foundation. We thank Jacques Mehler and Gerry Altmann for helpful information about speech compression algorithms. We are especially grateful to Ignatius Mattingly for his generous expert advice on speech compression and design of the experimental materials, and also for recording all the speech stimuli. Julie Boland, Brian McElree, Janet Nicol, and Martin Pickering provided very helpful advice on an earlier draft.

[1] CUNY Graduate Center, New York, New York 10036.
[2] Haskins Laboratories, New Haven, Connecticut 06511.
[3] University of Maryland, College Park, Maryland 20742.
[4] University of Connecticut, Storrs, Connecticut 06269.
[5] Address all correspondence to Weijia Ni, Haskins Laboratories, 270 Crown St., New Haven Connecticut 06511.

0090-6905/96/0100-0025$09.50/0 © 1996 Plenum Publishing Corporation

only grammatical principles (syntax and semantics) are operative, with effects of discourse, plausibility, and world knowledge not occurring until later. If this could be demonstrated, would it entail that the language processing mechanism is a module? Or that, within the language system, syntax is autonomous? Some discussions imply that it would.

For example, Mitchell, Corley, and Garnham (1992) presented data showing the absence of a rapid contextual influence in resolving the temporary ambiguity in sentences such as (1), and then concluded (p. 85, our emphases): "The findings show that discourse information is *ignored* at first, even though this information becomes *available* (in the context paragraph) well before the point at which it could usefully have made a contribution to the process of ambiguity resolution."

(1) The politician told the woman that had been meeting him that he was going to see the minister.

We agree completely with the logic of this conclusion: To prove modularity of linguistic processing one would show that outside-the-module information (a) is available but (b) is ignored, at least temporarily. But the meaning of the word *available* is crucial here. It has to mean "available to the parsing routines," not just "present in the situation and potentially accessible to any device with the ability to calculate it." The information has to be available inside the perceiver, at the time in question.

There is no guarantee that this will be so. One of the motivating arguments presented by J. A. Fodor (1983) in favor of a mental module for language was that thinking (the kind of central inferencing that draws on knowledge of the world) is open-ended and is therefore apt to be slow. Syntactic processes (and semantics in the strict sense) are narrower in scope, and therefore can proceed much more quickly—which is why it is good design for them to be unshackled from central processes and allowed to go at their own pace. But notice how the reasoning here turns around on itself. If pragmatic processing is *inherently* slower than syntactic processing, then there is no need to invoke the modularity thesis to explain an experimental finding that pragmatic processing lags behind syntactic processing. To put it more strongly: It would be improper to claim such a time lag as evidence for a modular structure of the mind.

A note on terminology is needed here. For brevity, we will be contrasting what we call syntactic processing and pragmatic processing. But these terms have been used with various meanings, and the choice of interpretation makes a difference to the empirical content of the modularity thesis for language. In what follows we will use the word *syntax* as a shorthand to denote syntax and formal compositional semantics and those aspects of reference, etc., that fall within the grammar. We will use the term *pragmatics* for matters of plausibility and all inferencing based on general knowledge and information provided by the discourse.

The substantive point we wish to make is that, if pragmatic inferences are just slow to be computed, there will be a delay before the use of pragmatic information whether or not the mental architecture is modular. This is what we call *de facto modularity*: Syntax precedes pragmatics as a matter of fact, for whatever reason. Given a finding of de facto modularity, further work is needed to establish its source. In order to argue that the use of pragmatic information in parsing is held up on principled grounds, it would be necessary to show that, whatever delay there might or might not be in computing the relevant pragmatic properties, there is an even greater delay in the use of them in making parsing decisions.

Mitchell et al.'s (1992) experiment did not include a check on whether the potentially disambiguating pragmatic information in the discourse had been processed by subjects in time to be useful, but the point can be illustrated with an experiment by Ferreira and Clifton (1986). They tested sentences as in (2).

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

Of interest here is the word *examined*, when it follows *defendant* and when it follows *evidence*. Following *defendant*, it is sensible (on both structural analyses), and the reading-time data showed that it was easy to process. Following *evidence*, *examined* is anomalous (if they are parsed as subject and predicate), and reading times showed that difficulty was high here. This reading-time difference on *examined* shows that the incongruity of the subject–verb analysis for *evidence examined* has already been detected at the word *examined*. With this established, it becomes interesting that the *evidence examined* sentence is still difficult at the *by*-phrase, which syntactically blocks the subject–verb analysis and forces a shift to the reduced relative analysis. At the *by*-phrase, Ferreira and Clifton's data indicated that the *evidence examined* sentence and the *defendant examined* sentence were both significantly harder to process than their unambiguous controls. Thus apparently the anomaly of *evidence examined* is recognized, but this information is not used right away to shift attention to the other analysis. This is the classic syntactic autonomy demonstration.

Later versions of this experiment have had different outcomes (cf. MacDonald, 1994; Trueswell, Tanenhaus, & Garnsey, 1994) and we will not dwell on the data here. It is also true that some methodological problems remain, problems which are difficult to avoid in experiments on modularity.

For instance, a positive result in the availability test [at *examined* in (2b)] creates a disturbance of processing; therefore a finding of difficulty at the disambiguation point [the *by*-phrase in (2b)] might be due merely to persistence of this disturbance over the next word or two before it declined. In that case it would not show that the garden path continued until the *by*-phrase, and so it would not show non-use of available pragmatic information prior to that point.[6] This account of the data could be eliminated by delaying the disambiguation point for several more words, until the upset due to *examined* has demonstrably declined. But by a psycholinguistic Catch-22, a delayed test point might be too late to detect a genuine garden path if one were present, since all models allow pragmatic information to have an influence at some point. Thus some delicate adjustments of materials or presentation might be needed in order to provide proof that processing difficulty first rises (at anomaly detection), then falls to baseline, and then rises again (at garden path detection). Also, Martin Pickering (personal communication) has pointed out the importance of ensuring that this double sensitivity occurs within trials; it is not sufficient that for some subjects or on some trials the anomaly is detected, and for other subjects or trials the garden path occurs. To satisfy all these practical demands is not easy. Nevertheless, the basic logic of Ferreira and Clifton's (1986) experimental design is surely just right. We believe that this logic should be applied to all purported modularity demonstrations: Unless the availability criterion for pragmatic information is shown to be satisfied at the relevant point in the sentence, that test point may reveal de facto modularity but it cannot provide evidence for a modular architecture for language.

[6] It might be countered that the recognition of anomaly (e.g., of *evidence examined*) at the ambiguous attachment site is itself evident for syntactic autonomy. The argument would be as follows: Anomaly detection at *examined* indicates that the attachment ambiguity for *examined* was resolved on a syntactic basis and only then subjected to pragmatic evaluation; if the more plausible reduced relative analysis of (2b) had been selected immediately, there would have been no anomaly to detect. If correct, this would undercut our claim in this paper, that it is necessary to give *independent* evidence of timely access to pragmatic information. However, this counterproposal presupposes that processing is slowed by an anomaly only if the parser has *adopted* the anomalous analysis. If, on the other hand, processing can be slowed by the anomaly of any analysis that the parser is contemplating as an option, then a difficulty at *examined* would not prove that a single analysis had been selected (by syntax) at that point. Hence, difficulty at *examined* (ignoring data for the *by*-phrase) would be equally compatible with a "weakly interactive" parser, in which syntactic analyses are computed in parallel and pragmatic factors select between them (cf. Crain & Steedman, 1985). But then the data for *examined* would not bear straightforwardly on modularity, since a weakly interactive parser is also architecturally modular.

AVAILABILITY OF PRAGMATIC INFORMATION: EXPERIMENTAL EVIDENCE

Rationale

Despite years of experimental research (see Tanenhaus & Trueswell, 1994, for a recent review), there is still considerable disagreement about whether the data do in general support a modular or nonmodular organization of the language faculty. We take no stand here on which interpretation of the experimental findings is correct; our concern in this paper is solely with the methodological issues involved in testing modularity or autonomy hypotheses. It seemed possible to us that neglect of the availability factor might be responsible for some of the apparent inconsistencies, especially if the time course of pragmatic processing is not constant but varies with the complexity of the inference, the speed of sentence presentation, and so forth. It could be useful, therefore, to establish how rapidly pragmatic implausibilities are computed, independently of the issue of how rapidly they are put to work in resolving syntactic ambiguities.

Eye movement studies like those by Ferreira and Clifton (1986) certainly seem to show sensitivity to pragmatic problems within as little as 300 to 400 ms. And in event-related brain potential (ERP) studies the negative shift at approximately 400 ms in response to pragmatic anomalies suggests much the same time frame. But current evidence is incomplete in several respects, and particularly so with respect to the speed of pragmatic processing *relative* to syntactic processing.[7] Few studies have compared pragmatic and syntactic processing in situations other than ambiguity resolution. Recent exceptions are the work of McElree and Griffith (1995) and Boland (1996). McElree and Griffith used an anomaly judgment task in which they forced the pace by training subjects to respond promptly at a signaled time following the sentence. This resulted in a speed–accuracy trade-off (SAT) whose time course could be tracked. Using this method, McElree and Griffith found a delay of pragmatic judgments relative to syntactic judgments; their results ‘‘indicated that thematic role violations began to be detected 50–100 ms later than either constituent structure or subcategorization vio-

[7] In ERP studies, N400 responses to pragmatic anomaly contrast with late positivity (P600) in response to syntactic anomalies (Osterhout & Holcombe, 1992; Osterhout, Nicol, McKinnon, Ni, Fodor, & Crain, 1994). This probably should not be taken as evidence that pragmatic processing is faster than syntactic processing. Possibly the P600 represents a reanalysis process which follows breakdown of the first-pass parse, as suggested by Friederici, Mecklinger, Steinhauer, and Hahne (1995). If so, it does not illuminate current concerns.

lations'' (p. 152). The speed–accuracy trade-off procedure is an interesting innovation in experimental methods for investigating modularity, and it permits estimates of very small temporal differences. However, for reasons discussed below it is not entirely certain that the observed lag in pragmatic judgment reflects a lag in pragmatic anomaly detection per se (see section on Experiment 3 below for discussion). Boland (1996) used a cross-modal naming task with auditory sentence fragments (at normal speed) including examples such as (a) *Which necklace did Nancy describe* and (b) *Which salad did Jenny toss*, followed by visual presentation of a name such as *Bill*. If the name is integrated into the sentence, it creates a syntactic anomaly with (a) and a pragmatic anomaly with (b). The data for these sentence versions appear to show a stronger and earlier influence on naming response time of the syntactic anomaly; the pragmatic anomaly version shows a significant influence on naming only for targets presented 300 ms later than the offset of the verb. (Note: The auditory input stopped at the end of the verb.) However, Boland's own interpretation of her data, including a comparison of naming and lexical decision on the visual stimulus, is that the naming task is not sensitive at all to pragmatic anomaly for reasons other than its temporal properties.

In the research reported here, our strategy for detecting any small timing differences that may exist in the availability of information during sentence comprehension was to put the processing mechanism under severe pressure by using compressed speech input. This approach was proposed some years ago by Chodorow (1979), but his suggestion has not previously been followed up. Chodorow's rationale for using compressed speech was as follows (p. 88): ''If component processes in comprehension have different performance characteristics (e.g., resource requirements or maximal rates) then increasing the overall rate of input ought to affect them differentially. In this way, we might reasonably expect to be able to pull apart otherwise entangled components.'' Chodorow compared lexical and syntactic processing, not syntactic and pragmatic processing. For syntactic processing his experiments showed that, when speech input is speeded up to twice the normal rate, the processor cannot keep up but builds up a backlog of processing to be done after the sentence is over.[8] His results showed that sentence recall declined for compressed input relative to normal input but, at least

[8] On the basis of the data previously available on sentence processing at approximately 50% compression, Chodorow observed (1979, p. 95): ''Early investigations . . . of time-compressed speech revealed that intelligibility of individual words spoken in isolation remains quite high. . . . By contrast, *comprehension* of passage of connected text . . . is very poor at such compression rates. . . . These results mirror the subjective feeling often reported by listeners who experience a sensation of 'falling behind' the input when they are given time-compressed passages of text.''

for unambiguous sentences, the provision of extra time between the compressed input and the recall task allowed performance to return to normal levels. The extent of the backlog was bracketed between 200 ms (since this provided insufficient time for recovery) and 750 ms (which was shown to be sufficient). Based on this finding for syntax, we speculated that forcing the pace of processing by using compressed speech would exaggerate any difference in the timing of syntactic and pragmatic processes, so that, if such a difference exists, it would be measurable however slight it normally is. Having set up this situation where a pragmatic delay is most likely to occur, and having shown that our methods could detect it, it would then be possible to vary the task demands in various ways to see under what circumstances the delay would diminish or disappear. At issue would be whether it fluctuates widely enough to account for some of the apparently contrary outcomes of ambiguity resolution experiments. In fact, we have not completed this broad program of investigation, and for reasons that will become clear below, it has somewhat changed its character as a result of the data reported here.

The experimental sentences were the same for all experiments described below, except for minor differences of constituent length as necessitated by different tasks. They were not ambiguous, either globally or temporarily. They contained syntactic anomalies or pragmatic anomalies of a kind similar to those commonly used in experiments on ambiguity resolution to disqualify one of the potential analyses, as illustrated in (3):

(3) It seems that the cats from across the road . . .

Syntactic anomaly	*won't eating*
Pragmatic anomaly	*won't bake*
Baseline (no anomaly)	*won't eat*

. . . the food that Mary puts out on the porch every morning as soon as she gets up.

The syntactic anomaly always involved a modal verb followed by an *-ing* form. The pragmatic anomalies involved unsuitable pairings of agents and actions, such as *cats–bake*, compared with the acceptable *cats–eat*. The unsuitability was often but not always due to a mismatch of animacy. The anomalously paired noun and verb were also not just low associates, compared with high associates in the acceptable baseline sentence. For example, the anomalous pairing *kangaroo–swear* contrasted with the baseline pairing *kangaroo–sit*, though the latter are not high associates; also, the anomalous *songs–learn* contrasted with the baseline *songs–tend*, which is not a high-associate pair. The anomaly did not depend on words following the verb: For example, cats baking something is anomalous regardless of what is being baked. No pretest was run to validate that the anomalous versions were indeed perceived by subjects as anomalous; the results of the main experiments sufficiently confirmed that they were. Thirty matched triples of experimental sentences like (3) were used in all three experimental paradigms reported here. The verbs were matched for mean frequency across sentence versions (pragmatic anomaly version 89.46; baseline version 79.96; Francis & Kučera, 1982). In each experiment, 30 experimental sentences were presented to a subject, 10 of each of the three versions (syntactic anomaly, pragmatic anomaly, baseline). They were dispersed semirandomly among 76 filler sentences, with at least one filler between neighboring experimental sentences. The filler sentences were of various syntactic types; some were similar and most were dissimilar to the experimental sentences; 58 were acceptable and 18 were anomalous in various ways (additional or missing words, noun/verb substitution, etc.). Half the subjects received the sentences in one order, and the other subjects received them with the first and second halves of the list interchanged. (Order of presentation was not a significant factor in any of the outcomes and will not be discussed further.) A practice session preceded each experiment.

Experiment 1 was an eye movement study designed to establish that our materials gave results similar in temporal profile to those standardly reported in the literature. Experiment 2 used compressed speech input to put the processing routines under time pressure; the time of detection of the anomaly was established by measuring its effect on a concurrent lexical decision task. Experiment 3 used the same compressed speech input but required subjects to monitor for anomalies, as in several recent studies with visually presented materials. [See for example, Boland, Tanenhaus, & Garnsey (1990) and Boland, Tanenhaus, Garnsey, & Carlson (1995) for discussion of the "stop-making-sense" task.] We would emphasize that none of these experiments was designed to resolve the modularity issue for language processing. Their aim was merely to clear away one possible source of uncertainty so that modularity questions can be asked and answered more clearly.

Experiment 1: Eye Movements

Due to space limitations we present Experiment 1 in outline only. Further details and all statistical analyses are reported in Ni, Fodor, Crain, and Shankweiler (1996). Sentences as in (3), adjusted to a maximum length of 76 characters, were presented visually on a single line and eye movements were recorded. Subjects were 24 college students who were paid for their participation. For purposes of data analysis the sentences were divided into regions as shown in (4):

(4) It seems that/the cats/won't usually/VERB the/food we/
1 2 3 4 5
put on the porch.
6

Region 1 was the beginning of the sentence prior to region 2; it was 0 to 3 words long. Region 2 consisted of the last two words of the subject NP of the main verb; this was always the head noun plus a preceding word (determiner or adjective). Region 3 was the modal verb plus a following adverb (inserted for purposes of this experiment to make it possible to measure regressions from the main verb to the modal). Region 4 was the main verb plus the next word, regardless of category. Region 5 was the next two words, regardless of category. Region 6 was the remainder of the sentence, 0 to 4 words.

First-pass residual reading times[9] and percent of regressions (i.e., the percentage of all first-pass fixations which resulted in regression to a prior region) are shown in Figs. 1 and 2. At the verb (region 4), there was no effect of either the syntactic or the pragmatic anomaly on reading times. However, there was an increase in regressions compared with baseline in

[9] See Trueswell et al. (1994) for the rationale for using this measure.

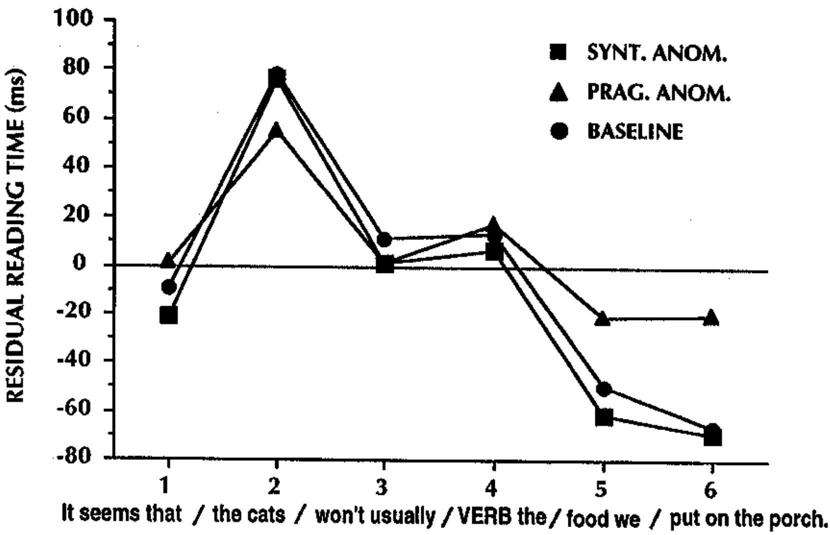

Fig. 1. Experiment 1: Mean first-pass residual reading times. SYNT. ANOM. = syntactic anomaly; PRAG. ANOM. = pragmatic anomaly.

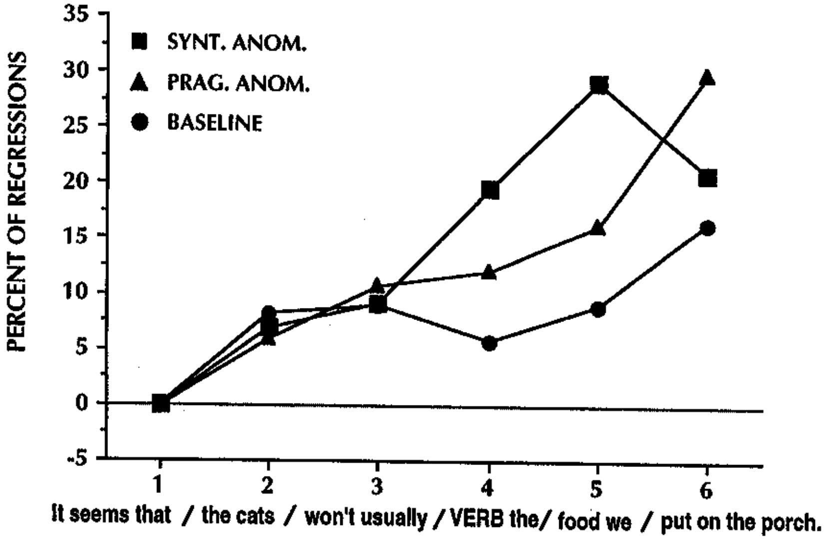

Fig. 2. Experiment 1: Percent of regressive eye movements. SYNT. ANOM. = syntactic anomaly; PRAG. ANOM. = pragmatic anomaly.

this region for both anomaly types, though only approaching significance for the pragmatic anomaly. The difference between the syntactic and pragmatic effects is not significant. Thus, eye movement responsiveness to both types of anomaly was essentially immediate and simultaneous, within the limits of measurement imposed by the paradigm; that is, at least within the time taken to read region 4 (mean 401.55 ms). To obtain finer information about timing it would be necessary to divide the sentence into even smaller regions, but since readers do not normally fixate on every word, results would become erratic and would be complicated by uncertainty about the extent of parafoveal effects.

For the remainder of the sentence (regions 5 and 6), responsiveness to the two types of anomaly was differently distributed between reading time and regressions. For the syntactic anomaly, reading times did not depart significantly from baseline at any point. For the pragmatic anomaly, by contrast, reading times rose relative to baseline at region 5 and remained high at region 6. The difference fell just short of significance for each region separately, but for both regions combined it was significant. For syntactic processing, regressions increased, relative to baseline, at region 5 but dropped to baseline level by region 6. For pragmatic processing, by contrast, regressions continued to increase relative to baseline throughout the sentence.

In summary: The syntactic anomaly had little effect on first-pass fixation durations; it caused an immediate increase in regressions, but this was short-term. The effects of the pragmatic anomaly were divided between reading time and regressions, and both effects became progressively stronger as the sentence continued. It seems reasonable to attribute these qualitatively different patterns to different strategic responses by the parsing mechanism to on-line problems. (We do not mean to imply by this that eye movements are controlled by deliberate decisions on the part of the reader, but merely that some part of the mental mechanism involved in eye movement planning and control is responsive to the outcomes of higher level linguistic processing.) The syntactic anomaly was easily detectable. It triggered immediate regression to check the source of the problem, and then, since the error did little damage to sentence comprehension, there could be a quick return to normal processing. For the pragmatic anomaly, we suppose that the parser had some uncertainty about the source of the problem, or some hope that it might be resolved by later words in the sentence. So though it did check back to confirm that all was not well, its main strategy was to keep pressing forward, but slowly, hoping that matters would eventually resolve themselves. At the end of the sentence, when it became clear that no resolution was forthcoming, regressions took an upward turn. Thus in general, the results of Experiment 1 show rapid sensitivity to both types of anomaly in our experimental materials, with qualitatively different profiles for syntax and pragmatics that are, on reasonable assumptions, in keeping with previous findings in the literature on eye movement patterns in response to anomalous sentences (for example, Ni, Crain & Shankweiler, in press; Pearlmutter, Garnsey, & Bock, 1995).

Eye movements provide generous quantities of on-line information and show very rapid response to sentence properties of interest. However, they are not ideal for the purpose of establishing *relative* timing relations between syntactic and pragmatic processes, for several reasons. As noted above, it is not easy to narrow down the intervals within which events occur. The division of effects between forward reading and regressions differs across anomaly types and may be susceptible to properties of the materials that are not of central interest. Because the reading is self-paced, the comprehension system can take as much time as it needs to complete all levels of processing, so minor differences in timing could go overlooked. There is some hint of a pragmatic processing lag in Experiment 1, but it is far from clear. To sharpen up the evidence we turn to compressed speech input.

Experiment 2: Cross-Modal Lexical Decision, Compressed Speech

A cross-modal dual-task paradigm (cf. Shapiro, Zurif, & Grimshaw, 1987) was used to establish the time course of anomaly detection under conditions of rapid processing. The sentences were produced by a male speaker at a normal rate (average 330 ms per word) and recorded using DigiDesign, Inc.'s Sound Designer II, an audio editing application for the Macintosh computer. Sound Designer II was then used to produce versions of these utterances compressed to approximately half of their original duration (54% on average, range 53% to 55%, the variability due to the fact that the program safeguards quality of the signal by refraining from compression at certain points). We checked the output of the program by making spectrograms of the stimuli both before and after compression, and observed no untoward changes. The timing was uniformly scaled, and the pitch contour and formant frequency contours were essentially unchanged. On the intelligibility of compressed speech, Altmann and Young (1993) reported: "We have found that there is virtually no loss of intelligibility, or subjective 'quality,' when sentences are compressed to, for instance, 50% of their original duration." Gerry Altmann (personal communication) has informed us that at 50% compression, intelligibility (after a brief period of adaptation) is around 85% to 90% for plausible sentences, where intelligibility is established by the percentage of words correctly reported by subjects after hearing the sentence. (The sentences were 8.5 words long on average, shorter than ours. The compression program differed from the one we used but, as far as we know, not in any consequential way.)

Sixty college students were paid for their participation in the experiment. All reported normal hearing. The compressed speech was played to subjects through headphones. The experimental sentences were as in (3). A minimum of 10 words followed the anomaly in all cases, to ensure that the concurrent task (lexical decision) did not overlap with sentence-final wrap-up effects. The first 14 sentences after the practice session were fillers, to provide subjects with the chance to adapt to the rapid speech. Data from three other subjects were excluded from the analysis due to poor performance on the comprehension task (described below).

A visual lexical decision target was presented at five different time points, both before and after the critical verb [*eat*, *eating*, *bake* in (3) above] was heard. Previous work (Ni, Fodor, Crain, Shankweiler & Mattingly, 1993) had shown the importance of tracing the rise and fall of sensitivity to the anomaly. A single test point is incapable of revealing timing differences between anomaly types if it happens that there is overlap between their time envelopes. The target word appeared at −81 ms, 0 ms, 81 ms, 162 ms, or 243 ms relative to the offset of the verb. (These test points were 150 ms apart before speech compression.) For filler sentences the target words appeared at a wide variety of sentence positions, randomly determined. The lexical decision targets for experimental sentences were all low-frequency words (mean 7.82 per million, range 0 to 36; Francis & Kučera, 1982) of medium length (mean 6.33 letters, range 3 to 10). Targets for the 76 filler sentences were 23 words, and 53 nonwords created by changing one letter in a word. Each target word appeared centrally on a computer screen until the subject responded, up to 700 ms. All target words were unrelated in meaning to the sentences during which they appeared. All three versions of an experimental sentence were associated with the same target word; each subject heard only one version. There was a complete rotation of sentence versions and test points: for each of the five test points, a subject heard six experimental sentences (two syntactically anomalous, two pragmatically anomalous, two baseline); thus each subject was presented with two tokens of each of the 15 presentation conditions. There was also an end-of-sentence task to ensure that subjects would take the trouble to listen to the sentences under these very demanding conditions; we will discuss the nature of this task below.

On the basis of Chodorow's (1979) findings, and our own impressions from listening to the compressed materials, we expected that the computation of content would fall behind the processing of structure, so that the lexical decision task would show interference from the syntactic anomaly at an earlier point than it would show interference from the pragmatic anomaly. As will become clear, this prediction was not confirmed.

The results are given in Fig. 3, where lexical decision time is shown for the anomalous sentence versions relative to lexical decision time for the baseline version. Note that Fig. 3 and all statistics associated with it derive from an analysis not of absolute RTs but of *z*-scores (= RTs expressed in terms of standard deviations from each subject's mean RT across all items). This is because the data were pooled from subjects divided into two groups whose mean RT differed as discussed in detail below. (Mean RTs in milliseconds for all three sentence versions are given by group in footnote 10; Synt Anom = syntactic anomaly; Prag Anom = pragmatic anomaly). In all analyses, only RTs for correct lexical decision responses were included, and lexical decision RTs were trimmed by changing those that were more than 2 standard deviations above or below the subject's mean RT to exactly 2 standard deviations above or below the mean, respectively.

Analysis of variance showed that the only position at which the anomalous sentences differed from the baseline was at the 0-ms test point. For the syntactic anomaly the mean difference in RT at this point was 56 ms [$F_1(1, 59) = 11.46$, $p = .0013$; $F_2(1, 29) = 5.54$, $p = .0256$]. For the pragmatic anomaly the mean difference in RT at this point was 52 ms [$F_1(1, 59) = 8.07$, $p = .0062$; $F_2(1, 29) = 7.610$, $p = .0100$]. There was no difference between the two anomaly versions at any test point. There was an apparent rise in RT for the syntactic anomaly version at the 243-ms test point but it was not significant relative to either the baseline ($p > .1$) or the

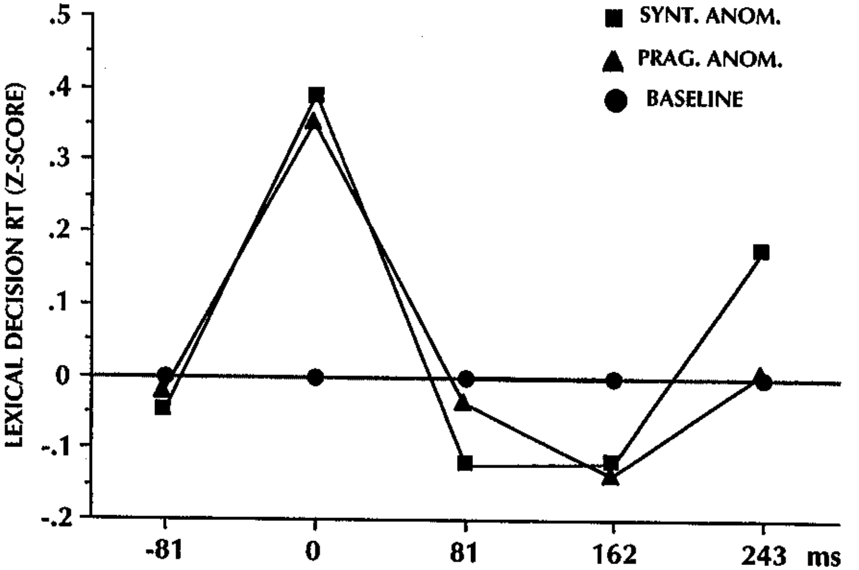

Fig. 3. Experiment 2: Lexical decision response times (RTs) (z-score) for anomalous sentence versions relative to baseline (all subjects). SYNT. ANOM. = syntactic anomaly; PRAG. ANOM. = pragmatic anomaly.

pragmatic anomaly ($p > .1$). Thus the result seems very clear. The peaks in lexical decision RT indicate the increase in sentence processing load due to the anomaly. And the peaks for syntactic and pragmatic processing lie on top of each other. There is no sign here of any delay in pragmatic processing relative to syntactic processing, at least within the grain provided by the 81-ms intervals between the test points.

The data in Fig. 3 are from all 60 subjects. These numbers combine results from subgroups of subjects who performed different postsentential comprehension tasks. Group A did an oral paraphrase task: After 20% of the filler sentences a bell sounded and the word *PARAPHRASE* appeared on the screen; the subject then had to speak into a microphone, giving the meaning of the preceding sentence in his or her own words. For Group B subjects, following the same sentences, a bell sounded and a simple comprehension question appeared on the screen; the subject answered it with the same *yes* and *no* keys as for the lexical decision task. Performance on these ancillary tasks was recorded but not analyzed except for purposes of screening out inattentive subjects. However, when the lexical decision RT data were analyzed for the two groups separately, it appeared that the comprehension tasks made an interesting (unanticipated) difference to lexical

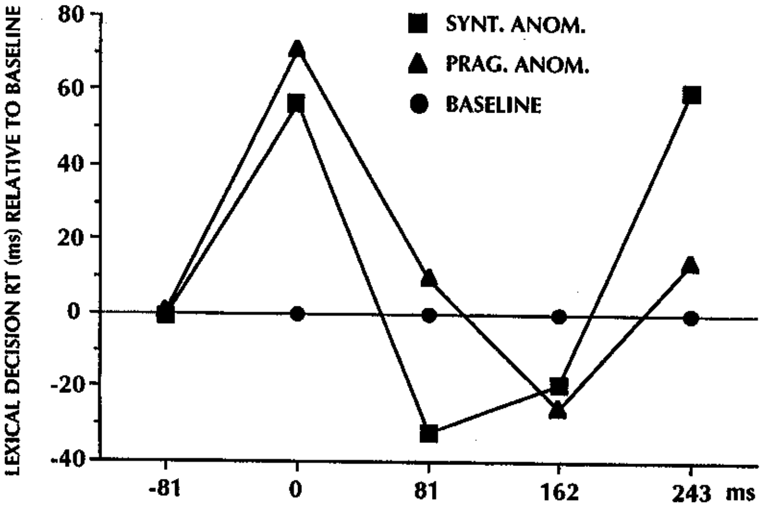

Fig. 4. Experiment 2: Lexical decision response times (RTs) (ms) for anomalous sentence versions relative to baseline: Group A (paraphrase task). SYNT. ANOM. = syntactic anomaly; PRAG. ANOM. = pragmatic anomaly.

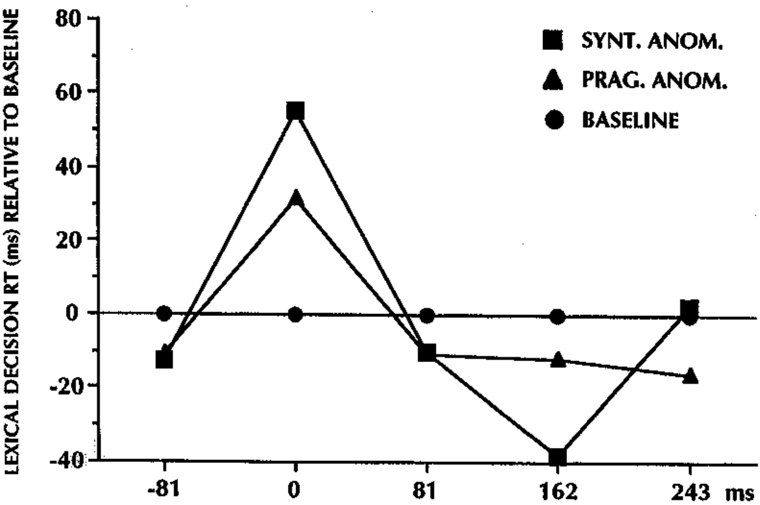

Fig. 5. Experiment 2: Lexical decision response times (RTs) (ms) for anomalous sentence versions relative to baseline: Group B (comprehension question). SYNT. ANOM. = syntactic anomaly; PRAG. ANOM. = pragmatic anomaly.

decision performance. It should be noted that this difference between groups is not statistically significant, but it exhibits a trend which is of sufficient theoretical interest to be worth discussing, even though conclusions must necessarily be tentative. Figures 4 and 5 show the results for the two groups separately.[10] The timing of the peaks for syntax and pragmatics is still identical, in each group. But the height of the peaks differs across groups, indicating a difference in the degree of sensitivity to the two types of anomaly. For the paraphrase subjects (Group A), sensitivity to the pragmatic anomaly was greater than sensitivity to the syntactic anomaly (Fig. 4). For the comprehension question subjects (Group B), the opposite was the case (Fig. 5). (Neither of these differences was significant; $p > .1$ in both cases.) In fact for Group B the pragmatic peak was not reliably a peak at all; the difference from the baseline dropped to 32 ms and was not significant ($p > .1$). The other three peaks remain significantly different from baseline. [For Group A, syntax vs. baseline, $F_1(1, 29) = 4.61, p = .0402$; $F_2(1, 29) = 4.64, p = .0397$; pragmatics vs. baseline, $F_1(1, 29) = 4.83, p = .0361$; $F_2(1, 29) = 6.16, p = .0191$. For Group B, syntax vs. baseline, $F_1(1, 29) = 4.82, p = .0363$; $F_2(1, 29) = 3.88, p = .0584$.]

What is of interest is the difference across groups in the sensitivity to the pragmatic anomaly. For pragmatics, the mean RT difference from baseline for Group A was 71 ms, but for Group B it was 32 ms. The difference between these (i.e., the interaction between subject groups and sentence type, for pragmatic anomaly vs. baseline) approached significance. This contrasted with sensitivity to the syntactic anomaly, which was almost identical across groups. For syntax, the mean RT difference from baseline for Group A was 57 ms; for Group B it was 55 ms. Plausibly, this apparent difference in pragmatic sensitivity between the subject groups was no accident. The paraphrase task was more demanding than the question answering task. It required more careful attention to the content of the sentence. This is reflected in the fact that for the paraphrase group the overall mean RT for lexical decision was 61 ms higher than for the *yes/no* question group. [Mean RT for Group A was 792 ms; for Group B it was 731 ms; $F_2(1, 29) = 52.30$; $p = .0001$.] Sensitivity to the pragmatic anomaly rose and fell as the task demands did. The syntactic effect, by contrast, stayed constant in magnitude

[10] Mean lexical decision RT (ms) for each sentence version and test point, by group, are as follows:

	Group A					Group B				
Test point	−81	0	81	162	243	−81	0	81	162	243
Synt anom	794	803	770	802	810	726	745	730	717	739
Prag anom	795	817	812	796	763	728	722	729	743	720
Baseline	794	746	802	822	749	738	690	740	755	736

across tasks. This suggests that when processing resources are limited, as in the compressed speech situation, the processor concentrates efforts on syntax. It seems that syntactic processing is mandatory, but pragmatic processing is not; it may be sacrificed when time is short. This is one interesting sense in which it seems that syntax does take priority over pragmatic processing, and it is worth noting that mandatoriness is another of the properties that J. A. Fodor (1983) proposed as indicative of a module at work. However, this is the *only* evidence we found in this experiment for the priority of syntax over pragmatics. The temporal delay of pragmatics that might have been expected with speeded speech was not evident at all.

To sum up: It can be concluded that differences in time of availability of syntactic and pragmatic information are *not* the source of timing differences in their use for ambiguity resolution, at least for materials such as these where detecting the pragmatic anomaly requires no complex inference. Of course, the data reported here do not exclude the possibility that pragmatics runs some milliseconds behind syntax, as might be expected on any model in which the pragmatic analysis is fed by the syntax. (Pragmatic analysis not fed by the syntax occurs in "strong interaction" models; see McClelland, St. John, & Taraban, 1989, and Bates & MacWhinney, 1989.) Our experimental results do not discriminate between a very short delay and none at all. But they do appear to exclude a delay on the scale of those standardly cited in favor of syntactic autonomy in ambiguity resolution. Such effects are not always precisely timed but seem to be on the order of several hundred milliseconds. The classic paper by Rayner, Carlson, and Frazier (1983; p. 371) concluded that "probabilistic semantic and pragmatic information does not influence the processor's initial choice of a syntactic analysis . . . [but] semantic and pragmatic plausibility information does influence the ultimately preferred analysis of a sentence"; however, the data as presented do not permit computation of specific temporal relations between the initial and the ultimate analyses. Mitchell et al. (1992) looked for an influence of discourse on ambiguity resolution at two points in their material: at the beginning and end of the ambiguously attached *that*-clause (see Example 1 above). As noted in the first section, they found no discourse influence at the earlier disambiguation point immediately after *that*, but they did find a marginal effect of discourse at the later point (e.g., at the second *that*-clause in *The politician told the woman that he had been meeting that he was going to see the minister*, which disambiguates the first *that*-clause as a relative). The data as presented do not afford an exact calculation of when this second disambiguation occurred, but it appears to have been at least 1200 ms later than the early test point. It is not possible to tell whether discourse information had been used for ambiguity resolution any earlier than that. A recent study by Urbach, Pickering, Branigan, and Myler (1995) was designed to provide more precise timing information. Urbach et al. tested syntactic and pragmatic cues for disambiguation of the familiar main clause/reduced relative clause ambiguity (e.g., *The cook helped/helping was busy; The teacher/language taught was Spanish*). ERP patterns indicated that both cues were helpful in staving off a garden path at the word *was* when sentences were presented visually one word at a time at 550 ms per word, but only the syntactic anomaly averted the garden path when the presentation rate was 400 ms per word. This could suggest a pragmatic lag in ambiguity resolution of somewhere between zero and 550 ms. However, some caution is necessary here since no data are presented for the "anomalous" word (e.g., *taught* following *language*), and without this it is not possible to estimate whether the pragmatic anomaly had been detected prior to *was* at the faster presentation rate, i.e., whether these materials meet the availability criterion. We return to the relationship between ambiguity resolution and unambiguous processing in our general discussion of the findings.

With appropriate provisos, we conclude from Experiment 2 that even at a twice-normal input rate, pragmatic computations can keep pace with syntactic computations. Perhaps we should have guessed that this might be so, on the basis of results reported by Young, Altmann, Cutler, and Norris (1993). They were looking for prosodic effects on intelligibility of compressed speech, but they concluded: "Overall, the only consistent predictor of intelligibility across the two experiments was plausibility. The more plausible a sentence, the better recognized are the component words." Obviously, plausibility could not have influenced word recognition unless plausibility were being calculated rapidly. However, the plausibility effect was not necessarily occurring in the word recognition stage; it could have been due instead to response bias, since the task was to write down the sentence after hearing it. Young et al. noted:

> . . . the present results do not allow us to decide whether this is because a word can be better predicted, and hence better recognized, on the basis of the preceding words as the sentence is heard, or because during the subsequent transcription of the sentence it is easier to reconstruct words which had not been recognized originally. The latter explanation, invoking reporting biases based on listeners' experience, is certainly consistent with the finding that the effect of plausibility remains constant across compression rates.

It also, of course, does not imply rapid pragmatic processing. However, the results of Experiment 2 make it more likely that Young et al.'s plausibility effect was not (just) a response bias.

The outcome of Experiment 2 does not conform to our expectation, based on Chodorow's (1979) findings, that syntactic and pragmatic processing would be "pulled apart" by time compression of the input. But here too, hindsight offers explanations. Our paradigm differed substantially from Chodorow's. Chodorow estimated processing load postsententially. With compressed speech it seems especially likely that the processor would want to resurvey the whole sentence during "wrap-up" operations; this could create a postsentential lag at some level of processing, even if none occurred on-line. Thus there is no contradiction between Chodorow's results and our own. Note also that Chodorow compared the extent of the postsentential spillover of processing for compressed speech with that for normal speed speech, and found that (for difficult sentences) the former was greater than the latter. Our data do not speak to this. They show that pragmatics was not *differentially* slowed by the speech compression, but they do not exclude the possibility that all levels of processing were retarded (by an equal amount).

It is important that other varieties of pragmatic anomaly be tested to establish whether it is generally the case that pragmatic processing is as rapid as syntactic processing, regardless of how intricate the pragmatic reasoning involved. The anomalies we tested resemble the sentence-internal subject–predicate incompatibilities investigated in many studies, including Ferreira and Clifton's (1986) experiment described above. But other mismatches, such as between a sentence and the prior discourse, might be established more slowly. What complicates matters is that it is often unclear what initiates the relevant computations. For instance, in Example 1 above from Mitchell et al.'s (1992) experiment, the singular noun *woman*, if it had no modifying phrase, would be referentially anomalous following a discourse which established two women as potential referents. As Mitchell et al. noted, the existence of two women was established in the context several words before the noun *woman* in the test sentence. However, there is no guarantee that the processor would make use of this time to deduce that *if* an unmodified singular noun *woman* were to occur it would be infelicitous. Perhaps such an inference is drawn only when it becomes relevant to some later decision. In that case the computation would not start until the word *woman* was received. Even this presupposes an active processor that anticipates what may follow the current input. (To what extent the human parser is anticipatory has never been very clearly established, but see Gorrell, 1989, for discussion.) A merely passive parser would process structure and meaning in step with the input and would not react to problems until they arose. In that case, the discourse information about two women would not initiate relevant inferencing until the following word, *that*, was received—but this is the very word whose attachment needs to be disambiguated.

The only safe course would be to test all materials designed to be used for experiments on modularity. The question then arises of what experimental paradigms are suitable for this purpose. Working with compressed speech cannot be commended on grounds of convenience. It is considerably more demanding in terms of both labor and technical resources than the use of normal spoken or visual stimuli. Moreover, a dual-task paradigm (at least, if the secondary task provides just a single RT, not a continuous response measure) is not economical as a way of monitoring changes in processing load over an extended stretch of a sentence. It is of practical importance, therefore, to establish whether an adequate test of information availability could be based instead on a direct response to a sentential event such as an anomaly, and at normal presentation rates. Arguably, the latter is insufficient because of the importance of catching small timing differences in availability that could have a powerful effect in a winner-takes-all system for ambiguity resolution (see the general discussion below). Unless more precise time-sensitive response measures are used than in most current experiments, there may be no way to achieve this other than by forcing the pace of the processing routines, as in Experiment 2 or in the manner of McElree and Griffith (1995), so that small timing differences cannot be absorbed and become measurable. It is possible, though, that the measurement could be made by means of other on-line tasks than the cross-modal lexical decision task of Experiment 2. To check this, we investigated the value of a simple anomaly monitoring task. (See also Osterhout et al., 1994, for an ERP study of sentences like those in the experiments reported here.)

Experiment 3: Anomaly Monitoring, Compressed Speech

The same compressed speech materials as in Experiment 2 were presented to nine subjects in a monitoring task. They were instructed to "push the NO button as fast as you can, if you hear a mistake." A practice session gave examples of "mistakes," both syntactic and pragmatic. All sentences continued to completion whether the button was pressed or not. Table I shows RT and percentage of sentences judged anomalous for each sentence version. The pragmatic anomalies were detected almost as often as the syntactic ones; the difference in accuracy was not significant ($p > .1$). We may conclude that subjects' judgments were roughly equally secure for both

Table I. Experiment 3: Sentences Judged Anomalous

	Percent	RT[a] (msec)
Syntactic anomaly	83.3	1,011
Pragmatic anomaly	75.3	1,535
Baseline	9.9	2,186

[a] RT = response time.

kinds of example. However, subjects' reactions to the pragmatic anomalies were very much slower than to the syntactic anomalies; the difference of 524 ms is highly significant [$F_1(1, 8) = 13.33$, $p = .0001$; $F_2(1, 29) = 12.62$, $p = .0001$]. However, in view of the striking lack of difference that we found in Experiment 2, where an explicit judgment on the sentences was not called for, we believe that this delay of response to the pragmatic anomaly when the subjects' task was monitoring for anomalies must be attributed to judgment processes rather than to architectural constraints on information flow between processing units. Indeed, the monitoring task would be expected to show the same sort of strategic effect as we saw for eye movements in Experiment 1, perhaps in heightened form because the button push in Experiment 3 depended on a more deliberate decision than the control of eye movements. The sentence processing routines, not in doubt about when a syntactic error has ruined a sentence, would respond promptly in the judgment experiment. But in the case of a pragmatic anomaly, the processor might anticipate that the apparent problem would fade away, or that it would be rescued somehow or other by later parts of the sentence; what at first seemed absurd might turn out to make sense when the message was eventually grasped as a whole. So the processor would wait for further words, to see if things improved. If so, then the button-pressing *response* to the pragmatic anomaly would be delayed even if there were no delay in pragmatic processing at all.

Other research shows that when the processor is not allowed to wait for more words before making a pragmatic anomaly decision, processing is slowed overall. Tanenhaus, Garnsey, and Boland (1990) noted that in the "stop making sense" paradigm, self-paced reading time increased by about 200 ms per word (for "makes sense" judgments) relative to normal self-paced reading. Response times for negative ("stop making sense") responses were not reported, but they were presumably at least as slow. Therefore, this paradigm, like the anomaly monitoring paradigm of Experiment 3, is not optimal for purposes of establishing precise timing of events in normal sentence processing. As noted earlier, an explicit judgment methodology for studying anomaly detection has also been employed by McElree and Griffith (1995), who observed a significant delay in pragmatic processing relative to syntactic processing. They noted (p. 152): "For the active conditions, the average estimate of when thematic role processing began was 279 ms as compared with 233 ms for syntactic processing. In the passive conditions, the estimate was 289 ms as compared with 172 ms." McElree and Griffith's experiment differed in several ways from both Experiment 2 and Experiment 3. Their sentences were only 4 to 6 words long, and the anomalous word was always the final one. Therefore subjects would know that the anomaly could not be reprieved by anything later in the sentence.

Also, the judgments would be made in the region of "wrap-up" processing, which may differ from on-line processing. Sentences were presented visually at 200 ms per word (somewhat slower than our compressed speech materials). The test points (at which a bell signaled subjects to respond, whether ready or not) were at 14, 157, 300, 557, 800, 1,500, and 3,000 ms after the onset of the final word. Thus, the early test points were 143 ms apart, and this is where sensitivity to the anomalies was estimated to have begun. The estimates of when, within one such interval, the different types of anomaly first became discriminable were made on the basis of curve fitting that included the later test points as well.

Since the goals of this project and our own were similar it is of interest to consider their findings together, even though comparison across such different paradigms can only be suggestive. McElree and Griffith's (1995) thematic role violations were similar to our pragmatic anomalies (e.g., *Some people alarm books*), and their category and subcategorization violations were not unlike our syntactic anomalies (e.g., *Some people rarely books, Some people agree books*). Tentatively, therefore, the small but significant estimated difference in onset of sensitivity to the thematic versus (sub)category anomalies in McElree and Griffith's experiment may be compared with the lack of an observed pragmatic/syntactic difference in our Experiment 2 and the large difference in our Experiment 3. One way to reconcile these findings is to assume, as suggested by Brian McElree (personal communication), that Experiment 2 would also have shown a pragmatic lag if the intervals between test points had been shorter. As noted earlier, this is certainly possible. It is clear that, in future experiments, the test points should be even closer together; we know now that if there is a lag it is very brief. However, we believe that it is at least as likely that the pragmatic lag that McElree and Griffith reported was induced by the anomaly judgment task, under the stress of the speed–accuracy trade-off test procedure; that is, that it represents a delay in the *judgment*, but not necessarily a delay in *detection* of the anomaly by the processing routines.

The reason for using anomalous sentences for purposes of comparing syntactic and pragmatic processing is that anomaly is a property that both levels of processing can exhibit regardless of differences in other respects. But even so, there are possible disparities between the two types of anomaly that must be guarded against. Syntactic errors are often more sharply defined than pragmatic ones, and may be judged more confidently. This is a difference that is not easy to eliminate, so it is important to protect experimental designs against its influence, as far as is possible. However, requiring subjects to make an overt anomaly judgment response may magnify this difference between anomaly types. The subject is faced with what amounts to a double judgment: (i) Do I, the subject, find this sentence odd? (ii) Is its oddity such that the experimenter classifies it as a mistake? One judgment is about the sentence; the other is about the likelihood of interspeaker agreement about the sentence. For the materials in our experiments and in McElree and Griffith's (1995), subjects were probably more certain that the experimenter agreed with them on syntactic judgments than on pragmatic judgments. We have no data on this matter, but it would not be difficult to test experimentally. If it is correct, it would explain why subjects might need a little more time before overtly classifying a sentence as containing an error if the error is pragmatic than if it is syntactic. The fact that the final judgment is the same in both cases (as the identical asymptote implies) does not mean that there was not more doubt along the way in one case than in the other.

The issues are complex, but one outcome that is clear is that the subjects' task makes a difference. Experiments 2 and 3 gave very different results for identical sentence materials. It seems clear that some experimental paradigms are less suitable than others for establishing timing relations among processing events, because the measurement task may itself introduce causes for differential delays in response. It is crucial to know what is controlling response time, to avoid misinterpreting any delayed response as evidence of delayed sensitivity to the input. The data reported here suggest that when explicit recognition of an anomaly is called for, as in the monitoring task of Experiment 3, the judgment of pragmatics is quite tardy, though there is no evidence of pragmatic delay when subjects listen to sentences for comprehension only, as in Experiment 2. Reading for comprehension, as in Experiment 1, falls somewhere between these poles. There is no overt judgment to be made and correspondingly no delay due to deliberation. But there are "decisions" to be made by a lower-level mechanism that controls the scanning of the text, and it does not treat pragmatic and syntactic problems alike. The methodological conclusion to be drawn would seem to be that, for purposes of investigating autonomy of linguistic processing, measurement should be as indirect as possible. There must be some means of establishing when an anomaly has been detected, but this should not require subjects to indicate their evaluation of the linguistic status of the sentences. Dual-task paradigms are well suited to this purpose. And they have the additional advantage of allowing the sentential input to be paced by the experimenter, so that the perceiver cannot slow processing down to a pace at which all levels of processing accommodated and no differences between them could emerge.

GENERAL DISCUSSION: AVAILABILITY AND MODULARITY

The data reported here suggest that there is virtually simultaneous (unconscious) processing of syntax and pragmatics, even under tough conditions where pragmatics might have been expected to fall behind. This is the most reasonable interpretation of the results of Experiment 2, and Experiments 1 and 3 do not in any way contradict it. Though perhaps unexpected, this is a welcome outcome for both sides of the modularity debate. Our results support the rapid availability of pragmatic facts. And this, as we will argue, is a precondition for using timing data to establish *either* the autonomy *or* the nonautonomy of syntactic processing.

For interactionists, the present findings have at least the virtue that they do not contradict empirical claims of immediate effects on parsing of all relevant forms of information. They validate the assumption that human language processing could be interactive not just architecturally but de facto as well: Pragmatic information is accessible in time for it to have an influence on ambiguity resolution, if the structure of the system permits it to do so. (Whether or not it does permit it to do so—and what counts as pragmatic in the relevant sense—is the issue that has drawn most attention in recent years, but it is not addressed here.) Of course, we have availability data so far only on simple subject–predicate mismatches like *the cats wouldn't bake the food*, and as noted above, more complex pragmatic inferences might become available at later times. We cannot now estimate at what point greater complexity of inference, or lack of sharp contrasts, gives rise to detectable delays.[11] But at least the current results imply that materials demanding fairly simple pragmatic processing can be used in experiments designed to reveal the interaction of all information sources on an equal footing, without danger that an availability delay will inadvertently mask the legitimacy of putting pragmatic information to work whenever and wherever it is useful.

The immediate availability of pragmatic facts is advantageous for the modularity hypothesis also. It protects it against objections of mere de facto autonomy compatible with an underlying interactive architecture. It implies that, whenever a delay of pragmatic influence on ambiguity resolution is successfully demonstrated (with a suitable task), it is most likely due to structural restrictions on how subprocessors are permitted to commune with one another. This saves proponents of modularity from having to demonstrate an even longer delay (a delay greater than the delay in availability of

[11] See Swinney and Osterhout (1989) on the speed of various types of inference involved in language comprehension. They distinguished "perceptual" inferences (rapid and mandatory, e.g., essentially immediate determination of antecedents for pronouns) from "conceptual" inferences (slow and "nonautomatic," e.g., metaphorical reasoning, presupposition, implicit instruments). Altmann (1988) presented data indicating a delay of approximately 300 ms due to the complexity of an inference relating a relative clause to its context.

pragmatic information) to prove the same point. The possibility that pragmatic processing is slow for materials where the relevant inferences are unclear or complex must of course be borne in mind, as noted above. But at least it appears that simple pragmatic anomalies can confidently be used to look for signs of architectural restrictions on parsing.

Exactly how these restrictions operate needs to be explicated in a modular model. If there is essentially no pragmatic lag in normal processing of unambiguous strings, why and how should a detectable lag occur in the processing of ambiguous strings? We may distinguish three answers to this question. (For purposes of this discussion let us suppose that a temporal advantage of syntax over pragmatics in ambiguity resolution *has* been empirically established.)

(i) *Time delay.* It might be that some fixed number of milliseconds is allotted to syntactic resolution strategies before pragmatics is consulted.
(ii) *Domain delay.* The lag might be defined by some significant linguistic domain such as a clause or a theta domain.
(iii) *Priority delay.* There might be a priority ranking which requires nonsyntactic factors to wait until syntax has done all that it is capable of doing with the current input item, however much or little that might be.

A priori, (ii) and (iii) seem more credible than alternative (i). A pure time delay just in the case of ambiguity resolution cannot be ruled out, but it has no obvious rationale. Note that it would have to be something other than a simple blockage or detour in the pipeline carrying input from the syntactic to the pragmatic processor since that would delay the pragmatic processor equally in ambiguity resolution and in normal unambiguous processing. But even if (i) is implausible, either (ii) or (iii) could provide the modularity hypothesis with an adequate account of why a pragmatic lag shows up specifically in ambiguity processing.

Arguments for a domain delay have been given by Frazier and colleagues (e.g., Frazier, 1990; Rayner et al., 1983). The proposal is that syntactic and thematic/pragmatic analyses are conducted in parallel and that the two units compare notes at the end of each thematic domain. The thematic analysis therefore cannot be based on a full syntactic analysis but it could be fed by a crude recognition of lexical categories such as noun and verb. The thematic processor determines their optimal relations on the basis of plausibility, and reports its views to the syntactic processor. If they agree, the syntactic analysis proceeds without alteration; this would be the normal case for unambiguous nonanomalous sentences. For ambiguous or anomalous sentences the two processors may disagree, in which case the syntactic processor then looks for a different analysis that does not conflict with pragmatics. The delay before this reanalysis begins would be variable; it would be determined by the distance between the ambiguity and the end of the relevant thematic domain in the particular sentence under analysis. Note that no delay is predicted in the *registering* of a pragmatic anomaly (with impact on a concurrent task as in Experiment 2), but only in when it is *acted on* in establishing the structure for the sentence. This point was not explicitly discussed by Rayner, et al. and Frazier, and it may seem incompatible with their assumption that the thematic processor ignores syntactic constraints and arranges arguments and predicates in whatever way best suits itself. A thematic processor of this kind would not immediately spot a pragmatic anomaly consisting of a reversal of a plausible predicate–argument structure, such as *The ice cream ate the children from the orphanage*; this would be detected only when the syntactic and thematic processors confer at the end of the clause. The model could be adjusted to predict anomaly detection in this case [e.g., by assuming the thematic processor employs heuristics such as a Noun-Verb-Noun strategy], but there is no need to do so to account for the present results, since the pragmatic anomalies in our materials were not curable by reversing them. Both a cat baking food and food baking a cat would offend the thematic processor. (Only one out of the 30 examples would be improved by reversing its arguments: *The pacifier we bought in Japan will drop the cranky baby. . . .*)

Arguments for a priority delay model are given by Meltzer (1995) based on comparison of the processing of the empty categories PRO and *pro* of government binding theory (Chomsky, 1981) in Spanish. Meltzer's results suggest that pragmatic selection of an antecedent for a dependent element may be either immediate or delayed depending on how much the grammar has to say about the choice of the antecedent. If syntactic principles do not determine the antecedent (as in the case of *pro*), pragmatic selection of the antecedent may begin immediately. But pragmatic selection must hold back if syntactic principles are relevant, e.g., if syntax entails that an obligatory antecedent will occur later in the sentence, as is the case for a controlled PRO in a clause that has been fronted. More generally, pragmatic delays would be expected to vary in length depending on whether or not the input word enters into a syntactic dependency with material to its right, and if so, how distant that material is. Such a dependency extends the domain in which syntax has information to contribute, and thereby postpones the point at which pragmatic guessing may begin. For model (iii), there would be essentially no pragmatic delay in cases where local information is fully sufficient to establish the correct analysis. Syntax tells pragmatics: *Wait until I'm through*; but where there is no local ambiguity, and no rightward dependency, syntax could be finished so quickly that there is no measurable waiting period for pragmatics at all. Why would there be delay in the case of ambiguity? It would result from the fact that syntax will have made its choice between the alternative analyses before pragmatics is given a chance to vote. When pragmatics enters on the scene, there is no decision left to be made: its only option is to accept the decision that syntax made, or else to override it. And overriding a decision once it is in place presumably takes more time and effort than initial decision making does. Therefore (depending on how difficult the revision process is[12]), effects of pragmatics on ambiguity resolution would be observed quite late.

This priority-delay variant of the modularity hypothesis raises a new methodological challenge. It reconciles the absence of a detectable pragmatic delay in normal processing with the possibility of a significant delay in ambiguity resolution, by the simple assumption that once syntax has started its work on an input item, it will not stop until it has done all that it can. As a result, even a very brief headstart for syntax can be magnified into a much greater one because of the difference between making a decision and overthrowing a decision. But now, this same kind of winner-takes-all priority system could be combined with a *non*modular design in which *any* processor can compete for priority and the winner is determined by a race. Whichever gets there first with useful information to contribute gains the right to proceed and is permitted to complete its work before the other begins. The observable manifestations of such a system would resemble those of modular model (iii) if, de facto, syntax is usually the first to contribute relevant information—first by any tiny interval, and for whatever practical reason. In that case syntax would usually have the headstart and pragmatics could at best struggle against it and so would be delayed, though not as a matter of principle, and not because of fixed architectural barriers restricting information flow.

Once again, then, we see that the deductive link between overt timing facts and underlying design characteristics is very fragile. Unless this case can be excluded somehow on other grounds, it raises the stakes considerably on the investigation of timing relations in unambiguous sentence processing. It would no longer be sufficient, as a precondition on being able to dem-

[12] Fodor and Inoue (1994) have proposed, on the basis of informal judgments of processing difficulty in Japanese and English, that when the "symptom" of a garden path is a pragmatic anomaly (as in *They told the boy that the girl met the story*) the parser is less successful at finding the correct analysis than when the symptom is a syntactic anomaly (as in *They told the boy that the girl met not to go home*). If true, this suggests that pragmatic information cannot easily overthrow an established syntactic structure; only a syntactic anomaly has the power to face down a prior syntactic decision. This does not deny that pragmatic cues may be helpful in shaping the direction of a structural reanalysis (cf. Carlson & Tanenhaus, 1988) but it implies that they will be more effective if the need for reanalysis has been signaled syntactically.

onstrate modularity, to establish the availability of pragmatics prior to the point at which pragmatics could have been useful but was not used. It would be necessary now to entirely eliminate the possibility of any systematic tendency for syntactic facts to become accessible slightly before pragmatic ones. This is hardly feasible with current methods. Yet anything less than simultaneity of availability could lend itself, as we have seen, to a winner-takes-all priority model compatible with either a modular or a nonmodular design for the language processing system.

As observed above, this kind of indeterminacy weakens both sides of the debate. Interactionism might be true, and yet be obscured by delays overlaid on it by a headstart priority system. Syntactic autonomy might be true, and yet not demonstrable because its manifestations could always be nonautonomously accounted for. This is unfortunate. It points up the perilousness of the project that so much of psycholinguistics has occupied itself with in recent years—the project of deducing underlying mental organization from its one-dimensional projection onto chronometric relations between observable operations. The goal is to infer mental *structure* from facts about the timing of mental *processes*. This is always chancy, but it is particularly tricky when it is modal notions that are under test. We have to distinguish "these processes cannot interact" from "these processes could interact but do not." Unless we are prepared to give up on the quest for architectural conclusions, other sources of evidence may need to be found.

APPENDIX: MATERIALS

Notes: For Experiment 1, the sentences as illustrated in (3) above were shortened by deletion of words at the beginning and end, but included the adverbs shown here in curly brackets. For Experiment 2, the words shown here in parentheses following the sentences were used as lexical decision targets. The alternative verb forms for all three experiments (syntactically anomalous, pragmatically anomalous, baseline) are given here in square brackets. Filler sentences are not included here.

1. It seems that the <u>cats</u> from across the road won't {usually} <u>[eat/eating/bake]</u> the food that Mary puts out on the porch every morning as soon as she gets up. (RINK)
2. Apparently the <u>argument</u> given by the astronomer might {even} <u>[prove/proving/shout]</u> that there are many canals on the moon though this has not been widely believed for over a decade. (MOTH)
3. In case of a break-in, the <u>alarm system</u> we just installed will {surely} <u>[warn/warning/swear]</u> that there is an intruder in the building and alert the local police department. (BISCUIT)

4. The new species of orchid that was discovered in Peru will {only} [grow/growing/sing] in tropical regions of South America or the islands around Madagascar. (HEXAGON)
5. This very expensive ointment from South East Asia will {supposedly} [cure/curing/loathe] all known forms of skin disease but only if it is used in accord with the instructions on the package. (METEOR)
6. This old electric blender that the bartender uses doesn't {really} [crush/crushing/own] icecubes any more but the management can't afford to spend money on a new one. (TRIBE)
7. This exotic spice from Aunt Ellen's kitchen may {possibly} [add/adding/seek] the subtle oriental flavor that John enjoys and is so difficult to find in this country. (HERMIT)
8. According to reports, the new fighter-plane that was tested in Nevada can {apparently} [fly/flying/walk] faster than anyone had expected it to when it was originally designed. (HUT)
9. The big wooden boxes in the attic may {still} [hold/holding/find] many old photographs and souvenirs from our trips abroad in the sixties when we hitchhiked all around Europe and India. (ELBOW)
10. Despite its other merits, this new test of mathematical reasoning might {occasionally} [fail/failing/hate] to discriminate between students of quite different abilities or aptitudes. (BLOUSE)
11. The inspector asserts that helicopters taking off from the roof may {repeatedly} [shake/shaking/paint] the walls and windows of the top floor and do considerable damage to the building. (TOOTH)
12. The plumber warned us that the leaking water he noticed yesterday might {slowly} [seep/seeping/speak] out from behind the refrigerator and ruin the linoleum tiles in the kitchen. (OUNCE)
13. They are confident that the fingerprints on the gun next to the body could {clearly} [prove/proving/judge] that the defendant is innocent though he had both motive and opportunity to commit the crime. (MATTRESS)
14. A family of beavers that lived in our duckpond would {sometimes} [chew/chewing/melt] the garden hose beside the shed so that we were unable to water the lawn. (SAUCER)
15. The fancy French clock that was selected by the mayor doesn't {always} [tell/telling/ask] the time during a power failure because the special batteries that it is designed to take are unavailable. (PEACH)
16. Critics say the latest rap songs that are played on MTV might {supposedly} [tend/tending/learn] to lead impressionable young people into immoral or indecent forms of behavior. (PETAL)

17. Those small red spiders with very long legs would {often} [spin/spinning/burn] beautiful webs in the rose bushes beneath the old maple trees near the barn. (SYNTHETIC)
18. I am sure that the pacifier we bought in Japan will {immediately} [soothe/soothing/drop] the cranky baby within a few minutes and then we will all be able to get some peaceful sleep at last. (SPLINTER)
19. People complained that the skyscraper being built by the city would {eventually} [block/blocking/seek] out the sunlight even in the middle of the day because it is positioned too close to the other buildings in the street. (BRIDESMAID)
20. Unfortunately, these grape vines from Southern France don't {usually} [grow/growing/jog] well in sandy regions where the topsoil is too loose to provide enough support for their long roots. (NICOTINE)
21. One elderly kangaroo in the San Diego Zoo would {just} [sit/sitting/swear] all day at the gate that the keeper usually enters through when he brings food and fresh water. (EXPLOSION)
22. The full-length portrait of great uncle Henry doesn't {really} [look/looking/talk] like him or like anybody else in the family but it is extremely handsome. (NATIONALISM)
23. Hopefully the new heater in the maid's room should {quickly} [dry/drying/kick] the laundry that she hangs over the towel rack when the weather is too bad to use the out-door clothesline. (EPISODE)
24. Don't you think that the strawberry beds being planted by the gardener might {soon} [tempt/tempting/lift] rabbits and other animals into the backyard and create a serious problem of pest-control for the future? (IDIOTS)
25. The exquisite colors woven into these sweaters shouldn't {ever} [fade/fading/cry] when they are washed in hot soapy water but I think it is always safer to send things to the dry cleaner. (GLOBE)
26. A chemical additive now being tested may {also} [tend/tending/want] to lower the freezing point of sea water so that ships can be kept ice-free all winter long. (SANCTUARY)
27. The sleek black sea lions that inhabit the little bay can {happily} [bask/basking/read] on the beach all day long when the weather is fine and can sleep on the rocky ledge at night. (GHOST)
28. We are pleased to report that the security camera at the bank will {now} [take/taking/tear] photographs of everyone who uses the automatic cash machine or the overnight deposit box. (MEDAL)
29. Sam is scared because the bull that escaped could {easily} [smart/smashing/mend] the wooden fence around the meadow and get into the field where the sheep are grazing. (MAST)

30. It is clear that the lever on the basement wall does not {reliably} [shut/shutting/leap] off the air-conditioning unit or the power supply to the elevators. (GEOGRAPHY)

REFERENCES

Altmann, G. T. M. (1988). Ambiguity, parsing strategies, and computational models. *Language and Cognitive Processes, 3*, 73–97.

Altmann, G. T. M., & Young, D. (1993). Factors affecting adaptation to time-compressed speech. *Proceedings of Eurospeech, 93.*

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

Boland, J. E. (1996). The relationship between syntactic and semantic processes in sentence comprehension. Manuscript submitted for publication.

Boland, J. E., Tanenhaus, M. K., & Garnsey, S. (1990). Evidence for the immediate use of verb control information in sentence processing. *Journal of Memory and Language, 29*, 413–432.

Boland, J. E., Tanenhaus, M. K., Garnsey, S., & Carlson, G. (1995). *Journal of Memory and Language, 34*, 774–806.

Carlson, G. N., & Tanenhaus, M. K. (1988). Thematic roles and language comprehension. In W. Wilkins (Ed.), *Syntax and semantics, Vol. 1: Thematic Relations*. London: Academic Press.

Chodorow, M. S. (1979). Time-compressed speech and the study of lexical and syntactic processing. In: W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett.* Hillsdale, NJ: Erlbaum.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht, The Netherlands: Foris.

Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological parser. In D. R. Dowty, L. Karttunen, & A. M. Zwicky (Eds.) , *Natural language parsing: Psychological, computational, and theoretical perspectives.* Cambridge, England: Cambridge University Press.

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

Fodor, J. A. (1983). *Modularity of mind.* Cambridge, MA: MIT Press.

Fodor, J. D., & Inoue, A. (1994). The diagnosis and cure of garden paths. *Journal of Psycholinguistic Research, 23*, 407–434.

Francis, W. N., & Kučera, H. (1982). *Frequency analysis of English usage.* Boston: Houghton Mifflin Company.

Frazier, L. (1990). Exploring the architecture of the language-processing system. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives.* Cambridge, MA: MIT Press.

Friederici, A. D., Mecklinger, A., Steinhauer, K., & Hahne, A. (1995, March). *Processing violations of syntactic structure versus violations of syntactic preference: Evidence from ERP studies.* Paper presented at the 8th CUNY Conference on Human Sentence Processing, Tucson.

Gorrell, P. (1989). Establishing the loci of serial and parallel effects in sentence processing. *Journal of Psycholinguistic Research, 18*, 61–73.

MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes, 32*, 692–715.

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

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

Meltzer, S. G. (1995). *Issues in the processing of null subjects*. Unpublished doctoral dissertation, City University of New York.

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

Ni, W., Crain, S., & Shankweiler, D. (in press). Sidestepping garden paths. Assessing the contribution of syntax, semantics and plausibility in resolving ambiguities. *Language and Cognitive Processes.*

Ni, W., Fodor, J. D., Crain, S., & Shankweiler, D. (1996). Anomaly detection: Eye movement patterns. Haskins Laboratories. Manuscript submitted for publication.

Ni, W., Fodor, J. D., Crain, S., Shankweiler, D., & Mattingly, I. G. (1993, March). *Evidence of the autonomy of syntax and pragmatics*. Poster presented at the 6th Annual CUNY Sentence Processing Conference, University of Massachusetts, Amherst.

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

Osterhout, L., Nicol, J., McKinnon, R., Ni, W., Fodor, J. D., & Crain, S. (1994, March). *An event-related brain potential investigation of the temporal course of sentence comprehension*. Poster presented at the 7th Annual CUNY Sentence Processing Conference, City University of New York.

Pearlmutter, N. J., Garnsey, S. M., & Bock, K. J. (1995, March). *Subject-verb agreement processes in sentence comprehension*. Paper presented at the 8th CUNY Conference on Human Sentence Processing, Tucson.

Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction between 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.

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

Swinney, D., & Osterhout, L. (1989). Inference generation during auditory language comprehension. In A. Graesser & G. H. Bower (Eds.), *The psychology of learning and motivation*. Academic Press.

Tanenhaus, M. K., Garnsey, S. M., & Boland, J. E. (1990). Combinatory lexical information and language comprehension. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives*. Cambridge, MA: MIT Press.

Tanenhaus, M. K., & Trueswell, J. C. (1994). Sentence comprehension. In J. Miller & E. Eimas (Eds.), *Handbook of Perception and Cognition*: Volume II: Speech, Language, and Communication.

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

Urbach, T., Pickering, M., Branigan, H., & Myler, L. A. (1995, March). *Event-related potential effects and parsing*. Poster presented at the 8th CUNY Conference on Human Sentence Processing, Tucson.

Young, D., Altmann, G. T. M., Cutler, A., & Norris, D. (1993). Metrical structure and the perception of time-compressed speech. *Proceedings of Eurospeech, 93*.