Journal of Psycholinguistic Research, Vol. 27, No. 5, 1998

Anomaly Detection: Eye Movement Patterns

Weijia Ni,[1,5,6] Janet Dean Fodor,[1,2] Stephen Crain,[1,3] and Donald Shankweiler[1,4]

Accepted May 3, 1998

The symptom of a garden path in sentence processing is an apparent anomaly in the input string. This anomaly signals to the parser that an error has occurred, and provides cues for how to repair it. Anomaly detection is thus an important aspect of sentence processing. In the present study, we investigated how the parser responds to unambiguous sentences that contain syntactic anomalies and pragmatic anomalies, examining records of eye movement during reading. While sensitivity to the two kinds of anomaly was very rapid and essentially simultaneous, qualitative differences existed in the patterns of first-pass reading times and eye regressions. The results are compatible with the proposal that syntactic information and pragmatic information are used differently in garden-path recovery.

THE ROLE OF ANOMALY IN SENTENCE PROCESSING

Much of the experimental work on the human sentence processing mechanism (the parser) has been concerned with how structural ambiguities are

The research reported here was supported in part 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 would like to thank David Swinney, Janet Nicol, and an anonymous reviewer for their helpful comments.

[1] Haskins Laboratories, New Haven, Connecticut, 06511.
[2] City Unversity of New York Graduate Center, New York, New York, 10036.
[3] University of Maryland, College Park, Maryland, 20742.
[4] University of Connecticut, Storrs, Connecticut, 06269.
[5] Yale University, New Haven, Connecticut, 06520.
[6] Address all correspondence to Weijia Ni, Haskins Laboratories, 270 Crown St., New Haven, Connecticut 06511.

0090-6905/98/0900-0515$15.00/0

resolved. Structural ambiguities can cause garden paths; that is, at the point of ambiguity, the parser may select an analysis which subsequently proves to be incompatible with later words of the input. The symptom of a garden path is this incompatibility, an apparent anomaly in the input. The error actually responsible for the incorrect garden-path analysis occurs earlier, but the parser does not discover that it has made an error until the occurrence of the anomaly reveals that something has gone wrong. Recovery from the garden path consists in finding an alternative analysis without the anomaly.

A well-known example of a garden path is Sentence (1), created by Bever (1970), which illustrates the general point.

(1) The horse raced past the barn fell.

The parser initially analyzes the ambiguous verb *raced* as the main verb, and when the real main verb *fell* is encountered, the sentence appears to be syntactically anomalous. In fact, *fell* is compatible with the earlier words, but since they have been misstructured there is no way to integrate *fell* into the current analysis. Either the parser will reject the sentence as ill formed, or else—ideally—it will be provoked by the apparent anomaly into reanalyzing the earlier part of the sentence to make it compatible with the "anomalous" input. How easy it is for the parser to achieve this has been shown to differ for different kinds of garden-path construction. In this particular example, reanalysis is difficult; typically, the parser remains trapped in the garden path. But in other instances, recovery is quite easy and rapid. Indeed, there is reason to believe that garden-path sentences are common in ordinary discourse and text, and that the parser is constantly getting into and quickly recovering from minor garden paths without their causing any disruption or surfacing in consciousness. (For examples and discussion, see for instance, Frazier & Clifton, 1995; Inoue & Fodor, 1994; Frazier & Fodor, 1978.) Though there are theories which advocate at least partial parallel processing (e.g., Gibson, 1991; Gorrell, 1989), or delayed decisions (Pritchett, 1992), all current models assume that the parser makes some decisions in advance of decisive evidence, and thus sometimes selects a garden-path analysis.

There is growing interest in modeling the process of garden-path recovery, with the aim of explaining why some garden paths are more difficult to recover from than others (see papers in Ferreira & Fodor, in press). In some models, the parser returns to an earlier point in the sentence and reparses it using just the same mechanisms as before but selecting a different analysis. In other models, the parser retains the structure it has computed so far and tries to alter parts of it to fit the "anomalous" input. In either case, the apparent anomaly not only triggers the reanalysis attempt, but also provides the best information the parser has about the nature of its prior error and what to do to put it right.[7] Thus the parser's ability to identify anomalies is likely to be a very important factor in the ease with which it can recover from garden paths. In order to understand better how the parser resolves garden paths (and why it sometimes fails to), we need to know more about how it responds to anomalies in sentences where issues of ambiguity do not arise. This is the focus of the present paper.

We will examine the parser's responses to different types of anomalies, namely, syntactic anomalies and pragmatic anomalies.[8] This is of particular interest because there is some evidence that syntactic and pragmatic anomalies may differ in their efficacy as triggers of successful reanalysis. It appears that, when the parser must override a previously built syntactic structure, it gives less weight to the information provided by a pragmatic anomaly than to the information in a syntactic anomaly (Fodor & Inoue, 1994, in press). One example discussed by Fodor and Inoue is shown in Example (2). Sentence (2a) is a garden path disambiguated by a pragmatic cue (the anomaly of *meeting a story*), while (2b) is disambiguated by a syntactic cue (the ungrammaticality of an infinitive clause *not to go home* as the complement of *met*. In both sentences, the parser at first prefers to analyze *that the girl met*. . .as a complement clause rather than a relative clause in the null context (Crain & Steedman, 1985). In both cases recovery from the garden path requires reattaching the *that*-clause as a relative inside the noun phrase (NP) headed by *boy*. According to informal intuitions of processing cost, (2b) is easier to reanalyze than (2a) is.

(2) a. They told the boy that the girl met the story.
Garden-path analysis: They told [the boy][that the girl met the story].
Correct analysis: They told [the boy that the girl met][the story].
b. They told the boy that the girl met not to go home.
Garden-path analysis: They told [the boy][that the girl met. . .].
Correct analysis: They told [the boy that the girl met][not to go home].

[7] In some models, though not all, the parser keeps a record of choice points during the parse, which could provide another source of useful information for reanalysis.

[8] For brevity, we will describe the contrast as between *syntactic anomaly* and *pragmatic anomaly*. These terms have been used with various meanings. We will use the term syntactic anomaly as a shorthand to denote violations of rules of syntax and formal compositional semantics and those aspects of reference, etc, that fall within the grammar. We will use the term *pragmatic anomaly* for matters of (im)plausibility and all inferencing based on general knowledge of the world and information provided by the discourse (see Crain, Ni, Shankweiler, Conway, & Braze, 1996)

This and other examples support the generalization that pragmatic anomalies are less effective cues for garden-path reanalysis than syntactic anomalies are.

Why this might be so is a difficult but important question, whose answer may shed light on issues of modularity of the language faculty (J. A. Fodor, 1983). One goal of the present paper is to open this question for investigation by articulating some possible answers and considering how they could be empirically evaluated. A strong modularity hypothesis would imply that the syntactic component of the parser does not know about, or does not care about, a problem experienced by the pragmatic component; altering the syntactic structure is work, and the effort might be undertaken only if the syntactic parser itself stands to gain from it. Alternatively, an explanation considered by Fodor and Inoue (1994) is that pragmatic cues carry less information than syntactic cues do concerning the nature of the repairs that need to be carried out. Though interesting, it is not clear how these explanations might be distinguished experimentally, and therefore they will not be addressed directly in the current paper. In the next section, we consider three specific hypotheses raised by previous research on anomaly detection that bear on the modularity issue and that do lend themselves to experimental testing.

THREE HYPOTHESES ARISING FROM PREVIOUS RESEARCH

If all research on garden-path resolution involves responses to apparent anomalies, as we have argued that it does, the research literature potentially relevant to our topic would be too vast to review here. To focus this material for present purposes, we organize it in relation to three types of testable hypotheses for why pragmatic and syntactic anomalies might differ in their effectiveness in initiating garden path reanalysis: A lag in pragmatic processing; a difference in the perceived degree of anomaly; or different estimates of the source of the anomaly.

First Hypothesis: Pragmatic Processing Lag

If a pragmatic anomaly does not immediately trigger garden-path revision, this might be because pragmatic processing is slow, so that the parser's detection of a pragmatic anomaly lags behind syntactic anomaly detection. The sources of this lag might be either a modular architecture, which entails that syntactic decisions take precedence over extramodular pragmatic processing, or it might be simply the practical fact that pragmatic processing is potentially open-ended while syntactic processing draws on a restricted set of computations, limited to their domain. Distinguishing between these two explanations of a pragmatic processing delay is extremely difficult in practice, in part for methodological reasons discussed in Fodor, Ni, Crain, and Shankweiler (1996). We refer interested readers to that discussion; for space reasons we do not review it here. For present purposes, these two proposals can be treated as variants of a single pragmatic delay hypothesis.

Proponents of constraint-based parsing models have emphasized the *rapidity* of pragmatic computations, including the detection of pragmatic anomalies. If that is correct, it would clearly tell against this account of the ineffectiveness of pragmatic cues for reanalysis. For example, Boland, Tanenhaus, Garnsey, and Carlson (1995) claimed that in filling "*wh*-gaps" in questions, effects of the (im)plausibility of thematic role assignments are "clear and immediate" (p. 774). They found that subjects judged sentences like (3a) to be anomalous ("stop making sense") right at the verb *visit*; this happened significantly more often than at the verb *visit* in the baseline Sentence (3b)

(3) a. Which prize did the salesman visit while in the city?
  b. Which client did the salesman visit while in the city?

However, Boland et al. did not give response times for the anomaly judgment to *visit* in (3a). They noted only that "word by word judgment times are slower than in 'normal' self-paced reading" (p. 782). Boland, Tanenhaus, and Garnsey (1990) reported that positive responses ("makes sense") in this sense-monitoring task are made approximately 500–800 ms after word presentation, but they gave no latency data for negative ("stop making sense") responses. It is not possible, therefore, to gauge how much time it took the parser to detect that a word was pragmatically anomalous in its context. Furthermore, these and related studies have not in general compared the speed of pragmatic anomaly detection with the speed of syntactic anomaly detection, which is what matters for the first hypothesis.

Studies of event-related brain potentials (ERPs) may bear on the issue of the relative speed of syntactic and pragmatic anomaly processing. It has been found that pragmatic/semantic anomalies elicit an enhanced negative wave with an onset around 200 ms and a peak amplitude at about 400 ms (the so-called N400; see Kutas & Van Petten, 1988, for a review). The N400 results have been replicated by many researchers for different languages such as English, Dutch, German, and French (e.g., Brown & Hagoort, 1993; Holcomb & Neville, 1990; Osterhout & Holcomb, 1992; for further references, see Friederici & Mecklinger, 1996). More varied ERP effects have been found for the processing of syntactically ill-formed sentences. A late centroparietal positivity (P600, or syntactic positive shift; see Osterhout &

Holcomb, 1995, for a review) is elicited by anomalies involving phrase structure, verb subcategorization, and subject-verb number agreement (e.g., Hagoort, Brown, & Groothusen, 1993; Neville, Nicol, Barss, Forster, & Garrett, 1991; Osterhout & Mobley, 1995). An early left anterior negativity (LAN), possibly less robust, has also been observed at around 250 to 350 ms following some syntactic anomalies. Friederici and Mecklinger (1996) suggested that P600 reflects processes of syntactic *re*analysis while the left anterior negativity reflects the detection of syntactic anomalies (see also Mecklinger, Schriefers, Steinhauer, & Friederici, 1995).

These studies, however, did not make direct comparisons of syntactic versus pragmatic anomalies. In a study using sentence materials which contain syntactic and pragmatic anomalies in closely matched contexts (based on the materials for the experiment to be reported below), Osterhout, et al. (1994) found that pragmatic anomalies elicited the expected N400 effect, while for syntactic anomalies the earliest sensitivity was a left anterior negativity effect within a 300- to 500-ms time window (but not statistically significant). A robust P600 effect was observed for the syntactic anomalies. A persistent negative shift at the sentence final word was found for both anomaly types. In sum, it seems clear that syntactic and pragmatic anomalies elicit qualitatively distinct ERP responses, but more research is needed to clarify the precise temporal relations between them. There is some evidence suggesting that, at least within approximately 300 to 400 ms, both syntactic and pragmatic anomalies are registered by the parser. However, the early negative response to syntactic anomaly, and its exact time frame, need to be established more firmly before we can assess whether or not syntactic and pragmatic anomaly detection have different time signatures.

Aside from ERP findings, two recent studies have addressed the timing issue. A study by Fodor et al. (1996) used the cross-modal lexical decision paradigm with compressed speech input (approximately 178 ms per word). The purpose of using compressed speech for presentation of the sentence materials was to force the processing routines to work at their maximum rate in the hope that this would more clearly reveal any timing differences between pragmatic and syntactic processing: If pragmatic computations are intrinsically slow, they might not be able to keep pace with the rapid phonetic input, and so would fall progressively behind syntactic processing (see Chodorow, 1979). The subjects performed a lexical decision task to semantically unrelated visual targets while listening to the test sentences (cf. Shapiro, Zurif, & Grimshaw, 1987). The lexical decision latencies were assumed to reflect differential processing loads imposed by the sentences, which contained either a syntactic anomaly, a pragmatic anomaly, or no anomaly (the baseline), in matched sentential contexts. These sentences were identical except for necessary length manipulations to those of the eye movement experiment to be reported below. The lexical targets appeared at five time points, ranging from 81 ms before the occurrence of the anomaly in the sentence to 243 ms after it. The results showed a significant effect of the anomaly relative to the baseline, focused very sharply at the test point simultaneous with the anomaly. This was true for both types of anomaly. Thus, detection was essentially simultaneous for pragmatic and syntactic anomalies, within the limits of measurement imposed by the 81-ms timing grid.

This bracketing of a possible timing difference may be narrowed further by an experiment by McElree and Griffith (1995) in which subjects made an acceptability judgment at the final word of a simple sentence such as (4a) to (4c)

(4) a. Some people love books.
b. Some people agree books.
c. Some people alarm books.

Sentence (4a) is fully acceptable; (4b) is syntactically anomalous (a subcategorization error: *agree with/about books*); (4c) is pragmatically anomalous. Sentences were presented visually one word at a time. Subjects were trained to respond promptly at a cue following the sentence at a variety of intervals (at 14, 157, 300, 557, 800, 1,500, and 3,000 ms after the onset of the final word). This method permits measurement of speed-accuracy trade-offs, and allows estimation of the earliest time at which an anomalous word affects performance. Data for the seven test points were input to a curve-fitting algorithm to project performance between those points. This gave an estimate of 233 ms for the earliest sensitivity to syntactic anomalies as in (4b), and 279 ms for sensitivity to pragmatic anomalies as in (4c), a significant difference of 46 ms. This is consistent with the Fodor et al. (1996) results discussed above, but seems to provide positive evidence for a pragmatic processing lag. It will be extremely interesting if such a positive timing difference, however small, can indeed be firmly established. However, it is not entirely clear whether the effect in McElree and Griffith's experiment might not be due to the fact that the dependent measure was response time for anomaly *judgments* on the sentences. In our own research, we have found reason to believe that explicit judgments of syntactic and pragmatic anomalies have very different time courses, which do not necessarily reflect the speed of linguistic processing per se. In an anomaly monitoring experiment reported by Fodor et al. (1996, Experiment 3), syntactic anomalies were responded to 524 ms faster than pragmatic anomalies. This is in contrast to the lack of an observable timing difference for exactly the same sentence materials in the dual-task cross-modal paradigm, where judgment of the anomalies is not called for. Thus it is still an open question whether there is any significant delay at all in pragmatic processing relative to syntactic processing.

To summarize: Timing considerations do not definitively rule out the first hypothesis as a potential explanation for why pragmatic symptoms might be ineffective in garden-path repair. However, if there is a timing difference it is evidently very small. Could such a small difference have a significant impact on garden-path recovery? This is not out of the question, at least if the small difference were magnified in a ''winner-take-all'' system, such that if syntax gains even a slight headstart, then pragmatics must wait until syntax has completed all of its work. This possibility is discussed in Fodor et al. (1996), but as yet there is little empirical evidence that bears on it.

Second Hypothesis: Degree of Anomaly

Pragmatic anomalies may be perceived as less severe than syntactic inconsistencies. If so, the different effectiveness of syntactic and pragmatic cues to reanalysis might really be a difference in cue strength. This may indeed be so in some cases, but the data reported in the literature provide little support for it. The qualitative noncomparability of the two types of anomaly probably precludes direct comparative judgments by subjects. However, the study by McElree and Griffith (1995) cited above found that asymptotic levels of anomaly judgments were identical for syntactic and pragmatic anomalies: Regardless of the speed of detection, an equal number of ''anomalous'' judgments were made for sentences like (4b) and (4c). In the cross-modal lexical decision experiment reported by Fodor et al. (1996), responsiveness to pragmatic anomalies (as measured by the amount of interference with the concurrent task) was damped when the postsentence comprehension task was easy (a simple yes/no question). This suggests that under demanding conditions (caused by the unusually rapid sentence input in that experiment) the parser may sacrifice pragmatic evaluation when it can get away with doing so. However, a shift to a more stringent secondary task (giving a verbal paraphrase of the sentence) brought sensitivity to the pragmatic anomaly up to the same level as for the syntactic anomaly.

Again, a final conclusion is not easy to draw from the evidence available in the current literature. It is obviously important to check whether the superiority of syntactic cues for garden-path reanalysis is observed for materials in which syntactic and pragmatic cues have been closely matched for severity of the anomaly, insofar as this is practically possible. However previous research offers no reason to believe that pragmatic anomalies are systematically underrated by perceivers.

Third Hypothesis: Qualitatively Different Responses

It could be that a pragmatic anomaly is detected just as quickly as a syntactic inconsistency is, and is perceived as equally severe, but that perceivers react to it differently. For instance, in a modular system, as noted above, a syntactic problem could be perceived as a sign of a prior syntactic error, while a pragmatic problem is taken as a sign of a prior pragmatic error. Alternatively, a pragmatic problem may be misattributed to the producer. When an apparent anomaly is encountered there are two possibilities: Either it really exists in the sentence, or it is due to the parser's own error in analyzing the sentence. The parser has no basis for deciding on-line which is the case. It needs some policy to guide its response to the anomaly. A wise strategy would be to routinely attempt reanalysis, just in case there is a parsing error that needs correcting, but the parser might be less inclined to do so in the case of a pragmatic anomaly because speakers do say odd things. It is not out of the question that the intended message was about some bizarre event (e.g., *the wolf spoke to Red Riding Hood; the girl met the story*). And messages that seem strange at first may make sense as more of the situation is known (e.g., *I couldn't do my homework because my brother swallowed an ant*). We know of no prior experimental studies that address the third hypothesis. If it is correct, then a structural revision would not be initiated by a pragmatic problem at least until the parser has had a chance to evaluate whether the problem might really be in the input rather than in the parse assigned to it. Hence, though the pragmatic anomaly is registered as it occurs, the parser would continue to process the incoming sentence, and only later make a decision about reanalysis.

The experiment reported here was designed to evaluate all three hypotheses. The technique of recording eye movements during reading makes it possible to explore whether responses to syntactic and pragmatic anomalies differ in timing, or magnitude, or in a qualitatively distinct manner. The dependent measures in eye movement recording are less constrained than in many other paradigms, such as the cross-modal lexical decision paradigm we have previously used. Patterns of eye movements can vary in several ways in response to an unexpected element in a sentence: Fixation on that elements may be prolonged, or subsequent elements may be given more attention, or there may be regressions to preceding words, or to the beginning of the sentence, or to a point in the sentence suspected of being the origin of the problem (see Frazier & Rayner, 1982; Ni, Crain, & Shankweiler, 1996, etc.). The eye-movement recording technique employed in the present experiment provides two on-line measures of sentence processing, namely, the measure of fixation durations and the measure of frequency of eye regressions. Different profiles of eye movement responsiveness may thus be obtained for different types of anomaly.

EXPERIMENT

Purpose

The aim of the experiment was to find out whether pragmatic and syntactic anomalies in matched contexts disrupt reading at different times, or to different degrees, and/or in qualitatively different ways. The anomalies tested appeared in fully unambiguous sentences. Though not themselves garden-path constructions, the materials were not unlike garden-path constructions from a perceiver's point of view, since a garden path is perceived as anomalous until or unless successful reanalysis occurs. In our materials, however, the anomalies were not eliminable by restructuring earlier portions of the sentence.

The subject's only task was to read the sentences and answer a comprehension question following some of them; eye movements were recorded. As noted above, we consider it best to avoid as far as possible any deliberative judgment of sentences as anomalous or nonanomalous since this might engender some difference between the syntactic and pragmatic conditions that is not characteristic of normal sentence processing (see Fodor et al., 1996). Therefore, no attention was drawn to the anomalies in the instructions or the practice session prior to the experiment. Nonetheless, subjects were exposed to some clearly anomalous sentences, which is a potential disadvantage of this approach. The failure of expectations could lead subjects to distrust the experimental stimuli, so normal processing routines may not be tapped. However, though not ideal, this is a feature of nonjudgment paradigms in the testing of ''difficult'' garden paths too. Thus, our results may not speak to garden-path processing in case of ambiguities so easy that recovery is effortless and invariably successful; but our experimental situation is in keeping with that in classic investigations of garden paths such as *The horse raced past the barn fell*, or *When Mary was mending the sock fell off her lap* (Gorrell, 1989; Traxler & Pickering, 1996; Trueswell, Tanenhaus, & Garnsey, 1994).

Method

Subjects

Twenty-four college students, all native speakers of English, participated in the experiment. They all reported normal vision or normal vision with contact lenses. They were not informed of the purpose of the experiment, and had not been exposed to the test materials prior to the test.

Materials

Thirty triads of sentences were used. All sentences were unambiguous, both locally and globally. Within a triad, sentences were identical except that they contained either a syntactic anomaly, as illustrated by (5a), a pragmatic anomaly (5b), or no anomaly (baseline, 5c). The three versions of each test sentence differed by only one word, always the verb (italicized in the examples in 5, though not in the experiment). In the two anomaly versions, this word was the first item that was incompatible with previous words in the sentence.

(5) a. It seems that the cats won't usually *eating* the food we put on the porch.
b. It seems that the cats won't usually *bake* the food we put on the porch.
c. It seems that the cats won't usually *eat* the food we put on the porch.

As noted, these materials were used in the Fodor et al. (1966) experiments discussed above, and they were a subset of those tested in the ERP study by Osterhout et al. (1994). They were edited to reduce their length so that they could be presented on a single line (maximum 76 characters). Mean frequencies of the verb stems in question were matched across sentence versions (for the pragmatic anomaly version = 89.46 per million; for the syntactic anomaly version and the baseline = 79.96 per million; Francis & Kucera, 1982). A complete list of the test sentences is given in the Appendix.

Three stimulus lists were generated, each containing 10 tokens of each of the three versions of the test sentences. These sentences were pretested for acceptability. Acceptability ratings were obtained from 12 native speakers of English who did not participate in the eye movement experiment. Each stimulus list was given to four scorers. They were asked to circle a number on a scale from 1 to 7 after reading each sentence, 1 denoting *fully acceptable* and 7 denoting *totally unacceptable*. Unacceptability was defined very broadly, to include defects of any kind perceived by the scorer. The rating scores are as follows: The mean score for the baseline sentences was 2.183, ranging from 1 (Sentence 4) to 4.5 (Sentences 19 and 25). (The sentence numbers here refer to sentences listed in the Appendix.) Scorers' comments indicated that this somewhat high rate of unacceptability of baseline sentences was due to the adverbs following the modal verbs in the sentences. The scores for the syntactically anomalous sentences were significantly higher than those for the baseline, ranging from 4.5 (Sentence 22) to 6.5 (Sentences 2 and 16), with a mean of 6.025. Likewise, the scores for the pragmatically anomalous sentences were also significantly higher than those for the baseline, ranging from 3.5 (Sentence 13) to 6.75 (Sentences 27 and 30), with a mean of 5.542. The difference between the scores for the two anomalous versions was statistically significant on the item analysis but not on the subject analysis. These judgments indicated that the pragmatic anomalies were indeed perceived as somewhat less severe than the syntactic anomalies, thus not entirely excluding an explanation along the line of the second hypothesis. However, the degree of overlap in the rating scores between the two anomaly versions was considerable, so it appeared that a severity difference was not likely to be a major factor in determining eye-movement patterns.

For the eye-movement experiment, the test sentences were interspersed quasirandomly among 76 fillers, with at least one filler between consecutive test sentences. The fillers were constructed with a variety of syntactic structures in order to camouflage the test sentences; 58 of them were acceptable while 18 were anomalous. Each stimulus list thus consisted a total of 106 sentences, and was presented to eight subjects. The two halves of each stimulus list were interchanged so that of the eight subjects who saw one of the lists, four received it in one order and the other four in the other order. Order of presentation made no significant difference in the data analysis, and therefore will not be discussed further.

Equipment

Subjects' eye movements were recorded using the IRIS infrared (IR) light eye-movement system (SKALAR model 6500). IRIS uses a differential reflection method of eye-movement recording. Eye movements are registered by directing an invisible infrared light beam into the subject's eye. Infrared-light emitters and detectors are positioned in front of the subject's eyes so that the receptive fields match the iris-sclera boundary, both on the nasal and on the temporal side. Upon horizontal rotation of the eye, the nasally positioned detector measures an increase (or decrease) in scleral IR reflection, while the temporally placed detector measures a corresponding decrease (or increase) in IR reflection. Subtraction of the nasal and temporal detector signals gives eye position with respect to head position. Eye positions are sampled every millisecond and they are transmitted to a computer equipped with an analog-to-digital conversion board.

Since eye positions are measured relative to head positions, the subject's head was stabilized by using a bite bar and a forehead rest. The subject's jaw rested on a wax-coated metal plate (secured on a shaft) made up with his/her dental impression. A foam-coated semicircular headrest was firmly secured in front of the subject's forehead so that he/she could lean against it. The subject was seated in a chair with height adjustment to achieve maximal comfort. The computer monitor was placed on a platform approximately 64 cm (25 in.) away from the subject's eye. The visual angle of each character was slightly greater than 12 min of arc, allowing for a resolution of eye position better than one character. All sentences appeared in mixed case in Courier 14-point font.

Procedure

Each subject was given instructions that contained a brief description of the eye-movement recording device and a diagram of the head restraint platform at which s/he would be seated. Before the experiment began, the brightness of the monitor was adjusted to the subject's comfort. The infrared emitter and receiver were aligned in front of the subject's right eye, but viewing was binocular. The experiment began with a calibration procedure in which the subject fixated on a series of screen positions while the computer sampled eye fixations. These samples were used by the computer to derive a set of linear equations that converted the horizontal eye position signals into horizontal screen coordinates. The calibration process usually required 2 to 3 min.

Before proceeding with the experimental materials, the experimenter emphasized to the subject that each sentence should be read at a normal pace. The subject was instructed to fixate on a cross at the center left of the screen. The sentence display was initiated by the press of a computer mouse. The whole sentence appeared on the screen. When the subject finished reading the sentence, he/she pressed the mouse button and the sentence disappeared. After each trial, a series of screen coordinates appeared for the experimenter to check the calibration. Minor adjustments were performed occasionally. Recalibration was seldom necessary.

Subjects were instructed to read the sentences for understanding; a short comprehension question asking about the topic of the sentence followed 22 filler sentences. These questions were quasirandomly scattered during the test session. For example, after the sentence ''The stranger at the scene of the robbery had a scar on his right cheek,'' a question followed: ''TOPIC = CRIME?'' Subjects answered the question by clicking the mouse at a box marked ''YES'' or a box marked ''NO'' on the screen. Audio (speech) feedback was given by the computer to inform the subject whether or not the answer was correct. This task was intended to ensure that subjects actually processed the sentences. All subjects performed well on this task (at least 90% correct). A practice session preceded the experimental session, after which subjects were given the opportunity to pause and ask questions.

Results

Data analyses were conducted on experimental sentences only. Reading times and regressions were calculated for six regions, identical across the three versions, shown below:

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

Region 1 was the beginning of the sentence prior to Region 2; it was zero to three words long. Region 2 consisted of the last two words of the subject NP; this was always the head noun plus one preceding word (determiner or adjective). Region 3 was the modal verb plus a following adverb. Region 4 was the critical 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, zero to four words long.

The dependent variables were first-pass reading time and percent of regressions.[9] The critical comparison between versions of the test sentences started from the region that contained the main verb, and the verbs in each triad varied in length (e.g., *eating, bake, eat*); therefore, total fixation times would be expected to vary. For this reason we used a measure of residual reading time instead of total fixation durations (see Trueswell et al., 1994, for a detailed discussion of this measure). To compute residual reading times, a regression analysis was performed on each subject's reading time data, using the number of letters and spaces in each region as the independent variable, and fixation durations at each region of each sentence for that subject as the dependent variable. This permitted calculation of a "residual" for each region of each sentence, representing the deviation from the reading time expected if reading time were fully determined by region length. (Thus region length is removed as a factor in subsequent data analyses.) The measure of regressive eye movements for each region was the percentage of subjects' first-pass fixations that resulted in a leftward glance out of that region.

Figure 1 illustrates the profile of first-pass residual reading times for the two anomalous versions relative to the corresponding baseline residual reading times. Thus the baseline scores are effectively set to zero in this figure. Actual residual reading time scores for all three versions are shown in Table Ia; and raw reading time scores are shown in Table Ib. No significant differences between versions were found at Regions 1 through 3. At

[9] Second-pass reading times are sometimes reported in the literature in addition to regressions, but these data are not uniformly interpretable (for discussion see Ni et al., 1996)

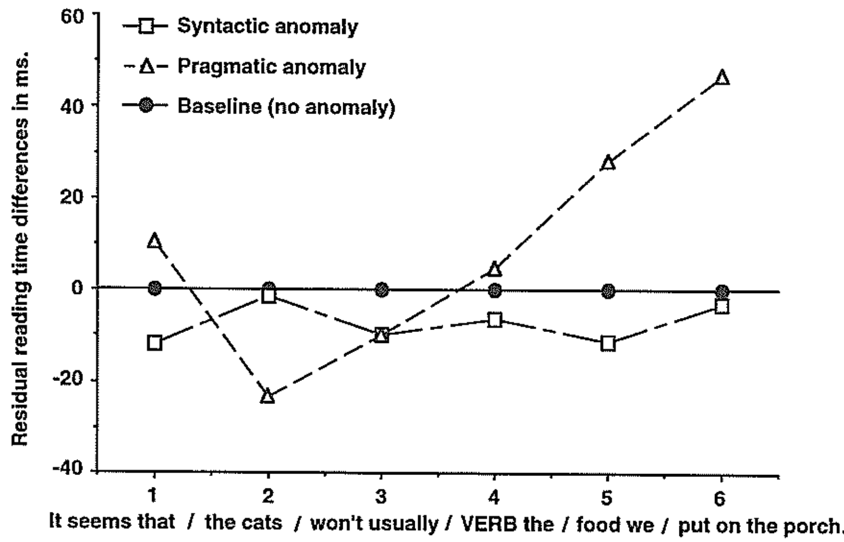

Fig. 1 Mean first-pass residual reading times for anomalous sentences relative to baseline.

Table Ia. Mean First-Pass Residual Reading Times (in Milliseconds) at Each Region for the Three Versions of the Test Sentences

Region	1	2	3	4	5	6
Syntactic anomaly	−20.72	76.97	1.31	6.91	−60.99	−69.65
Pragmatic anomaly	1.34	55.03	1.41	18.04	−21.31	−19.67
Baseline (no anomaly)	−8.84	78.35	11.15	13.23	−49.56	−66.62

Table Ib. Mean First-Pass Raw Reading Times (in Milliseconds) at Each Region for the Three Versions of the Test Sentences

Region	1	2	3	4	5	6
Syntactic anomaly	315.17	598.68	492.66	444.80	385.75	410.58
Pragmatic anomaly	337.15	577.63	492.07	375.52	426.64	473.35
Baseline (no anomaly)	334.55	598.01	497.67	373.80	407.44	412.48

At Region 4, where the verb created an incompatibility in the syntactic anomaly and pragmatic anomaly sentences, there were still no differences between versions. Thus, there was no immediate effect of either type of anomaly on the reading time measure. At Region 5, reading times for the pragmatic anomaly sentences began to show signs of slower processing; reading times for this version differed numerically, though not statistically, from both the baseline and the syntactic anomaly version. The final region revealed even slower reading of the pragmatic anomaly sentences: The increase in reading time compared to baseline was significant in the analysis by subjects, and approached significance in the analysis by items [F_1 (1, 23) = 6.221, p = .0202; F_2 (1, 29) = 3.263, p = .0820]. Reading times for pragmatic anomaly were also longer than for the syntactic anomaly version, though the difference was significant only in the analysis by subjects [F_1 (1, 23) = 5.141, p = .0335; F_2 (1, 29) = 2.059, p = .1628]. This strong tendency to slow down following the pragmatic anomaly version resulted in a statistically significant difference between the pragmatic anomaly sentences and the baseline for Regions 5 and 6 together [F_1 (1, 23) = 5.450, p = .0287; F_2 (1, 29) = 8.395, p = .0071]. In addition, there was a significant difference between the pragmatic anomaly version and the syntactic anomaly version when Regions 5 and 6 were combined [F_1 (1, 23) = 7.411, p = .0121; F_2 (1, 29) = 7.311, p = .0113]. The syntactic anomaly version did not differ significantly from the baseline at any sentence position.

The regressive eye movements showed a different pattern. Figure 2 depicts the profile of regressions for the two anomalous versions relative to the baseline. As in Fig. 1, this reduces scores for the baseline sentences to zero; see Table II for actual scores. Regression patterns showed no differ-

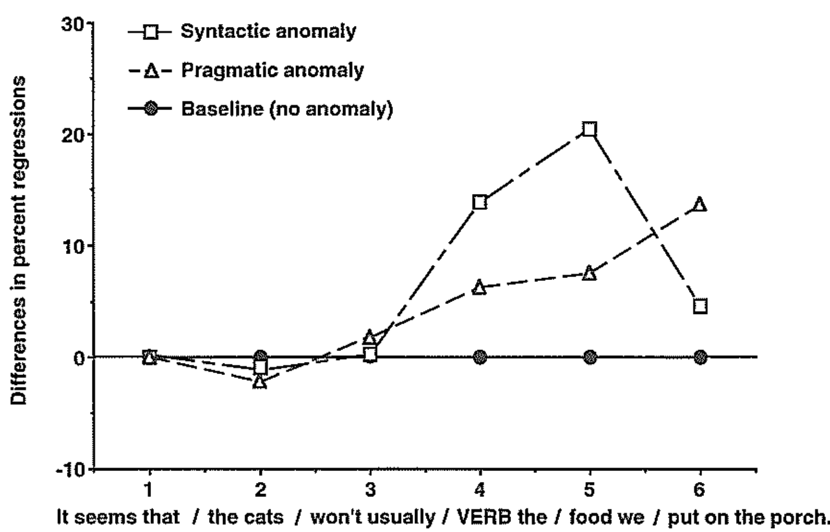

Fig. 2 Percent of eye regressions from anomalous sentences relative to baseline.

Table II. Mean Percent (%) of Regressions from Each Region for the Three Versions of the Test Sentences

Region	1	2	3	4	5	6
Syntactic anomaly	0.0	7.0	9.2	19.7	29.2	21.1
Pragmatic anomaly	0.0	6.0	10.7	12.1	16.4	30.1
Baseline (no anomaly)	0.0	8.2	8.9	5.8	8.8	16.5

ences between sentence versions at Regions 1 through 3 where the versions were lexically identical. An increase in regressions occurred for both anomaly types (compared with baseline) at Region 4, which contained the critical verb. Syntactic anomaly sentences induced significantly more regressions than baseline sentences at this region [19.7% vs. 5.8%; F_1 (1, 23) = 12.89, p = .0015; F_2 (1, 29) = 20.83, p = .0001]. This effect persisted at Region 5 [29.2% vs. 8.8%; F_1 (1, 23) = 17.64, p = .0003; F_2 (1, 29) = 41.15, p = .0001], but was absent at Region 6 ($p > .1$). The regression rate for the syntactic anomaly sentences had dropped to the baseline level by Region 6. For the pragmatic anomaly, more regressions occurred at Region 4 than for the baseline at that region (12.1% vs. 5.8%). The effect was significant in the analysis by items, and approached significance in the analysis by subjects. [F_1 (1, 23) = 3.81, p = .0631; F_2 (1, 29) = 4.35, p = .0459]. A similar strong trend was observed at Region 5 in the comparison between the pragmatic anomaly and the baseline [16.4% vs. 8.8%; F_1 (1, 23) = 3.21, p = .0862; F_2 (1, 29) = 7.82, p = .0088]. For Regions 4 and 5 together, the difference between the pragmatic anomaly version and the baseline did attain significance [F_1 (1, 23) = 6.674, p = .0166; F_2 (1, 29) = 16.811, p = .0003]. At Region 6, unlike the profile of the syntactic anomaly version, the regression rate for the pragmatic anomaly version was still climbing, differing significantly from that of the baseline [F_1 (1, 23) = 4.26, p = .0504; F_2 (1, 29) = 5.52, p = .0263]. This was the case despite the fact that the regression rate for the baseline also increased from Region 5 to Region 6 [from 8.8% to 16.5%; F_1 (1, 23) = 5.623, p = .0183; F_2 (1, 29) = 7.150, p = .0079], presumably due to a sentence "wrap-up" effect.

We now consider the comparison between the two anomalous sentence versions with respect to regressions. The syntactic anomaly induced numerically more regressions than the pragmatic anomaly at Region 4 (19.7% vs. 12.1%); only the analysis by items approached significance [F_1 (1, 23) = 2.938, p = .1000; F_2 (1, 29) = 3.616, p = .0675]. At Region 5, the difference between the two anomaly versions was significant [F_1 (1, 23) = 8.139, p = .0090; F_2 (1, 29) = 13.068, p = .0011]. When the results from Regions 4 and 5 were pooled, the difference between the two anomaly versions was highly significant [F_1 (1, 23) = 9.247, p = .0058; F_2 (1, 29) = 25.634, p = .0001]. Thus the syntactic anomaly triggered more regressions in this portion of the sentence than the pragmatic anomaly did. The different regression trajectories of the two anomalous versions over the last three regions (4, 5, and 6) gave rise to a statistically significant interaction of anomaly type (syntactic vs. pragmatic) by region (4, 5, and 6) [F_1 (2, 23) = 4.358, p = .0186; F_2 (2, 29) = 4.292, p = .0184].

To summarize: Eye-movement responsiveness to syntactic and pragmatic anomalies showed some similarities. Specifically, the onset of regressions was immediate for both kinds of anomaly within the limits of measurement imposed by the paradigm—that is, within the time taken to read the verb region (mean 401.55 ms). However, there were also differences between the anomaly types. Pragmatic anomaly, but not syntactic anomaly, resulted in progressively longer first-pass reading time than the baseline sentences. At least in this respect, the response to pragmatic anomalies was more extreme than that to the syntactic anomalies. Syntactic anomaly showed a higher regression rate than for pragmatic anomaly until the final region of the sentences. In short, the two types of anomaly induced somewhat different patterns of disruption in reading. Syntactic anomalies caused an immediate, but short-lived, spate of regressions. By contrast, the effects of pragmatic anomaly included both slower forward reading times and increased regressions; both effects were weak at first but became progressively stronger as the sentence continued.

Discussion

Patterns of eye movements obtained from this experiment did not provide strong support for the first two hypotheses; they were more in keeping with the third hypothesis. For both anomaly types, there was rapid response to the anomaly in the form of eye regressions; the only sign of a possible pragmatic lag was that the response at the verb for the pragmatic anomaly version was not so robust as that for the syntactic anomaly. Neither was there evidence for a difference in the perceived severity between the two types of anomaly. Both induced clear emergency responses, i.e., regressions occurred at the point of anomaly with negligible differences between the two versions. To further test this indication that it was not a difference in perceived degree that distinguished the syntactic and pragmatic processing, we conducted an analysis on a subset of the test materials (17 out of 30 test triads) whose mean acceptability rating scores were very closely matched between the two types of anomaly in each triad (mean difference = 0.04 on a 7-point scale). Results for this subset exhibited somewhat more varied patterns, as would be expected with fewer data points, but otherwise showed very close resemblance to the patterns presented in the previous section for the entire data set.

The results do provide some encouragement for the third hypothesis, that the parser perceives both types of anomaly equally well but reacts to them differently. In particular, the data are compatible with the idea that the parser is more inclined to keep an open mind on the source of an anomaly if it is pragmatic in nature than if it is syntactic. If it can be assumed that regressions reflect attempted reprocessing, the low initial rate of regressions for the pragmatic anomalies suggests that perceivers continued for a while to attempt to make sense of the sentence before rereading it.

It must be borne in mind, of course, that the anomalous sentences tested here were of limited types; many more varieties of anomalous sentences, both syntactic and pragmatic, need to be investigated before definite conclusions can be reached. Caution is in order until further research has been conducted to determine whether the profiles of response observed here are also associated with other sorts of syntactic and pragmatic problems. Subject to this understanding, these results demonstrate a qualitatively different pattern of eye movement in response to syntactic and pragmatic anomalies, suggestive of a tendency to more extended evaluation of pragmatic problems before a revision is attempted.

No other studies, as far as we know, have directly compared eye-movement responses to different types of anomaly in comparable sentence contexts. This is difficult to obtain in experiments on garden-path sentences because of difficulties in constructing garden-path constructions that systematically differ only in whether their symptoms are syntactic or pragmatic anomalies. However, if we look at data for one type of anomaly or the other, there are some comparisons to be made between the results of prior research and those of the present study.

For syntactic anomalies, Pearlmutter, Garnsey, and Bock (1995) obtained results similar to those reported here. They found a significant increase in regressions but no significant increase in first-pass reading times in response to ungrammaticalities consisting of mismatches of agreement features.[10] However, other researchers have reported an increase in both regressions and reading times for syntactic symptoms in garden-path constructions. For example, Ferreira and Clifton (1986) observed both effects in response to the *by*-phrase in sentences such as (6a), compared with (6b):

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

[10] However, a summed gaze measure (the "sum of first-pass times for each word in a region") did show significant sensitivity to the ungrammaticality.

Of interest is why this construction elicits a mixed response involving both regression and reading time. It may be relevant that the anomaly in (6a) is not purely syntactic, since it turns on the agentive reading of the lexically ambiguous item *by*. Also, the anomaly of the *by*-phrase in (6a) shows that the meaning assigned earlier (*the witness doing the examining*) was wrong, while in the materials of our present study and those of Pearlmutter et al., the syntactic anomaly did not affect the semantic interpretation of the word string. It could be that, if the needed reprocessing is simultaneously syntactic and pragmatic, an increase in both reading time and regression occurs.

For pragmatic anomalies, prior results appear to lean even more strongly than ours do, toward an increase in reading time as the characteristic effect. Ferreira and Clifton (1986) observed increased first-pass reading times but no increase in regressions at a word such as *examined* in (6b), which is pragmatically anomalous as a main verb. Ni et al. (1996) tested sentences such as (7), which are pragmatically anomalous on the attachment of the prepositional phrase *with large cracks* to the verb *painted*:

(7) The man painted the doors with large cracks before the festival. (cf. The man painted the doors with new brushes before the festival.)

An immediate increase in first-pass reading times was found at the point of anomaly, i.e., at the noun phrase *large cracks*, but there was no significant overall increase in regressions. The meaning of the noun phrase indicates that the instrumental meaning of the preposition *with* is inappropriate, so *with* must be reconstrued in its accompaniment or property sense. This sense is compatible only with attachment to NP, so a structural reattachment must follow as a secondary consequence of the pragmatic/semantic revision. The fact that the primary revision here is pragmatic/semantic would explain why the response to it shows up in longer reading times.[11] This garden path also

[11] There is evidence, however, that persons with low memory span may deviate from this pattern. Ni et al. (1996) found that subjects showed a significant increase in regressions, but no increase in reading times, in response to the pragmatic anomaly in sentence whose analysis is quite demanding, such as (i):

(i) The man painted only doors with new brushes before the festival.
(cf. The man painted only doors with large cracks before the festival.)

Ni et al. contended that in a sentence like (i), the focus operator *only* changes the parser's preference so that it attaches the prepositional phrase to the object NP rather than to the verb, creating a pragmatic anomaly at the noun phrase *new brushes*. The proposal is that, for this particular group of subjects, memory capacity was completely overloaded by the referential processing triggered by *only*, by the time the anomaly was encountered. On-line reanalysis was therefore impossible, hence the high rate of regressions.

clarifies the scope of Fodor and Inoue observation on the relative ineffectiveness of pragmatic cues for restructuring. Since revision evidently does occur in examples like (7), it appears that a pragmatic anomaly *can* trigger semantic revision. Normally, however, (e.g., in 2a) a semantic revision does not point to a specific syntactic repair, as it does in (7).

A common and very reasonable assumption about when readers in difficulty continue to read forward slowly and when they regress is that revision is conducted on-line if possible, but rereading is necessary if the task becomes difficult. The current research has raised the possibility that the determinant of eye movements is not (or not only) the *degree* of difficulty, but the *type* of difficulty involved. On the basis of our data and a new look at results reported in the literature, the following generalization begins to emerge: that syntactic work in response to an anomaly triggers regression, while semantic/pragmatic work in response to an anomaly is more likely to invoke slower forward reading (but see footnote 11). In resolving a typical garden path, both syntactic and semantic/pragmatic work are required, so the consequence is a mix of both prolonged reading time and increased regression. Further evaluation of these ideas must await additional research.

SUMMARY AND CONCLUSION

The present results confirm that the eye-movement recording technique is sensitive to differences in levels of sentence processing. They are novel in that they establish qualitative differences in the parser's responses to syntactic and pragmatic anomalies in closely matched sentence contexts and in a within-subject design. An added advantage is that these materials (or related ones) have also been tested in ERP, cross-modal lexical decision, and anomaly monitoring studies, as reported above. Important next steps are to broaden the range of anomaly types tested, and where possible, to examine similar anomalies when they occur as the error signals in garden-path constructions.

Implications of the current data for garden-path recovery processes are not conclusive but of interest. We considered the proposed generalization that only syntactic anomalies reliably trigger structural reanalysis. There is some preliminary evidence for this in informal judgments of difficulty for a varied range of different garden path constructions (see Fodor & Inoue, 1994, in press). There is also experimental evidence, for examples such as *The girl realized the man cheated on the exam*, tested by Ferreira and Henderson (1991). These authors concluded:

> Even if the anomalous nature of the sequence *realized the man* is detected by the language processing system, this information does not initiate reanalysis.

> Instead, the parser begins reanalysis only after receiving information [i.e., *cheated*] indicating that its initial analysis is syntactically ill-formed. . . .Thus, a more plausible model would be one where reanalysis processes are initiated most reliably by a syntactic error signal. (p. 316)

It should be noted that this goes beyond the more familiar proposition that pragmatic anomaly fails to block the development of a first-pass analysis of a sentence on-line but may guide subsequent reanalysis (see, for example, Rayner, Carlson, & Frazier, 1983). The stronger claim is that, even in the reanalysis stage, pragmatic information does not play a leading role in bringing about sentence structure alterations. The experimental results reported here indicate that there is a qualitative difference in how syntactic and pragmatic information is made use of by the sentence processing routines. And the specific nature of the difference adds the strength to the view that the human language perception mechanism tends to put its trust in the *form* of a sentence and accept whatever sentence *content* may result—one of the fundamental characteristics of a modular language faculty proposed by J. A. Fodor (1983).

APPENDIX: TEST SENTENCES

Each sentence has three versions, represented by underlined verbs: <u>baseline/syntactic anomaly/pragmatic anomaly.</u>

1. It seems that the cats won't usually <u>eat/eating/bake</u> the food we put on the porch.
2. Apparently, his argument might even <u>prove/proving/shout</u> that there are canals on Mars.
3. The new alarm system will surely <u>warn/warning/swear</u> about intruders in the hallway.
4. Some species of orchid will only <u>grow/growing/sing</u> in tropical rain forests.
5. This expensive ointment will supposedly <u>cure/curing/loathe</u> all forms of skin disease.
6. That old electric blender doesn't really <u>crush/crushing/own</u> ice cubes any more.
7. This exotic spice might possibly <u>add/adding/seek</u> the subtle flavor she craves.
8. The new fighter plane can apparently <u>fly/flying/walk</u> faster than anyone had expected.
9. The large wooden boxes may still <u>hold/holding/find</u> many old photographs of the family.

10. This math test might occasionally fail/failing/hate to identify gifted students.
11. The roof-top helicopters may repeatedly shake/shaking/paint the windows of the building.
12. The plumber said the water may slowly seep/seeping/speak out from behind the sink.
13. Those three fingerprints could clearly prove/proving/judge that the defendant is guilty.
14. A family of beavers would sometimes chew/chewing/melt the garden hose beside the shed.
15. The fancy German clock doesn't always tell/telling/ask the time accurately.
16. The latest rap songs might supposedly tend/tending/learn to lead young people astray.
17. Those small red spiders would often spin/spinning/burn pretty webs in the rose-bushes.
18. This kind of pacifier will immediately soothe/soothing/drop the cranky baby at bedtime.
19. The partly-built skyscraper might eventually block/blocking/seek out all the sunlight.
20. These French grape vines don't usually grow/growing/jog well in dry sandy soil.
21. At the zoo, one elderly bear would just sit/sitting/swear all day by the cage door.
22. Great Uncle Henry's portrait doesn't really look/looking/talk like him at all.
23. The space heater should quickly dry/drying/kick the towels hanging on the rack.
24. Newly-planted lettuces might soon tempt/tempting/lift many rabbits into the yard.
25. These colored yarns shouldn't ever fade/fading/cry if you wash them in cold water.
26. This chemical filter may also tend/tending/want to remove salt from sea water.
27. In summer, the sea-lions can happily bask/basking/read on the beach all day long.
28. The bank's security camera will now take/taking/tear a photograph of every customer.
29. The fierce bull could easily smash/smashing/mend the wooden fence around the field.
30. The defective lever doesn't reliably shut/shutting/leap off the air-conditioner.

REFERENCES

Bever, T. G. (1970). The cognitive basis for linguistic structure. In J. R. Hayes (Ed.), *Cognition and the Development of Language*. New York: John Wiley.

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). Verb argument structure in parsing and interpretation: Evidence from wh-questions. *Journal of Memory and Language, 34*, 774–806.

Brown, C. M., & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience, 5*, 34–44.

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.

Crain, S., Ni, W., Shankweiler, D., Conway, L., & Braze, D. (1996). Meaning, memory and modularity. In C. Schütze (Ed.) *Proceedings of the NELS 26 Sentence Processing Workshop*. MIT Occasional Papers in Linguistics, Vol. 9.

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.

Ferreira, F. & Fodor, J. D. (in press). *Reanalysis in sentence processing*. Hingham, MA: Kluwer Academic Press.

Ferreira, F., & Henderson, J. M. (1991). Recovery from misanalyses of garden-path sentences. *Journal of Memory and Language, 30*, 725–745.

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.

Fodor, J. D., & Inoue, A. (in press). Garden path diagnosis: The grammatical dependency principle. In M. Ryan (Ed.), *CUNY Forum* (vol. 20).

Fodor, J. D., Ni, W., Crain, S., & Shankweiler, D. (1996). Tasks and timing in the perception of linguistic anomaly. *Journal of Psycholinguistic Research, 25*, 25–57.

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

Frazier, L. & Clifton, C. (1995). *Construal*. Cambridge, MA: MIT Press.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition, 6*, 291–325.

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

Friederici, A. D., & Mecklinger, A. (1996). Syntactic parsing as revealed by brain responses: First-pass and second-pass parsing processes. *Journal of Psycholinguistic Research, 25*, 157–176.

Gibson, E. (1991). *A computational theory of linguistic processing: Memory limitations and processing breakdown*. Unpublished doctoral dissertation, Carnegie Mellon University, Pittsburgh.

Gorrell, P. (1989). Establishing the loci of serial and parallel effects in syntactic processing. *Journal of Psycholinguistic Research, 18*, 61–74.

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift as an ERP measure of syntactic processing. *Language and Cognitive Processes, 8*, 439–483.

Holcomb, P. J., & Neville, H. J. (1990). Semantic priming in visual and auditory lexical decision: A between modality comparison. *Language and Cognitive Processes, 5*, 281–312.

Inoue, A., & Fodor, J. D. (1994). Information-paced parsing of Japanese. In R. Mazuka & N. Nagai (Eds.), *Japanese sentence processing*. Hillsdale, NJ: Erlbaum.

Kutas, M., & Van Petten, C. (1988). Event-related brain potential studies of language. In P. K. Ackles, J. F. Jennings, & M. G. H. Coles (Eds.), *Advances in psychophysiology*. Greenwich, CT: JAI Press.

Mecklinger, A., Schriefers, H., Steinhauer, K., & Friederici, A. D. (1995). Processing relative clauses varying on syntactic and semantic dimensions: An analysis with event-related potentials. *Memory & Cognition, 23*, 477–494.

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.

Neville, H., Nicol, J., Barss, A., Forster, K., & Garrett, M. (1991). Syntactically based processing classes: Evidence from event-related potentials. *Journal of Cognitive Neuroscience, 3*, 151–165.

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

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

Osterhout, L., & Holcomb, P. J. (1995). Event-related brain potentials and language comprehension. In: M. D. Rugg & M. G. H. Coles (Eds.), *Electrophysiology of mind: Event-related brain potentials and cognition*. Oxford, England: Oxford University Press.

Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language, 34*, 739–773.

Osterhout, L., Nicol, J., McKinnon, R., Ni, W., Fodor, J. D., & Crain, S. (1994). *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, New York.

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

Pritchett, B. L. (1992). *Grammatical competence and parsing performance*. Chicago: University of Chicago Press.

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

Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language, 35*, 454–475.

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.