THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 1997, *50A* (3), 481–497

Influence of Connectives on Language Comprehension: Eye-tracking Evidence for Incremental Interpretation

Matthew J. Traxler

University of Glasgow, Glasgow, U.K. and The Florida State University, Tallahassee, Florida, U.S.A.

Michael D. Bybee

St. John's College, Santa Fe, New Mexico, U.S.A.

Martin J. Pickering

University of Glasgow, Glasgow, U.K.

An eye-tracking experiment investigated whether incremental interpretation applies to interclausal relationships. According to Millis and Just's (1994) *delayed-integration* hypothesis, interclausal relationships are not computed until the end of the second clause, because the processor needs to have two full propositions before integration can occur. We investigated the processing of *causal* and *diagnostic* sentences (Sweetser, 1990; Tversky & Kahneman, 1982) that contained the connective *because*. Previous research (Traxler, Sanford, Aked, & Moxey, 1997) has demonstrated that readers have greater difficulty processing diagnostic sentences than causal sentences. Our results indicated that difficulty processing diagnostic sentences occurred well before the end of the second clause. Thus comprehenders appear to compute interclausal relationships incrementally.

Theories of discourse processing must specify the kinds of representations readers build from texts and the process by which those representations are constructed. Researchers working on discourse have paid particular attention to connectives. Writers use connectives like *and*, *because*, and *therefore* to signal relationships between different pieces of text. It is generally assumed, therefore, that connectives play a role in the specification of the discourse representation (Beck, McKeown, Sinatra, & Loxterman, 1991; Caron, Micko, &

Requests for reprints should be sent to Matt Traxler, Department of Psychology, The Florida State University, Tallahassee, FL 32306–1051, U.S.A. Email: traxler@psy.fsu.edu

We acknowledge the assistance of Keith Edwards, Simon Garrod, Simon Liversedge, and Tony Sanford. We also acknowledge Gerry T.M. Altmann and an anonymous reviewer for comments on a previous draft. This research was supported by ESRC Grant No. R000234542 and a British Academy Post-Doctoral Fellowship (both awarded to the third author).

Thuring, 1988; Goldman & Murray, 1992; Halliday & Hassan, 1976; Knott & Dale, 1995; Loman & Mayer, 1983; Lorch, 1989; Millis, Graesser, & Haberlandt, 1993; Sanders, Spooren, & Noordman, 1992; Segal, Duchon, & Scott, 1991). It is somewhat less clear how, precisely, connectives influence discourse processing (see, e.g. Britton, Glynn, Meyer, & Penland, 1982; Geva & Ryan, 1985; Haberlandt, 1982; Marshall & Glock, 1978; Meyer, Brandt, & Bluth, 1980; Murray, 1995; Noordman, Vonk, & Kempff, 1992), and so theories of discourse processing differ in terms of the processing consequences they attribute to the presence or absence of connectives. This article explores one proposal concerning connectives' influence on text processing, namely Millis and Just's (1994) *delayed-integration hypothesis.* According to this, comprehenders process two clauses linked by a connective by interpreting each clause separately and combining them when they reach the end of the second clause. We provide evidence against this proposal and suggest instead that the semantic processing needed to interpret such complex sentences takes place incrementally.

Incremental interpretation refers to the idea that semantic interpretations are assigned to fragments of text. Highly incremental accounts (e.g. Altmann & Steedman, 1988; Garrod, Freudenthal, & Boyle, 1994; Marslen-Wilson, 1973, 1975; Pickering, 1993, 1994; Tyler & Marslen-Wilson, 1977) propose that semantic interpretations are computed on an almost word-by-word basis as text is read. Less incremental accounts propose that some semantic operations are delayed until the ends of larger constituents, such as clauses or sentences (Just & Carpenter, 1980; Kintsch, 1988; Kintsch & van Dijk, 1978; Millis & Just, 1994).

There is good evidence that many aspects of language comprehension are performed very rapidly. Lexical access normally occurs at or before the end of a word (e.g. Marslen-Wilson, 1987; Rayner & Duffy, 1986). Prior context has rapid effects on this process, indicating that the meaning of each word is quickly integrated with previously processed text (e.g. Fischler & Bloom, 1979; Frazier & Rayner, 1990; Glucksberg, Kreuz, & Rho, 1986; Seidenberg, Tanehaus, Leiman, & Bienkowski, 1982; Swinney, 1979). Garden-path phenomena demonstrate that comprehenders build syntactic representations before encountering phrase or clause boundaries (e.g. Altmann & Steedman, 1988; Bever, 1970; Ferreira & Clifton, 1986; Ferreira & Henderson, 1990; Frazier, 1979; Frazier & Rayner, 1982; Mitchell & Holmes, 1985; Rayner, Carlson, & Frazier, 1983; Trueswell, Tanenhaus, & Garnsey, 1994; Trueswell, Tanenhaus, & Kello, 1993). Discourse context can rapidly influence lexical processing, syntactic parsing, anaphoric resolution, and other comprehension processes (Altmann, Garnham, & Dennis, 1992; Altmann & Steedman, 1988; Britt, 1994; Britt, Perfetti, Garrod, & Rayner, 1992; Ehrlich & Rayner, 1983; Garrod et al., 1994; Marslen-Wilson, Tyler, & Koster, 1994; Shillcock, 1982; Spivey-Knowlton, Trueswell, & Tanenhaus, 1993; Swinney, 1979). Many of these findings demonstrate that readers access and employ information from preceding text before reaching clause- or sentence-boundaries to resolve lexical, syntactic, and referential ambiguities.

A great deal of evidence suggests that comprehenders rapidly construct semantic interpretations, even for incomplete syntactic structures. "Off-line", readers can assess the plausibility of fragments whose interpretation within standard formal semantics (e.g. Montague Semantics: Dowty, Wall, & Peters, 1981) is non-propositional (Chater, Pickering, & Milward, 1995). During processing, plausibility effects occur before clause- or sentence-boundaries are reached (Boland, Tanenhaus, Garnsey, & Carlson, 1995; Clifton, 1993; Garnsey, Tanenhaus, & Chapman, 1989; Garrod et al., 1994; Holmes, Stowe, & Cupples, 1989; Marslen-Wilson et al., 1994; Stowe, 1989; Traxler & Pickering, 1996; Tyler & Marslen-Wilson, 1977). For instance, Traxler and Pickering monitored eye-movements while subjects read implausible sentences like (1):

> That's the garage with which the heartless killer shot the hapless man yesterday afternoon. (1)

Reading was disrupted as soon as subjects reached the verb *shot*, which is implausible in this context, in comparison with plausible control sentences containing the same verb.

Despite the existence of a great deal of evidence supporting very rapid semantic processing, some theories propose that certain semantic work is delayed until clause or sentence-boundaries are reached. For example, Millis and Just's (1994) *connective-integration* proposal suggests that when a connective intervenes between two clauses, integration of the propositions expressed by those clauses is delayed. They propose that when comprehenders encounter a connective, they store the content of the preceding clause in a working memory buffer while processing the clause that follows the connective. The preceding clause is less accessible during processing of the following clause. When the end of this second clause is reached, the connective causes comprehenders to reactivate the content of the preceding clause (if the clause is "related" to the content of the second clause) and to integrate their interpretations. Crucially, reactivation does not occur before the end of the second clause.

More generally, the construction-integration theory of discourse processing (Kintsch, 1988; Kintsch & van Dijk, 1978) also proposes that integration occurs at the ends of sentences or clauses (even though other kinds of processing are done "immediately"). This processing strategy follows naturally from the emphasis on propositions as the functional unit of semantic representation. On these accounts, text is interpreted as propositions, and text integration cannot occur before at least two complete propositions are presented to the processor that is responsible for integration. Construction-integration also assumes rough equivalence between clauses and propositions. In summary, construction-integration proposes that readers construct integrated semantic representations for multiple sentences or clauses at or very soon after they encounter the final word of the last sentence or clause.

This hypothesis accounts for *wrap-up* (Aaronson & Scarborough, 1976; Just & Carpenter, 1980; Mitchell & Green, 1978), which is the finding that reading times increase at the end of sentences and clauses. Just and Carpenter propose the *immediacy hypothesis*, under which many processes of language comprehension occur as soon as possible (in line with incremental processing). However, they argue that the construction of interclause relations, including those that require inferences, take place during wrap-up at the end of the sentence. Hence, Just and Carpenter's model is compatible with Millis and Just's (1994) more specific account.

To test connective integration, Millis and Just (1994) performed a series of experiments using self-paced reading combined with a recognition probe task. Subjects read either two sentences or a two-clause sentence (with the same content as the two sentences) with a connective (*because*) appearing between the two clauses. Responses to a recognition probe presented either near the beginning or at the end of the second statement (i.e. the second sentence or the second clause) indexed activation of the first statement. The recognition probes were either the main verb from the first statement or the main verb from the second statement, or a word that did not appear in the test sentences. In the first experiment, probes were presented at the end of the second statement. Readers responded more rapidly to the main verb from the first clause when a connective was present. In the second experiment, an early probe point was also tested. The connective produced no facilitation at the early test point, but the effect at the late test point was replicated. The third experiment manipulated causal relatedness between the two statements, and the fourth experiment tested a different connective (*although*), with compatible results.

A crucial component of connective integration, the *delayed-integration* hypothesis, proposes that readers "set the representation of the first statement aside in working memory as they construct a representation of the second statement, after which the two clausal representations are integrated together" (Millis & Just, 1994, p. 135). Thus, connective integration proposes that when readers encounter a connective, they shut off processing of the preceding clause, reactivating it only when they reach the end of the second clause. It follows that the construction of a unified semantic representation of two clauses joined by a connective is not an incremental process. If this is correct, then the semantic characteristics of the first clause should not influence processing of the second clause until the end of the second clause, when a connective like *because* intervenes between the two clauses.

Millis and Just (1994) found some support for the delayed integration hypothesis in their reading-time data. If delayed integration is correct, then increased processing times attributed to clause-integration processes should not occur until the end of the second clause. Experiment 3 manipulated the relatedness of the two clauses. Readers read "moderately related" or "low-related" pairs of clauses conjoined by *because* and responded to recognition probes. Reading times for moderately and low-related second clauses did not diverge until the final word of the second clause, and then only if a connective appeared between the two clauses. Millis and Just suggested that readers delayed integrating the clauses until the end of the second clause.

Millis and Just (1994) might have failed to detect earlier evidence for disruption in the low-related cases because of the limitations of self-paced reading combined with a secondary probe-recognition task. In addition, the probe-recognition task may not be sensitive to relevant aspects of early semantic processing (see Discussion). It is possible that a technique that is more sensitive to momentary disruptions in processing (like eye-movement monitoring) might have detected effects where Millis and Just's technique did not. Thus, the current experiment used eye-movement monitoring to test the delayed integration hypothesis.

We investigated the processing of *causal* and *diagnostic* sentences containing the connective *because* (Sweetser, 1990; Traxler, Sanford, Aked, & Moxey, 1997; Tversky & Kahneman, 1982). Traxler et al. demonstrated that diagnostic statements like (2) take longer to comprehend than causal statements like (3).

Heidi could imagine and create things because she won first prize at the art show. (2)

Heidi felt very proud and happy because she won first prize at the art show. (3)

They suggest that this occurs because diagnostic statements require an inference to establish the nature of the causal consequence. Usually, this consequence constitutes a belief or conjecture about events in the world. (Note that all of Millis & Just's [1994] sentences were causal, like [3].)

The crucial question that the current experiment addresses is when the difficulty associated with diagnostic statements occurs during processing. If readers employ a delayed-integration strategy, then increased reading times associated with diagnostic statements should be observed at the end of the second clause. If readers adopt a more incremental strategy, then reading times for diagnostic and causal statements should diverge before the end of the second clause. We employed an eye-tracking paradigm to assess processing of causal and diagnostic statements because it is sensitive to incremental processes and because it closely approximates natural reading (Rayner & Pollatsek, 1989).

Method

Subjects

Thirty-two students from the University of Glasgow were paid to participate. Some of them had participated in other eye-tracking experiments.

Materials

Subjects read 24 causal and diagnostic sentences like (2) and (3) (see Appendix). In causal sentences, Clause 2 constitutes a plausible cause for the events described in Clause 1. In diagnostic sentences, Clause 2 constitutes plausible evidence that Clause 1 is a true statement (Sweetser, 1990). These sentences comprised a subset of the stimuli used by Traxler et al. (1997). The items were controlled for degree of association between the two clauses across conditions. The crucial second clause was identical across conditions.

To generate the items, two groups of 20 subjects were presented with the second clauses of the experimental sentences. The first group was asked to describe what the events in the second clause would cause. The second group was asked to describe what they could infer from the events in the second clause. We used the subjects' responses to construct first clauses for the experimental items. Across causal and diagnostic sentences, we matched the first clauses for frequency with which the events they described occurred in the subjects' responses.

Pilot studies demonstrated that readers process the first clauses of diagnostic sentences as rapidly as the first clauses of causal sentences. Thus, differences in difficulty of the first clauses cannot account for differences observed during processing of the second clauses. Pilot work and the experiments performed by Traxler et al. (1997) demonstrated that differences between causal and diagnostic sentences that occur in the second clause are due to the presence of the connective *because*. Differences between causal and diagnostic sentences in processing of the second clause disappear when the coordinating conjunction *and* appears between the first and second clause.

Procedure

An SRI Dual-Purkinje Generation 5.5 eye-tracker monitored subjects' eye movements. The tracker has an angular resolution of 10′. The tracker monitored only the right eye's gaze location. A PC displayed items on a VDU 70 cm from subjects' eyes. The VDU displayed four characters per degree of visual angle. The tracker monitored subjects' gaze location every millisecond, and the software sampled the tracker's output to establish the sequence of eye fixations and their start and finish times.

Before the experiment started, subjects read an explanation of eye-tracking and a set of instructions. The instructions told them to read at their normal rate and comprehend the texts as well as they could. The experimenter then seated the subject at the eye-tracker and used bite bars and forehead restraints to minimize head movements. Next, subjects completed a calibration procedure. Before each trial, a small "+" symbol appeared near the upper left-hand corner of the screen. Immediately subjects fixated the "+" symbol, the computer displayed an item, with the first character of the text replacing the "+" on the screen. The "+" symbol also served as an automatic calibration check, as the computer did not display the item until it detected stable fixation on the "+" symbol. If subjects did not rapidly fixate the "+" symbol, the experimenter recalibrated the eye-tracker. When subjects finished reading each item, they pressed a key, and the computer either displayed a comprehension question, on about half of the trials, balanced across conditions, or proceeded to the next trial. Half of these questions had "yes" and half had "no" answers. Subjects responded to the questions by pressing a button and did not receive feedback on their answers. After subjects completed each sixth of the experiment, the experimenter recalibrated the equipment. Thus, the eye-tracker was calibrated a minimum of six times during the experiment and often more.

The computer displayed each experimental list in a fixed random order together with 56 filler sentences. The first two sentences and the first two after each calibration were of varied grammatical type. The other 48 fillers made up two further experiments of 24 items each. The computer displayed all sentences on two lines, splitting the experimental sentences just before the connective *because*. At least one filler sentence separated experimental sentences.

Fixations

In the first stage of analysis, we determined which line of text subjects were reading. This involved some judgement, because subjects' eyes could land between two lines of text, and because slight head movements sometimes caused small, systematic changes in the recording of fixation location. However, in the vast majority of cases, determining which line subjects actually fixated caused no difficulty whatsoever. After this, an automatic procedure pooled short contiguous fixations. The procedure incorporated fixations of less than 80 msec into larger fixations within one character space and then deleted fixations of less than 40 msec that fell within three character spaces of any other fixation. Following Rayner and Pollatsek (1989), we presume that readers do not extract much information during such brief fixations.

Measures

First-pass time is the sum of the fixations occurring within a region before the eye leaves the region, either to the right or to the left. *Total time* is the sum of all fixations in a region. We also employed two less standard measures of early processing, because first-pass times are sometimes truncated by subjects entering a region and rapidly regressing, which can occur when they encounter processing difficulty (Brysbaert & Mitchell, 1996; Hemforth, Konieczny, Scheepers, & Strube, 1994; Liversedge, Pickering, & Traxler, 1996; cf. Rayner & Pollatsek, 1989). Thus, two pieces of text, one that causes processing difficulty and one that does not, can have similar first-pass times. In the first case, readers rapidly exit to the left; in the second case, readers rapidly exit to the right. *Right-bounded time* is the sum of all fixations within a region before the eye fixates any region to the right of the region. *Regression-path duration* includes all of the fixations within a region and all subsequent fixations on prior regions, until the eye crosses the region's right boundary. Note that if the eye fixates a point beyond the end of a region before landing in the region for the first time, then first-pass time, right-bounded time, and regression-path duration are zero.

Regions

Region 4, the final region, had to contain at least two words, so that we could be certain that subjects had not finished reading the sentence while fixating Region 3. If a prepositional phrase appeared at the end of the sentence, then Region 4 started at the beginning of the prepositional phrase. Thus, 12 of the items had three words or more in Region 4. Region 1 comprised the word *because*. Region 2 comprised the subject of the embedded sentence, or the subject and the main verb when the subject was a pronoun. Region 3 comprised the main verb and the object of the second clause (except in cases where the main verb was part of Region 2). Note that one item did not contain a final prepositional phrase. In this item, Region 2 comprised the subject, Region 3 comprised the main verb, and Region 4 comprised the object of the main verb in the embedded sentence. The Appendix shows the region boundaries for all of the items.

Analyses

The analyses reported below excluded 0-msec fixations that occurred when readers skipped a region on the first pass. Analyses that included 0-msec fixations produced nearly identical results.

Results

Mean first-pass, total, and right-bounded time and regression-path durations for the regions of each sentence type appear in Table 1. Preliminary analyses demonstrated that no reliable differences between causal and diagnostic sentences occurred during subjects' first pass through the first clause, $F_1(1, 31) = 2.57$, $p > .11$, $F_2 < 1$.[1] The data from the second clause were subjected to a series of one-way ANOVAs testing for effects of sentence type (causal vs. diagnostic). First-pass time, right-bounded time, and regression-path duration did not differ significantly between causal and diagnostic sentences in Regions 1 and 2, all $F < 1$.

Data from the third region provide the crucial test of the delayed integration hypothesis. If differences between causal and diagnostic sentences occur here, then readers did

[1] First-pass time for Clause 1 of causal sentences was 50 msec less than for diagnostic sentences (note that right-bounded time and regression-path durations are identical to first-pass time for first regions). This trend probably reflects slight differences in length between the two sentence types. Clause 1 of causal stimuli averaged 31 characters and 5.8 words, whereas diagnostic stimuli averaged 33 characters and 5.9 words. Neither difference is statistically reliable: within-items ANOVA for character difference, $F(1, 27) = 1.94$, $p > .18$; for word difference, $F(1, 27) < 1$. When the reading time data are subjected to a msec/character transform, the trend toward a sentence-type effect disappears entirely, $F_1(1, 31) = 1.65$, $p > .20$, $F_2(1, 27) < 1$.

TABLE 1
Mean First-pass, Right-bounded, Total Reading Time and Regression Path Duration by Sentence Type and Region

		Region			
		1	*2*	*3*	*4*
First-pass	Causal	228	318	330	541
	Diagnostic	228	321	391	595
Right-bounded	Causal	249	365	357	758
	Diagnostic	257	337	394	838
Regression-path Duration	Causal	249	373	361	1013
	Diagnostic	257	348	411	1302
Total Time	Causal	274	444	467	770
	Diagnostic	321	512	554	834

Example Clause 1 (causal): *Heidi felt very proud and happy . . .* (diagnostic): *Heidi could imagine and create things . . .*
Example Target Clause: *because/ she won/ first prize/ at the art show.*
"/" marks indicate region boundaries.

not wait until the end of the second clause to begin integrating the two clauses. In fact, all three measures of early processing (first-pass, right-bounded, and regression-path duration) show differences between causal and diagnostic sentences in Region 3. Region 4 effects and total time effects are reported for the sake of completeness.

Readers responded to diagnostic sentences more slowly than causal sentences on their first pass through Region 3, $F_1(1, 31) = 6.28$, $p < .02$, $F_2(1, 23) = 7.13$, $p < .02$. There was a trend in the same direction in Region 4, $F_1(1, 31) = 3.17$, $p < .09$, $F_2(1, 23) = 2.76$, $p < .12$. Right-bounded time and regression path duration produce similar effects in Regions 3 and 4: right-bounded, Region 3, $F_1(1, 31) = 4.36$, $p < .05$, $F_2(1, 23) = 6.63$, $p < .02$; Region 4, $F_1(1, 31) = 2.75$, $p < .11$, $F_2(1, 23) = 1.00$, n.s.; regression-path durations, Region 3, $F_1(1, 31) = 6.94$, $p < .01$, $F_2(1, 23) = 9.67$, $p < .01$; Region 4, $F_1(1, 31) = 19.69$, $p < .0001$, $F_2(1, 23) = 7.28$, $p < .01$.

There were differences in total time between causal and diagnostic sentences in Region 3, $F_1(1, 31) = 11.00$, $p < .01$, $F_2(1, 23) = 10.56$, $p < .01$. Similar differences occurred in Regions 1 and 2: Region 1, $F_1(1, 31) = 6.81$, $p < .01$, $F_2(1, 23) = 5.31$, $p < .03$; Region 2, $F_1(1, 31) = 3.01$, $p < .10$, $F_2(1, 23) = 5.55$, $p < .03$. Effects in Region 4 were in the same direction but weaker, $F_1(1, 31) = 2.04$, $p < .16$, $F_2(1, 23) < 1$.

Many current models of eye movements in reading propose that readers do not do any semantic processing on words until they fixate them directly (e.g. Rayner & Morris, 1992). However, some models of eye movements suggest that readers extract some information from words to the right of the word they are currently fixating (e.g. Hyona, 1995; Underwood, Bloomfield, & Clews, 1989). Before moving on, we must rule out one possible counter-explanation of our results: perhaps readers processed the end of the sentence while fixating near the end of Region 3. This could be a problem, because some of our final regions contained two very short function words. To see whether this could be the case, we analysed separately the 12 items with three or more words in the final region. No model of eye movements in reading proposes that readers can process the end of the sentence while fixating four words away. Thus, if differences between causal and diagnostic sentences appear in these analyses, that rules out the possibility that the effects in the main analyses resulted from readers processing the end of the sentence while looking at Region 3. These analyses did produce differences between causal and diagnostic sentences in all of the measures, with first-pass $F_1(1, 31) = 4.58$, $p < .05$, $F_2(1, 11) = 8.08$, $p < .02$, right-bounded $F_1(1, 31) = 4.25$, $p < 0.05$, $F_2(1, 11) = 11.38$, $p < .01$, and regression-path duration $F_1(1, 31) = 8.02$, $p < .01$, $F_2(1, 11) = 17.87$, $p < .01$. Thus, the effects for the whole set of items cannot be a result of readers processing the end of the sentences while fixating Region 3.

Discussion

The experiment demonstrated that construction of a unified semantic representation of two clauses can proceed incrementally. In this experiment, the clauses were joined by the connective *because*, which preferentially signals a causal relationship between clauses but can also signal a diagnostic relationship. Disruption from when subjects reached the middle of the second clause onwards indicates that readers incrementally constructed a semantic interpretation of the second clause and assessed it as a cause of the state of affairs described in the first clause long before they reached the end of the sentence. The fact that incremental interpretation can take place at this level is particularly striking because most previous demonstrations of incremental interpretation have concentrated on lower-level processes, such as assessing the plausibility of verb-argument relationships (e.g. Traxler & Pickering, 1996; Trueswell et al., 1994) and of referents for anaphoric expressions within a discourse context (e.g. Altmann et al., 1992; Britt, 1994). In this respect, the finding is compatible with Marslen-Wilson et al. (1994), who demonstrated incremental interpretation in spoken-language comprehension taking place over sentence boundaries.

Our findings are incompatible with the delayed-integration hypothesis of Millis and Just (1994) and with any account that predicts that the interpretations of clauses cannot be combined until the end of the second clause (e.g. Just & Carpenter, 1980; Kintsch & van Dijk, 1978). If readers set the representation of the first clause aside during processing of the second and combined them only at the end of the second clause, then differences between causal and diagnostic sentences should have appeared at the end of the second clause and not before. However, differences in reading times between causal and diagnostic sentences emerged well before the end of the second clause. Thus, it does not appear that readers set aside their representation of the first clause while building a semantic representation of the second. The early difference between causal and diagnostic sentences in this experiment is entirely consistent with incremental interpretation at the clausal level, but is difficult to reconcile with delayed integration. That is, the data suggest that readers attempted to construct a unified representation for the sentences as soon as possible. These data are thus at odds with Millis and Just's results. What might account for this difference?

The most obvious possible reason why our data suggest incremental interpretation while Millis and Just's (1994) data suggest delayed integration is that the methods differed. Previous research suggests that probe recognition tasks tap into relatively superficial representations (Cloitre & Bever, 1988; see also Garrod et al., 1994; Garrod & Sanford, 1994) and thus may not always be sensitive to processes involved in constructing semantic representations for complex sentences. Cloitre and Bever's experiments demonstrated that tasks that require the recovery of conceptual information produce earlier effects than do tasks (like probe recognition) that require the recovery of surface information. They also found that different kinds of probes (e.g. nouns vs. pronouns) produce different patterns of results (see also Dell, McKoon, & Ratcliff, 1983; MacDonald & MacWhinney, 1990; O'Brien, Duffy, & Myers, 1986).

In probe-recognition studies, experimenters typically present probes at one or two points in the text. Thus effects that occur some distance from a probe point may not affect performance on the probe task. In Millis and Just's (1994) experiments, probes were only presented immediately after the connective and at the end of the sentence. By contrast, eye-tracking data provide a more continuous record of readers' behaviour.

Pronoun resolution studies provide an example of the differential sensitivity of probe tasks and eye-tracking to some language comprehension processes. Probe-recognition tasks sometimes find reactivation effects for pronouns no earlier than the end of the sentence in which a pronoun appears (e.g. Gernsbacher, 1989; but cf. Stevenson, 1986). Other secondary-task paradigms, such as cross-modal lexical decision, sometimes produce similar results (Tyler & Marslen-Wilson, 1982; but cf. Shillcock, 1982). By contrast, eye-tracking studies produce evidence for much earlier pronoun resolution (e.g. Ehrlich, 1983; Ehrlich and Rayner, 1983; Garrod et al., 1994). Thus, while secondary-task paradigms could lead one to believe that readers do not identify a referent for a pronoun before the end of a sentence, the eye-tracking data demonstrate that they do. Differences between the tasks could, therefore, easily explain why the current study provides support for incremental interpretation of clausal relations, whereas Millis and Just's (1994) probe-recognition data led them to conclude that delayed integration occurs.

Apart from its inconsistency with the results of the present experiment, Millis & Just's connective integration proposal may also fail to generalize to other connectives. Consider processing of the coordinating conjunction *and*. Connective integration proposes that readers close off the processing of clauses when they encounter a connective. If that is the case, then connective integration predicts a garden-path in sentence (4), but not in sentence (5) (Frazier, 1979).

John saw Mary and Ted yesterday. (4)
John saw Mary and Ted laughed. (5)

Contrary to connective integration's prediction, Frazier found difficulty in the processing of sentences like (5) but not in sentences like (4). She attributed this to the processing principle of *late closure.*

The data from the current experiment suggest that readers initially attempt to interpret diagnostic sentences containing *because* as simple causal statements and that this process begins soon after readers encounter the connective *because*. This makes sense, given that a causal relationship is central to the discourse model for many texts and that the conjunction *because* normally signals a simple causal relationship. As readers process more and more of the second clause of diagnostic sentences, the causal interpretation becomes less and less plausible. As the causal interpretation becomes less plausible, reading times increase. Subjects either switch to the diagnostic interpretation or else fail to produce a plausible interpretation of the sentence. This is wholly consistent with previous findings regarding processing of causal and diagnostic sentences (Traxler et al., 1997) and with other research supporting incremental interpretation.

The delayed integration hypothesis does not predict the result we obtained. Is it possible to modify the hypothesis in some way to account for the present results? One possible alteration would be to suggest that integration occurs at any potential end-of-clause or end-of-sentence position, rather than occurring only at actual clause- or sentence-endings. This change would accommodate the current results, but it would also make predictions that would seem to go against the principles on which connective-integration is based. Consider processing of sentence (6):

Fred was irritated because John was* reading* a book* about gardening* to Sue* in a loud voice* this afternoon. (6)

The "*" marks indicate points at which the sentence could end but does not. If readers attempt integration at any potential end-of-clause, then they would do so at every point in sentence (6) marked by a "*". However, readers would also have to put the first clause back into the buffer each time, because more material would have to be added to the semantic representation of the second clause. Re-buffering the first clause would presumably incur some cost in processing resources. Given that the purpose of buffering is to preserve working memory resources, reactivating the first clause at every potential end-of-clause would seem to be a non-optimal strategy. It seems unlikely that advocates of connective-integration, construction-integration, or related frameworks would want to propose such a strategy.

Green, Mitchell, and Hammond's (1981) proposal may offer another way to reconcile connective-integration and more incremental accounts. Green and colleagues propose that text serves as instructions to processors that construct representations of texts. They propose that readers first translate the text into processing instructions and then execute those instructions. In this second stage, readers modify the existing text representation, evaluate the plausibility of the result, and execute processes that increase the coherence of the representation, such as making elaborative or bridging inferences. On the basis of self-paced reading data, Green et al. propose that modification and evaluation processes can occur at any time during processing, but that coherence-enhancing operations occur only at clause- or sentence boundaries. This proposal could account for both disruption occurring in the middle of sentences, like that which occurred in the present experiment, and effects occurring at the ends of sentences, like that which occurred in Millis and Just's (1994) experiments. On this view, semantic representations are built and