ACADEMIC PRESS

Journal of Memory and Language 47 (2002) 530–547

Journal of Memory and Language

www.academicpress.com

Coercion in sentence processing: evidence from eye-movements and self-paced reading☆

Matthew J. Traxler,[a,*] Martin J. Pickering,[b] and Brian McElree[c]

[a] *Department of Psychology, University of South Carolina, Columbia, SC 29208, USA*
[b] *New York University, USA*
[c] *Edinburgh University, USA*

Received 11 June 2001; revision received 20 February 2002

Abstract

Seemingly simple expressions may require an enriched form of interpretative processing. Verbs like *began* and *finished* can be used felicitously only when one of their arguments denotes an event (e.g., *reading*). However, such verbs commonly appear with noun phrases whose literal interpretations denote entities (e.g., *the book*). It has been suggested that readers and listeners have to undertake additional computations to interpret strings like *began the book* that are not required when *the book* is interpreted as an entity (e.g., Pustejovsky, 1995). If so, *began the book* should be harder to process than strings like *read the book*, when the verb does not require an argument that denotes an event, or strings like *began the fight*, when the argument denotes an event. Experiment 1 found evidence from eye movements showing that entity noun phrases take longer to process following verbs that require event arguments than verbs that do not. Experiments 2 and 3, using eye-tracking and self-paced reading, respectively, found that difficulty did not appear when verbs like *began* had arguments that referred to events. We interpret the results with respect to accounts of semantic processing.

Keywords: Coercion; Semantics; Sentence processing

What processes are used to map *form* to *meaning* in language comprehension? A standard view holds that syntactic processes build structured expressions, and that semantic processes assign each expression an interpretation by combining lexical representations according to their position in syntactic structure. Recent work in lexical semantics has challenged this somewhat simplistic view, arguing that interpretation often involves richer and more complex forms of composition. Particularly, it has been suggested that composition is context-sensitive (e.g., Pustejovsky, 1991, 1995). For example, default interpretations of individual expressions can be modified by other elements in the sentence through a form of *type shifting* (Bach, 1986; Partee, 1992; Pustejovsky, 1991, 1995), in which one element coerces another

☆ Order of authorship is arbitrary. The research was supported by the NICHD grant 1R01HD040865-01 A1, the University of South Carolina SPAR program, and by the South Carolina Commission on Higher Education (awarded to M.T.) and by NIMH grant MH57458 (awarded to B.M.). We thank Steven Frisson for comments on the manuscript. We thank Don Donaldson for technical support.

* Corresponding author. Fax: 1-803-777-4302.
E-mail address: traxlerm@gwm.sc.edu (M.J. Traxler).

0749-596X/02/$ - see front matter
PII: S0749-596X(02)00021-9

to shift from one semantic type to a different type, and, moreover, interactions between expressions can introduce semantic structure that is not explicitly represented in the surface form of the sentence (Jackendoff, 1997; Pustejovsky, 1991, 1995).

Both the types of operations appear to be evident in the interpretation of seemingly simple sentences like (1):

(1) *The author began the book.*

Verbs like *began* semantically select for a complement that expresses an event (or activity; for supporting arguments, see Jackendoff, 1997; McElree, Traxler, Picketing, Jackendoff, & Seely, 2001; Pustejovsky, 1991, 1995; Pustejovsky & Bouillon, 1995). However, the default interpretation of the complement *the book* in (1) is not as an activity or an event, but as an entity. Despite this apparent conflict, readers and listeners do recover an interpretation for (1), the most typical being one that can be paraphrased as *The author began to write the book* (see McElree et al., 2001, and below for supporting normative data). In this interpretation, the complement appears to have been type-shifted from the semantic type *entity* to the type *event* by interpolating additional semantic structure, which, if it were to be overtly realized in syntax, would be expressed either as an infinitival clause (*...to write the book*) or as a participial clause (*...writing the book*).

The particular way in which the complement is coerced, and hence the specific interpretation that is recovered, appears to be determined by interactions between the lexical–semantic properties of the verb and its complement, as well as information in the broader context. Verbs like *begin*, *start*, and *finish*, when paired with a complement like *the book*, select particular events commonly associated with the noun phrase (NP), typically ones in which the entity is used for a designated purpose or ones which express the means by which the entity was created. For an NP like *the book*, *reading* is perhaps the most common designated purpose, while *writing* (but also *editing*, *printing*, etc.) are examples of the latter. That *the book* in (1) is typically interpreted as *to write the book* is due to the presence of the agent *the author*. Other agents, like *the student*, *the copywriter*, *the typesetter*, *the designer*, *etc.*, would not circumvent the need for type-shifting the complement from an entity to an event, but would alter the content of the interpolated semantic structure.

Knowing that readers are able to interpret sentences like (1) does not tell us how, exactly, that interpretation is accomplished. This paper presents evidence that supports what we term an *enriched composition* hypothesis, following terminology proposed in lexical–semantic research (e.g., Jackendoff, 1997; Pustejovsky, 1991, 1995). We assume that sentences like (1) require an enriched form of composition that involves a semantic coercion operation. Formally, coercion can be defined as an operation that converts an expression, α, to the semantic type expected by a governing function, β (Pustejovsky, 1995). The expression α (*the book* in example 1) is assumed to have a set of type shifting operations that may operate over it, which serve to alter its type and denotation. In cases like (1), the verb *begin* is assumed to select for an eventual function, and that the coercion operation converts the expression *the book* from its default semantic type *entity* to the type *event* by interpolating semantic structure that is synonymous with expressions like *began to write the book* or *began writing the book*.

To evaluate accounts of this class, we examined reading time measures for sentences like (1), which are assumed to require additional cognitive operations, comparing them to reading times for sentences in which the default interpretation of the complement matches or satisfies the requirements of the governing verb. The latter include cases like (2), in which the verb can semantically select for an entity, and cases like (3), in which the complement matches the verb's requirements for an event:

(2) *The author saw the book.*

(3) *The author began the lecture.*

The reported research extends earlier findings that constructions like (1) are associated with greater processing costs than control conditions that are not thought to require an enriched form of composition. In a self-paced reading (moving window) experiment, McElree et al. (2001) examined sentences like (4):

(4a) *The author was starting the book in his house on the island.*

(4b) *The author was writing the book in his house on the island.*

(4c) *The author was reading the book in his house on the island.*

In sentences such as (4a), *was starting*, like *began*, requires an event complement, a property not normally expressed by *the book*. Sentences like (4b) used a verb phrase (*was writing*) that explicitly expressed the event interpretation that readers typically ascribed to sentences like (4a), as confirmed by a norming task. Sentences like (4c) used a verb phrase (*was reading*) that expressed an

atypical, low frequency, yet nevertheless plausible event structure, again confirmed by a norming task. Reading times at the word *book* were longer for sentences like (4a) and (4c) than for sentences like (4b), indicating that there were costs associated with processing atypical relationships (4c) and with processing the semantic mismatch between the verb and its complement (4a). Reading times on the following word, *in*, produced a different pattern of results. Here, readers had greater difficulty with sentences like (4a) than with the other two types of sentences. This was argued to reflect on-going difficulty interpreting sentences like (4a), which require readers to generate an event-related interpretation of *the book*. The data suggest that there is a cost to construing an NP like *the book* as an event, over and above the cost associated with interpreting sentences that describe atypical relationships between thematic agents and patients.

The notion that structures requiring an enriched form of composition engender measurable costs has also received support from an investigation of the processing of aspectual relations. Pinango, Zurif, and Jackendoff (1999) used a cross modal lexical decision task to measure the processing load associated with strings like (5):

(5a) *The insect glided effortlessly until...*

(5b) *The insect hopped effortlessly until...*

The verb *glide* denotes a temporally unbounded activity, while the verb *hop* prototypically denotes a point-action activity with an intrinsic beginning and end. The preposition *until*, however, is incompatible with a point-action activity, and is argued to induce a form of coercion that, in this case, type shifts the aspectual form from point-action to repeated activity. Processing load was found to be greater for (5b) as compared to (5a) at the preposition, as hypothesized.

However, apparently not all nominal forms of type shifting induce measurable processing cost. Consider, for example, metonymic expressions like *Vietnam* in *A lot of Americans protested during Vietnam* (Frisson & Pickering, 1999, 2001 cf. Frazier & Rayner, 1990; Pickering & Frisson, 2001). Assuming that the literal interpretation of expressions like *Vietnam* is as a place (i.e., of the semantic type *entity*), the metonymic use involves a type-shift from an entity to an event (i.e., *Vietnam as war*). Frisson and Pickering (1999) used eye-movement monitoring to investigate processing of sentences containing place-for-event shifts (e.g., *protested during Vietnam*). If readers first computed the literal interpretation, then shifted to the metonymic interpretation, they should have had difficulty with sentences that require the latter. Readers experienced difficulty with sentences that had no appropriate literal or metonymic interpretation (e.g., *protested during Finland*), but sentences with appropriate literal and metonymic interpretations were equally easy to process. The same pattern emerged for place-for-institution metonymies (e.g., *answer to the convent*). Frisson and Pickering (1999, 2001) proposed that the fact that literal and familiar metonymic senses of NPs like *Vietnam* are related enabled readers to adopt an *underspecified* interpretation, and to choose between alternative senses during a later "homing-in" stage of processing.

In theory, NPs like *the book* could be treated in the same way. The initial interpretation of *the book* could be underspecified between entity and event interpretations, and then context could be used later to home in on the appropriate reading (e.g., the entity reading for *saw the book*, and the event reading for *began the book*). However, this explanation does not fit with McElree et al.'s (2001) results, which suggested that readers experienced difficulty when the event reading of *the book* was required.

There are several reasons why type shifting might be costly in cases like *began the book* but not in metonymic cases like *protested during Vietnam*. Notably, the type-shifting operations involved in interpreting these latter types of metonymic expressions do not appear to require the generation of additional semantic structure as do cases like *began the book*. A somewhat related reason is that common metonymic senses like *Vietnam as a war* may come to be stored in the lexicon along with the literal sense, whereas it seems unlikely that all relevant event senses of nouns like *book* are precompiled in the lexicon. These issues are considered more fully in General discussion. However, it is important first to determine whether the discrepancy might be due to methodological differences between the studies. One concern is that our prior results might be due to readers adopting unusual strategies. Word-by-word self-paced reading time tends to be considerably longer than reading time under more naturalistic conditions (Just, Carpenter, & Woolley, 1982). Participants may select a particular interpretation, rather than underspecify, when reading is slowed. Alternatively, they might simply delay choice of analysis under such circumstances. Although there is some indication that self-paced reading can produce similar results to

eye-movement monitoring (Rayner, 1998; Rayner & Pollatsek, 1989), the mode of presentation could still affect semantic processing. Experiment 1 therefore examined whether increased reading times associated with sentences like (4a) relative to sentences like (4b) and (4c) occur in an eye-movement monitoring paradigm, which more closely resembles natural reading.

Additionally, one might be concerned that the longer reading times for strings like *began the book* as compared to strings like *read/wrote the book* were due to the fact that verbs like *begin*, *start*, or *finish* are (arguably) less specific than verbs like *read* and *write*, and that the additional processing time reflected the time needed to develop an explicit semantic interpretation. In Experiments 2 and 3, we contrasted strings like *the boy started the fight*... with strings like *the boy started the puzzle*..., in which the same event verb is paired with an event NP or an entity NP. The enriched composition hypothesis holds that a processing disadvantage should occur only when the critical NP mismatches the required semantic form of the verb, and the NP must be type-shifted into a compatible form. For verbs like *begin*, *start*, or *finish*, no processing difficulty is predicted with an event complement like *the fight*. An account that attributes our previous findings to general processing difficulty associated with semantically underspecified verbs like *began* predicts that both strings should be of equal difficulty, others things being equal. Additionally, Experiment 2 used an eye-movement monitoring procedure and Experiment 3 used a self-paced reading procedure, which allowed us to compare two methods of examining semantic processing.

Experiment 1

Experiment 1 tested sentences like (6a–c).

(6a) The secretary began the memo about the new office policy.

(6b) The secretary typed the memo about the new office policy.

(6c) The secretary read the memo about the new office policy.

In sentences like (6a), verbs like *began* require events in the post-verbal argument slot, but the object NP's default interpretation was an entity. In sentences like (6b) and (6c), the verb specifies the activity and the default interpretation of the object NP is compatible with the verb's selectional restrictions. If interpretation of sentences like (6a) requires interpolation of additional semantic structure to type-shift the complement, then localized processing difficulty should occur soon after readers encounter the word *memo*. Processing difficulty might arise because the misanalysis between verb and complement temporarily disrupts processing, because more time is required to generate the additional structure necessary to effect the type shifting, or some combination of the two. If by contrast, readers adopt an underspecification or a delay strategy, then little or no processing difficulty should be observed around *memo*.

Method

Participants. Thirty-three undergraduates at the University of South Carolina took part in the eye-movement-monitoring phase of the study. All of the participants were native English speakers with normal vision and hearing. They participated to fulfill requirements in undergraduate Psychology courses.

Items. We tested 36 triplets like (6a)–(6c) (see Appendix). Sentences like (6a) contain a verb that requires an event or activity as its direct-object argument. The enriched composition hypothesis predicts that sentences such as (6a) will become difficult to process when readers realize that the default interpretation of the direct-object NP does not match the selection restrictions of the preceding verb. This should occur when (or shortly after) readers fixate the direct-object NP. According to the results of a norming pre-test (see below), sentences like (6b) represent the preferred interpretation of sentences such as (6a). Sentences such as (6c) represent a non-preferred but plausible alternative interpretation of sentences such as (6a).

We were constrained in our choice of verbs in the preferred and non-preferred conditions by participants' responses to the preference norming tasks. That is, we wanted to compare reading times from sentences where the verb requires an event (like (6a)) to those from sentences that instantiated the preferred interpretation (like (6b)) and sentences that instantiated an alternative interpretation (like (6c)). Thus, we were forced to replace the event verbs with the verbs that participants indicated represented the preferred interpretation and the next most likely interpretation. This resulted in the event verbs being longer on average than the verbs in the preferred and non-preferred conditions. The average length

was 6.9 characters in the coerced condition, 5.4 characters in the preferred condition, and 5.7 characters in the non-preferred condition [$F(2,70) = 10.4,\ p < .0001,\ MSe = 2.35$]. Verbs in the coerced condition therefore were longer than verbs in the preferred condition [$F(1,35) = 18.4,\ p < .0001,\ MSe = 2.35$] and in the non-preferred condition [$F(1,35) = 12.0,\ p < .001,\ MSe = 2.35$], but the preferred and non-preferred conditions did not differ ($F < 1$). To make sure that a difference in length on the verb did not affect the pattern of reading times on the subsequent target region, we performed a set of subsidiary analyses after removing those items that had the longest verbs in the coerced condition. Removing nine items eliminated the difference in length between conditions [$F(2,52) = 2.23,\ p > .11$]. These subsidiary analyses produced the same pattern of effects as the analyses with the entire set of 36 items (see Results and discussion).

We also assessed the frequency of the verbs between conditions using the Francis and Kucera (1982) corpus. Mean frequency was 308 occurrences per million words for the coerced condition, 266 for the preferred condition, and 471 for the non-preferred condition. A one-way ANOVA on the frequency data revealed no reliable effect of condition (coerced versus preferred versus non-preferred) [$F(2,70) = 1.28$, NS], and no simple effects [coerced versus preferred: $F(1,35) < 1$; coerced versus non-preferred: $F(1,35) = 1.46$, NS; preferred versus non-preferred: $F(1,35) = 2.29,\ p > .14$].

One version of each item was assigned to one of three lists for the eye-movement monitoring phase of the experiment. Items were assigned to lists such that equal numbers of each condition appeared on each list, exactly one version of each item appeared on each list, and so that no participant saw more than one version of any given item. The items were displayed along with 44 filler sentences of various types. At least one filler item intervened between each experimental item. Comprehension questions occurred after ten of the trials. Participants did not receive feedback on their answers. All of the participants in the analyses reported below scored at or above 90% accuracy on the comprehension questions.

Preference norming. To derive preferred and non-preferred sentences, we asked participants to provide one or two-word fill-in-the-blank responses to type-shifting sentences like "The editor finished_____the article" indicating how they would interpret the string. Seventy-nine candidate sentences were randomized and presented to 36 raters. Thirty-six test sentences were selected such that the dominant response occurred more than twice as often as the next most frequent response. The verbs selected for the preferred condition (e.g., *read*) occurred on average 18.7 times (out of 36), ranging from 10 to 35 times. Those selected for the non-preferred condition (e.g., *wrote*) occurred on average 5.6 times, ranging from 0 to 11 times.

Cloze norming. A group of 20 participants wrote sentence-completions for each of the versions of the sentences used in the experiment. The participants were presented with the experimental sentences up to and including the determiner following the verb (e.g., *The secretary began the...*, *The secretary wrote the...*, *The secretary typed the...*). Participants' responses were compared with the actual experimental sentences to assess how predictable the target words were. Cloze proportions (proportion of responses completed with the target words) for the coerced, preferred, and non-preferred conditions were .03, .19, and .14, respectively. A one-way ANOVA with condition (coerced versus preferred versus non-preferred) as a within-items factor produced a reliable effect of condition [$F(1,35) = 10.6,\ p < .001,\ MSe = 0.023$]. The coerced condition produced lower cloze proportions than the preferred [$F(1,35) = 20.1,\ p < .000,\ MSe = 0.023$] or the non-preferred conditions [$F(1,35) = 9.56,\ p < .01,\ MSe = 0.023$], which did not differ from one another [$F(1,35) = 1.95$, NS]. Note, however, that the magnitude of the cloze proportions indicates that the target words generally were unpredictable. In addition, the absolute size of the difference in predictability is much smaller than predictability differences that have previously produced null results in eye-movement studies (Hyönä, 1993). Note further that our sentences were generally low in degree of constraint prior to the target noun. All but four of the sentences had preferred continuations with cloze probabilities less than .6 (with the highest cloze probability being .8). Reading time effects caused by predictability differences tend to occur more strongly in high-constraint sentences (i.e., sentences where the most likely continuation has a cloze value greater than .8; Ehrlich & Rayner, 1981; McConkie & Zola, 1981; Rayner & Well, 1996; Zola, 1984) than in low constraint sentences (where multiple target words tend to benefit from the less constraining context Schwanenflugel & LaCount, 1988; Schwanenflugel & Shoben, 1985). Never-

theless, to assess whether predictable target words produced differences in the eye-movement data between conditions, we conducted a series of subsidiary analysis after removing the four items where any target word had a cloze probability of .6 or greater [the difference in predictability was still significant, with $F(2,62) = 8.69$, $p < .001$, $MSe = .012$]. These subsidiary analyses of the eye-movement data produced a nearly identical set of results as the main analyses (removing these items led to numerically larger F values in all but one of the analyses, and so we do not report them).[1]

Plausibility norming. The experimental sentences were pre-tested for plausibility by a set of 20 raters, who judged each sentence based on how likely they believed the events described by the sentence were (e.g., Pickering & Traxler, 1998; Traxler and Pickering, 1996). Sentences resembling (6a)–(6c) were randomized and presented to the raters. Raters assigned a number from 1 (the event described by the sentence is highly likely) to 5 (highly unlikely) to the sentences in each list. Mean plausibility ratings were 2.2 (coerced), 1.9 (preferred), and 2.1 (non-preferred), all $F < 1$.

Eye-movement-monitoring procedure. A Fourward Technologies Dual-Purkinje Image eye-tracker monitored participants' eye movements while they read sentences like (6a)–(6c). The tracker has angular resolution of 10′ of arc. The tracker monitored only the right eye's gaze location. A PC displayed materials on a VDU 70 cm from participants' eyes. The location of participants' gaze location was sampled every millisecond and the PC software recorded the tracker's output to establish the sequence of eye fixations and their start and finish times. At the beginning of the experiment, the experimenter seated the participant at the eye tracker and used a bite-plate and head rests to minimize head movements. The tracker was then aligned and calibrated before the experiment began. After reading each sentence, the participant pressed a key. After 10 filler sentences, the participant responded to a comprehension question. Participants received feedback on their responses. All of the participants in the analyses reported below scored at 90% accuracy or above on the comprehension questions. Between each trial, a pattern of boxes appeared on the computer screen along with a cursor that indicated the participants' current gaze location. If the tracker was out of alignment, the experimenter recalibrated it before proceeding with the next trial.

Results and discussion

We report five measures of eye-movement data: *first-pass time*, *first-pass regressions*, *second-pass time*, *total time*, and *regression-path time* (sometimes called *go-past*; e.g., Traxler, Bybee, & Picketing, 1997). First-pass time is the sum of all the fixation times beginning with the reader's first fixation in a region until the reader's gaze leaves the region (on one-word scoring regions, first-pass time is equivalent to *gaze duration*; e.g., Rayner & Duffy, 1986). A first-pass regression occurs when the reader's gaze crosses the left edge of the scoring region following a first-pass fixation. Second-pass time includes all of the time spent in a region following first-pass fixations, including time spent in the region after exiting to the left and time spent in the region after exiting to the right. Total time is the sum of all of the fixations within a region. Regression-path time for a region includes all fixation times from the first fixation in a region until the reader fixates to the right of the region. (Unlike most other measures, regression-path time can include time fixating outside the region.) Three scoring regions were examined (see Table 1). The *target* region included the determiner and the noun following the verb. The determiner was included in the scoring region because the reader probably processes the noun while fixating on the determiner, and hence fixation times on the determiner may be affected by characteristics of the noun (Rayner & Pollatsek, 1989). The *verb* region contained the verb of the sentence. As effects sometimes emerge on words following the critical region (Rayner & Pollatsek, 1989), we also examined the *post-target* region, which consisted of the two words immediately following the critical noun.

Verb region. No statistically significant differences were observed in first-pass time, first-pass regressions, or regression-path time in the verb region, suggesting that readers initially had no more difficulty processing verbs like *began* than they had processing the control verbs.

In contrast, a condition effect (coerced versus preferred versus non-preferred) did occur in both second-pass [$F1(2,64) = 8.93$, $p < .01$, $MSe = 2335$; $F2(2,70) = 12.0$, $p < .01$, $MSe = 1862$] and total time data [$F1(2,64) = 5.14$, $p < .01$, $MSe = 4264$; $F2(2,70) = 6.95$, $p < .01$, $MSe = 3550$].

[1] Full results available from the first-named author upon request.

Table 1
Experiment 1: first pass, second pass, total time, first-pass regressions, and regression-path time for Experiment 1 by region

Example item. The worker began/read/wrote the memo to the district managers					
	First pass	Regression path time	First-pass regressions (%)	Second pass	Total time
Scores for the verb region (e.g., *began* versus *read* versus *wrote*)					
Condition					
Coerced	365	382	7	85	444
Preferred	364	370	4	38	397
Non-preferred	367	367	5	47	405
Scores for the target region (e.g., *the memo*)					
Condition					
Coerced	453	476	11	64	516
Preferred	441	456	6	46	483
Non-preferred	439	455	7	49	476
Scores for the post-target region (e.g., *to the*)					
Condition					
Coerced	446	499	7	67	501
Preferred	424	451	4	49	466
Non-preferred	436	454	4	44	472

Simple effects demonstrated that readers spent more time re-reading the verb in the coerced condition than in either the preferred condition [$F1(1,32) = 16.0,\ p < .0001$; $F2(1,35) = 21.1,\ p < .0001$] or the non-preferred condition [$F1(1,32) = 10.1,\ p < .01$; $F2(1,35) = 14.2,\ p < .001$]. The preferred and non-preferred conditions did not differ (both $F < 1$). Total time data produced the same pattern of simple effects [coerced versus preferred: $F1(1,32) = 8.98,\ p < .01$; $F2(1,35) = 12.2,\ p < .001$; coerced versus non-preferred: $F1(1,32) = 6.19, < .02$; $F2(1,35) = 8.32,\ p < .01$]. The preferred and non-preferred conditions did not differ (both $F < 1$).[2] The second-pass time data, in particular, suggest that readers had difficulty later in the sentence in the coerced condition, and that this difficulty led to increased refixations on the verb.

Target region. First-pass time analyses did not demonstrate any differences between conditions in the target region (e.g., *the memo*). A marginal effect of condition emerged in the first-pass regressions analyses [$F1(2,64) = 3.13.,\ p < .05,\ MSe = 64.5$; $F2(2,70) = 2.93,\ p < .10,\ MSe = 66.3$], but not in the regression-path time data [$F1 = 1.60$, NS; $F2 = 1.42$, NS]. Simple effects showed that the coerced condition evoked more first-pass regressions than the preferred condition [$F1(1,32) = 5.81,\ p < .02$; $F2(1,35) = 5.25, p = .03$], and marginally more than the non-preferred condition [$F1(1,32) = 3.20,\ p = .08$; $F2(1,35) = 3.31,\ p = .07$]. The preferred and non-preferred conditions did not differ (both $F < 1$).

Second-pass time data did not demonstrate any differences, but total time data produced a marginal effect of condition [$F1(2,64) = 3.63,\ p < .05,\ MSe = 4444$; $F2(2,70) = 2.83,\ p = .07,\ MSe = 5629$]. Simple effects showed longer total times in the coerced condition than in either the preferred condition [$F1(1,32) = 4.07,\ p < .05$; $F2(1,35) = 3.86,\ p = .05$] or the non-preferred condition [$F1(1,32) = 5.95,\ p < .02$; $F2(1,35) = 4.59,\ p < .05$]. Again, the preferred and non-preferred conditions did not differ (both $F < 1$).

Post-target region. First-pass times data did not differ in the post-target region. In contrast, regression-path times did differ [$F1(2,64) = 5.13,\ p = .01,\ MSe = 4778$; $F2(2,70) = 4.99,\ p = .01,\ MSe = 4749$]. More specifically, the coerced condition differed from both the preferred condition [$F1(1,32) = 8.22,\ p = .01$; $F2(1,35) = 7.49,\ p = .01$] and the non-preferred condition [$F1(1,32) = 7.11,\ p = .01$; $F2(1,35) = 7.48, = .01$]. The first-pass regressions data produced weaker evidence for the same pattern, with an effect of condition that was significant by participants

[2] Imperfect matching for length of the verbs across conditions did not lead to significant first-pass reading time or first-pass regressions effects. Additional analyses conducted on the data after eliminating items with long verbs produced the same pattern of results.

[$F1(2,64) = 3.15,\ p = .05,\ MSe = 30.8$] but not by items [$F2(2,70) = 2.09,\ p > .10,\ MSe = 51.1$]. Simple effects showed a trend toward a difference between the coerced and preferred conditions [$F1(1,32) = 4.09,\ p = .05;\ F2(1,35) = 2.88,\ p = .09$] and between the coerced and non-preferred conditions [$F1(1,32) = 5.27,\ p < .05;\ F2(1,35) = 3.36,\ p = .07$]. These effects (together with the weaker regressions effects on the target word) are entirely consistent with the second-pass time effects on the verb discussed above. They suggest that readers experienced difficulty with the coerced condition soon after they encountered the noun phrase, and that they then tended to refixate the verb.

Second-pass data from the post-target region produced a trend toward a main effect of condition [$F1(2,64) = 3.62,\ p < .05,\ MSe = 1395;\ F2(2,70) = 2.71,\ p = .07,\ MSe = 2016$], with simple effects suggesting that the coerced condition differed from the non-preferred condition [$F1(1,32) = 6.62,\ p < .01;\ F2(1,35) = 4.97,\ p < .05$] and might differ from the preferred condition [$F1(1,32) = 3.89,\ p < .05;\ F2(1,35) = 2.87,\ p < .10$]. The effect of condition for the total reading time data was significant by participants only [$F1(2,64) = 3.05,\ p = .05,\ MSe = 4102;\ F2(2,70) = 2.19,\ p = .12,\ MSe = 5335$].

Overall, the results demonstrate rapid difficulty with the combination of verbs that require an event role and NPs that are typically used as entities rather than events. These results are compatible with previous results from self-paced reading which also indicated that those expressions cause difficulty (McElree et al., 2001). This could be taken as indicating that additional semantic processing took place in the coerced condition relative to the other two conditions, neither of which required enriched processing of the target NP. It could also be interpreted as indicating that the semantic mismatch between verb and complement—the property that presumably triggers the coercion operation—temporarily disrupts processing.

One point of conflict between the current results and those reported in McElree et al. (2001) is the absence of any difficulty associated with the non-preferred condition in the current study. In the McElree et al. study, reading times were longer at the noun for the non-preferred condition than for the preferred condition. The reasons for this discrepancy are not entirely clear. However, it should be noted that different materials were used in the two experiments, and there was a 3:1 difference in the likelihood of the two non-preferred readings in the two sets of materials. In the completion norms (*The editor finished_____the article*), the non-preferred response (*wrote*) was given 5% in the McElree et al. (2001) study as compared to 15% of the time in the current experiment. The stronger preferences in the current materials might have attenuated the differences observed in the McElree et al. (2001) study.

Experiment 2

Experiment 2 was designed to provide further evidence about the semantic operations that occur when readers process verbs like *began* with complements of different semantic types. In Experiment 1 and in McElree et al. (2001), all of the NPs represented entities as opposed to events, and readers had greater difficulty dealing with such NPs when they followed verbs that required event complements. We have suggested that these effects are due to the coercion operation that type-shifts an entity to an event by interpolating additional semantic structure. However, it is possible that verbs like *began* are simply semantically underspecified, and that they produce general processing difficulty regardless of what type of NP follows.

Experiment 2, an eye-movement monitoring experiment, contrasted NPs that denote events (e.g., *the fight*) with nouns that denote entities (e.g., *the puzzle*). Half of the sentences contained NPs that denoted events or activities (*the contest, the recital*, etc.) and half contained NPs that denoted entities (*the letter*, *the sandwich*, etc.). Half of the sentences contained verbs like *began* that require events or activities. The other half contained verbs like *saw* that do not require events or activities and can be used felicitously with either an event or entity NP without a coercion operation. Crossing these two factors produced four conditions, as in (7a)–(7d):

(7a) The boy started the fight after school today. Event verb + event NP.
(7b) The boy saw the fight after school today. Neutral verb + event NP.
(7c) The boy started the puzzle after school today. Event verb + entity NP.
(7d) The boy saw the puzzle after school today. Neutral verb + entity NP.

Experiments 2 (and 3, see below) allow additional hypotheses to be tested. Verbs like *start* may be semantically less specific and may require

more interpretative processing than, by hypothesis, more specific verbs like *see*. If so, we should observe a main effect of verb type (event versus neutral), with readers having greater difficulty processing sentences like (7a) and (7c), than (7b) and (7d). Alternatively, readers might have difficulty assigning event-related interpretations to NPs in general. If so, sentences like (7a)–(7c) should be more difficult to process than sentences like (7d).

The enriched composition hypothesis predicts an interaction of verb type (*started* versus *saw*) and NP type (*fight* versus *puzzle*). The complement *puzzle* requires type-shifting and the generation of additional semantic structure after *started* (e.g., *started solving the puzzle*. . ., but neither operation is required after *saw*. The complementary *fight* matches the semantic requirements of both *started* and *saw*, so no processing difficulty is predicted.

Method

Participants. Thirty-six participants from the same pool as Experiment 1 participated in the eye-movement phase of the experiment under the same terms.

Items. We tested 32 quadruplets of items like (7); see Appendix. The items were configured to cross two factors: whether the verb required an event or activity as its direct object and whether the direct-object NP's default interpretation was as an event or an entity. Items such as (7a) and (7c) contained verbs such as *started* that require entity NPs to be reinterpreted as events. Items such as (7b) and (7d) contained neutral verbs such as *saw* which can be used with entity NPs or event NPs. Items such as (7a) and (7b) contained direct-object NPs that represented an event (e.g., *fight*). Items such as (7c) and (7d) contained direct-object NPs that represented an entity (e.g., *puzzle*).

The verbs and direct-object NPs were matched across conditions for length and frequency on the Francis & Kucera (1982) corpus. The event verbs in conditions (7a) and (7c) averaged 6.75 characters, whereas the neutral verbs in conditions (7b) and (7d) averaged 7.41 characters [$F2(1,31) = 2.90$, $p = .10$, $MSe = 2.37$]. The event verbs averaged 242 appearances per million words, whereas the neutral verbs averaged 203 appearances per million words [$F2(1,31) < 1$]. Hence, any difference in reading times between conditions is not likely to be due to differences in length or frequency of the verbs. The direct-object NPs whose default interpretations were events averaged 6.19 characters, whereas the entity direct-object NPs averaged 5.94 characters [$F2(1,31) = 3.44$, $p > .07$, $MSe = 0.290$]. The event NPs averaged 65.4 appearances per million words, whereas the entity NPs averaged 79.7 appearances [$F2(1,31) < 1$].

Preference norming. We did not collect preference norming data for the items tested in Experiment 2, because the relevant controls did not involve sentences representing preferred or non-preferred interpretations of the coerced condition. Instead, the controls in this experiment involved sentences that expressed an entirely different meaning. For example, *The boy started the puzzle* will most likely be interpreted as *The boy started putting together the puzzle* or *The boy started working on the puzzle*, but here the control was *The boy saw the puzzle* as opposed to a string that expressed either of those coerced readings. Hence, it is not critical that we know precisely what event-related interpretation of *puzzle* readers settle on for *started the puzzle*. It is only critical that we know that they can come up with a sensible interpretation of *started the puzzle* (hence the plausibility pre-test, below).

Cloze norming. Twenty participants from the same pool as the eye-movement experiment provided cloze responses for the test sentences in Experiment 2 under the same instructions as the cloze norming participants from Experiment 1. The cloze data were subjected to a 2 (verb type: event versus neutral) by 2 (NP type: event versus entity NP) ANOVA. This analysis produced only a main effect of verb type (event versus neutral) as sentences with event verbs produced higher cloze proportions than sentences with neutral verbs [.04 versus. 01; $F2(1,31) = 4.22$, $p < .05$, $MSe = 0.006$]. As the cloze values are low, they are unlikely to have a measurable effect on reading time (e.g., Rayner & Well, 1996). Crucially, the contrasts of interest center on different types of NPs following the verbs. The analyses did not show a difference between NP types (.03 versus .02 for event and entity NPs, respectively, $F < 1$), nor an interaction of verb and NP types ($F < 1$). As in the preceding experiment, the target words were not predictable from the context.

Plausibility norming. We also assessed the mean plausibility of the test sentences across the four conditions using the same norming procedure and the same participants as the previous experiment. Sentences with event-related NPs were rated about as plausible as sentences containing entity NPs (2.2 for sentences with event

NPs versus 2.6 for sentences with entity NPs). Likewise, sentences with event verbs were rated about as plausible as sentences with neutral verbs (mean rating of 2.3 versus 2.5, respectively). Individual means for the four conditions were 2.0 for event verb + event NP, 2.4 for neutral verb + event NP, 2.5 for event verb + entity NP, and 2.7 for neutral verb + entity NP. When these plausibility data were subjected to a 2 (verb type: event versus neutral) by 2 (NP type: event versus entity) ANOVA, no main effects or interactions achieved statistical significance (all $p > .10$).

Eye-movement-monitoring procedure. The eye-movement-monitoring procedure was identical to Experiment 1.

Results and discussion

Mean values of the eye-movement measures by condition appear in Table 2. Initially, the data from the three scoring regions were subjected to 2 (verb type: event versus neutral) by 2 (NP type: event versus entity) ANOVAs treating both factors as within-participants and within-items.

Verb region. Data from the verb region in Experiment 2 did not produce any main effects or interactions that were reliable by both participants and items, in contrast with Experiment 1.

Target region. In the target region (e.g., *the fight/the puzzle*), the first-pass time data produced a main effect of NP type [event versus entity; $F1(1,35) = 8.20, p < .01$, $MSe = 4136$; $F2(1,31) = 6.92$, $p < .01$, $MSe = 4673$], as target regions containing event NPs had an average first-pass reading time of 450 ms, whereas the comparable regions containing entity NPs had an average first-pass reading time of 419 ms. This effect also appeared in the regression-path time data, with event nouns having longer regression-path times than entity nouns [$F1(1,35) = 5.88$, $p < .02$, $MSe = 4386$; $F2(1,31) = 4.54$, $p < .05$, $MSe = 5423$]. No other effects occurred in the regression-path time data. These effects most likely reflect some differential difficulty processing the NPs. The nouns in the event conditions were numerically longer (by .2 characters on average) and numerically less frequent than the nouns in the entity conditions (by 15 appearances per million on average). The first-pass times may reflect some small influence of these two factors on early processing. The first-pass regressions data from the target region did not produce any main effects or interactions.

The second-pass times on the target region also produced a main effect of NP-type, but in the opposite direction than the effect that occurred in

Table 2
Experiment 2: first pass, second pass, total time, first-pass regressions, and regression-path time for Experiment 2 by region

Example item. The boy started/saw the fight/puzzle after school today

	First pass	Regression path time	First-pass regressions (%)	Second pass	Total time
Scores for the verb region (e.g., *started* versus *saw*)					
Condition					
Event verb, event NP	339	349	4	74	417
Neutral verb, event NP	367	378	6	84	454
Event verb, entity NP	346	355	3	101	453
Neutral verb, entity NP	362	369	5	87	440
Scores for the target region (e.g., *the fight* vesus the *puzzle*)					
Condition					
Event verb, event NP	452	470	8	67	509
Neutral verb, event NP	448	462	6	84	531
Event verb, entity NP	421	443	7	134	550
Neutral verb, entity NP	417	436	10	96	501
Scores for the post-target region (e.g., *after school*)					
Condition					
Event verb, event NP	417	465	6	89	492
Neutral verb, event NP	422	482	9	93	509
Event verb, entity NP	430	490	9	82	497
Neutral verb, entity NP	425	479	6	79	495

the first-pass time data [$F1(1,35) = 14.4$, $p < .01$, $MSe = 3862$; $F2(1,31) = 5.74$, $p < .05$, $MSe = 8716$]. Sentences with entity NPs generated greater second-pass time in the target region (115 ms) than sentences with event NPs (76 ms). This main effect of NP type must be interpreted in the light of the interaction of verb and NP types [$F1(1,35) = 5.97$, $p < .05$, $MSe = 4396$; $F2(1,31) = 6.32$, $p < .05$, $MSe = 3150$]. Specifically, as predicted, readers experienced the greatest difficulty when entity NPs followed event verbs. This produced a simple main effect of verb type (event versus neutral) within sentences that contained entity NPs [$F1(1,35) = 5.76$, $p = .02$; $F2(1,31) = 5.50$, $p = .03$]. Within sentences that contained event verbs, entity NPs generated much longer second-pass times than event NPs [$F1(1,35) = 18.0, p < .001$; $F2(1,31) = 21.1, p < .0001$]. Hence, entity NPs were processed slower following event verbs than following neutral verbs. Sentences with event NPs did not show the same pattern. If anything, event NPs were processed more rapidly following event verbs than following neutral verbs (although the numerical difference between conditions did not achieve statistical significance). No other main or simple effects approached statistical significance in the second-pass time data from the target region.

Total time data produced a pattern of effects similar to the second-pass time data, except that there was no main effect of NP-type. The interaction of NP and verb type was statistically significant [$F1(1,35) = 5.95$, $p < .05$, $MSe = 7539$; $F2(1,31) = 8.92$, $p < .05$, $MSe = 4192$], as it was in the second-pass time data, and the same pattern of simple effects was obtained. Sentences with entity NPs produced longer total times in the target region when the sentence contained an event verb than when the sentence contained a neutral verb [$F1(1,35) = 5.83, p = .02$; $F2(1,31) = 8.74, p = .01$]. Further, sentences with entity NPs generated longer total times on the target region than sentences containing event NPs when the sentences contained an event verb [$F1(1,35) = 4.03$, $p = .05$; $F2(1,31) = 6.91$, $p = .01$]. No other main or simple effects approached statistical significance.

Post-target region. First-pass regressions data from the post-target region produced a reliable interaction of verb (coercing versus control) and NP types (event versus entity; $F1(1,35) = 4.42$, $p < .05$, $MSe = 66.4$; $F2(1,31) = 4.76$, $p < .05$, $MSe = 77.8$]. Tests for simple effects, however, did not produce any significant contrasts. No main effects or interactions occurred in any of the other dependent measures in the post-target region.

The main finding from Experiment 2 is the interaction of verb (event versus neutral) and NP types (event versus entity) in data from the target region in second pass and total time measures. These interactions rule out an account under which readers experience generalized difficulty semantically processing verbs like *began* and *finished*, or, alternatively, an account under which it is simply more difficult to process an event complement. There was no evidence for difficulty when NPs representing events follow event verbs, or when the verb did not require an event. Instead, the results suggest that readers experience difficulty only when the verb requires an event and the NP's default interpretation does not satisfy this requirement, as predicted by the enriched composition hypothesis.

Somewhat surprisingly, however, this difficulty only emerged on analyses that include late processing, such as second-pass time and total time. This contrasts with Experiment 1 and may have many explanations. One important difference between the experiments concerns the control conditions to which the coerced forms were compared. In the first experiment, the control conditions were selected to reflect one of the underlying interpretations of the coerced form (e.g., *read the memo* or *wrote the memo* for *began the memo*), thereby insuring that the coerced condition involved the same event as the preferred condition (e.g., *read the memo*) or potentially the non-preferred condition (e.g., *wrote the memo*). This was of course not possible in Experiment 2 because of the need to vary both the verb (*started* versus *saw*) and NP (*puzzle* versus *flight*). The tighter relationship between the events in Experiment 1 than Experiment 2 may therefore have allowed us to detect more subtle effects during earlier processing.

Experiment 3

The results of Experiment 2 support an enriched composition hypothesis, despite the fact that the predicted difficulty in processing for sentences requiring coercion emerged later than what was found in prior studies. Because of this discrepancy, however, it is important to verify that the results of Experiment 2 generalize to other processing measures. Experiment 3 involved a

self-paced reading (moving window) paradigm on the same set of materials as Experiment 2 but with a different group of participants.

Method

Participants. Thirty-six participants from the same population as the previous experiments took part under the same terms.

Stimuli. The stimuli were the same as those in Experiment 2.

Self-paced reading procedure. Participants were instructed to read at a normal, comfortable pace in a manner that would enable them to answer comprehension questions. Sentences were presented with a self-paced moving window procedure using a PC running custom made software. Each trial began with a series of dashes on the computer screen in place of the letters in the words. Any punctuation marks appeared in their exact position throughout the trial. The first press of the space bar replaced the first set of dashes with the first word in the sentence. With subsequent space-bar presses, the next set of dashes was replaced by the next word, and the preceding word was replaced by dashes. A yes-or-no question followed each sentence and participants did not receive feedback on their answers. All of the participants in the analyses reported below scored at 90% accuracy or greater on the comprehension questions. The computer recorded the time from when a word was first displayed until the next press of the space bar.

Results and discussion

Mean self-paced reading time by region and condition for Experiment 3 appears in Table 3. Reading times greater than 2000 ms (1% of the data) were excluded from the analyses. Here, we report results for the target word (*fight* versus *puzzle*), for comparison to Experiment 2, and the following word. No statistically significant effects occurred on any other words.

The data from the two scoring regions were subjected to 2 (verb type: event versus neutral) by 2 (NP type: event versus entity) ANOVAs that treated both factors as within-participants and within-items. No significant main effects or interactions occurred at the target word. A main effect of NP type (event versus entity) occurred on the word following the target word [$F1(1,35) = 13.3,\ p < .01,\ MSe = 6400$; $F2(1,31) = 10.3,\ p < .01,\ MSe = 6475$], but that effect was qualified by an interaction of NP and verb type (event versus neutral; $F1(1,35) = 9.63,\ p < .01,\ MSe = 5079$; $F2(1,31) = 4.12,\ p < .05,\ MSe = 5073$]. The interaction occurred because sentences with entity NPs and event verbs generated longer reading times than sentences with entity NPs and neutral verbs [$F1(1,35) = 7.25,\ p < .01,\ MSe = 5079$; $F2(1,31) = 3.07,\ p = .09,\ MSe = 5073$]. Sentences with event verbs and event NPs produced numerically shorter reading times than sentences with event NPs and neutral verbs, but that difference was not significant [$F1(1,35) = 2.88,\ p = .10,\ MSe = 5079$; $F2(1,31) = 1.25$, NS, $MSe = 5073$].

As with Experiment 2, these data are not compatible with an account under which event interpretations of NPs are in general difficult to compute. If this were the case, sentences like (7a), (7b), and (7c) should have been more difficult to process than (7d). Additionally, the data are not compatible with an account under which event verbs like *started* are more difficult than (arguably) more specific verbs like *see*. If this were the case, then strings like *started the fight* (7a) should have been just as difficult to process as strings like *started the puzzle* (7c) (when compared to *saw the fight*/*saw the puzzle*), but in fact *started the puzzle* was harder to process than *started the fight*. Finally, the results of this experiment provide

Table 3
Experiment 3: mean self-paced reading time by condition and region

Example item. The boy started/saw the fight/puzzle after school today		
	Target word (e.g., *fight*/*puzzle*)	Next word
Condition (verb-complement)		
Event, event	470	427
Neutral, event	497	455
Event, entity	467	512
Neutral, entity	486	467

further evidence that self-paced reading and eye-movement monitoring can lead to similar results for semantic processing.

General discussion

In Experiment 1, participants read expressions like *began the memo*, In which an NP that typically refers to an entity was preceded by a verb that requires an event, and expressions like *read the memo* and *typed the memo* that conveyed the preferred and non-preferred interpretations of the expression *began the memo*. The enriched composition hypothesis predicts that the mismatching semantic properties in cases like *began the memo* trigger a coercion operation that type shifts the NP into a compatible type by introducing additional structure that is semantically equivalent to an explicit expression like *began reading the memo*. Tests for simple effects on first-pass regressions and total time data from the critical NP showed that the coerced condition was more difficult to process than the preferred condition, although the difference between the coerced and non-preferred conditions was marginal. Second pass and total time data from the verb produced robust differences between conditions, providing further evidence that the coerced condition was more difficult than both the preferred and the non-preferred conditions.

Experiments 2 and 3 examined whether this effect was due to a general difficulty associated with interpreting what might be argued to be semantically underspecified verbs like *begin*, or whether it might to due to a general difficulty associated with interpreting event-denoting NPs. Neither factor affected processing difficulty; rather we found a reliable interaction between verb (event versus neutral) and NP (event versus entity) types in second pass and total time data from the critical NP (Experiment 2) and in self-paced reading times one word after the critical NP (Experiment 3). These interactions indicated that NPs that typically denote entities were more difficult to process than NPs that typically denote events when they followed verbs that require event complements. Event and entity NPs were equally easy to process following neutral verbs. Thus, in all three experiments, entity-denoting NPs were difficult to process following event verbs only. Moreover, the similarity of patterns across different procedures alleviates concerns that the results might be due to experimental artifacts. We argue that these studies, together with the McElree et al. (2001) and Pinango et al. (1999), provide strong support for the enriched composition hypothesis (Jackendoff, 1997; Pustejovsky, 1991, 1995).

One alternative account of the difficulty with strings like *began the memo* is simply that they are semantically ambiguous (e.g., between reading and typing the memo), and that ambiguity in itself leads to difficulty. For example, lexically ambiguous words can lead to difficulty in comparison to unambiguous words (e.g., Duffy, Morris, & Rayner, 1988; Rayner & Duffy, 1986). Presumably, readers are at some level comparing alternative interpretations, and this process of comparison is in itself costly. However, such difficulty only emerges for "balanced" words with two meanings of roughly comparable frequency (or biased words where context supports the non-preferred meaning). In our cases, one interpretation was strongly preferred. Additionally, this difficulty seems to be limited to lexical ambiguity. In balanced syntactic ambiguities, when two analyses are equally likely for a given sentence, processing difficulty is actually decreased relative to unambiguous controls (Traxler et al., 1998; Van Compel et al., 2001). For both these reasons, it is very unlikely that processing difficulty with strings like *began the memo* is due to ambiguity per se.

Formal work in lexical–semantics has argued that expressions like *began the book* are interpreted by coercing the complement NP in to the event type required by the verb (e.g., Jackendoff, 1997; Pustejovsky, 1991, 1995). That the verb *begin* selects for an event is evident from the fact that it can explicitly occur with complements that are marked in syntax as an activity or event (Levin, 1993; Quirk, Greenbaum, Leech, & Svartvik, 1985), such as a participial ("-ing" or gerundive) clause (e.g., *began reading/writing the book*) or an infinitival clause (e.g, *began to read/write the book*). The coercion operation, which is assumed to be triggered by the mismatching semantic properties of the verb and NP, is argued to convert expressions like *the book* from the usual semantic type *entity* to the type *event* by interpolating additional structure analogous to overt expressions like *began to write the book* or *began writing the book*. In Pustejovsky's (1991, 1995) treatment, the added structure is argued to be semantic in nature, although in alternative accounts the coercion operation also generates covert syntactic structure (cf. Grimshaw, 1979; Fodor & Lepore, 1998).

This process is thought to draw upon richly structured lexical representations associated with the NP, as well as on information in the discourse context. In Pustejovsky's (1991, 1995) treatment, the lexical representation for nouns like *book* partly consists of a Qualia Structure, which codes attributes of the entity, such as its constituent parts, telic properties (purpose and function), and agentive properties (mode of creation), among others. Qualia are assumed to be relational structures with variables that are bound during interpretation. For example, the telic representation for *book* is a two-argument event structure headed by the verb *read*; its agentive representation is also a two-argument event structure, but one that is headed by the verb *write*. An NP like *the book* can be readily type-shifted from entity to event by using the event structure in either the telic or agentive qualia to restructure the predicate into a form that is semantically equivalent to an explicit predicate like *began to read/write the book* (see Pustejovsky, 1995 for a formal treatment of the operators and representations). Qualia representations are intended to capture default interpretations for expressions like *began the book*. In the absence of other information, an expression like *began the book* is likely to be interpreted using the telic or agentive properties most commonly associated with the NP (here, *reading* and *writing*, respectively). Which property is selected clearly depends on other constituents in the sentence and in the discourse (consider, for example, likely interpretations for *The author/student/copywriter/designer began the book*). Properties not coded in the lexical representation or interpretations that require a lexically specified property to be overridden by pragmatic information in the discourse context (e.g., *the goat began the book*) are assumed to be generated by set of distinct but related coercion operations that utilize discourse structures instead of lexical structures (e.g., Asher & Pustejovsky, 2001; Lascarides & Copestake, 1998).

Our experiments do not of course provide unambiguous support for all aspects of Pustejovsky's (1995) account. For example, the experiments do not provide direct evidence for the specific lexical representations (qualia representations) that are posited. The putative qualia structure does give a principled explanation for our findings that mismatching semantic types between the verb and complement induce processing difficulty. However, there are three potential sources for this increase in processing cost. The costs might be associated with: (1) the detection of an anomalous relationship between the verb and its complement, (2) the time to type shift the complement from one semantic type to another, and/or (3) the time required to generate the additional structure necessary to effect the type shift. However, we believe that our findings are best attributed to the third alternative.

It is possible that the observed processing difficulty simply reflects the detection of an anomalous relationship between verb and complement rather than the cost of generating new structure. Disruption in processing does occur when a computed interpretation does not make sense (e.g., Clifton, 1993; Picketing and Traxler, 1998; Speer & Clifton, 1998; Traxler & Picketing, 1996). On this account, the difficulty that occurred following the target word would occur because readers access the default meaning of the target word and attempt to integrate this meaning with the sentence context. This default interpretation would produce an anomalous result, triggering semantic reanalysis. This account is attractive in that no special operations are presupposed, but it leaves unanswered how readers recover a viable interpretation of expressions like *began the book*. Moreover, this account is less compatible with the particular effects that we found. The differences in Experiments 2 and 3 emerged one word after the target word (Experiment 3) and in measures typically considered to reflect integrative processing in eye-movement data (Experiment 2). Although it is possible to interpret the patterns we observed as arising from the detection of a mismatch, to do so would require postulating substantial delays in composition, which would be inconsistent with the common view that interpretation is rapid (e.g., Marslen-Wilson, 1975; Marslen-Wilson & Tyler, 1980; Traxler & Picketing, 1996).

One reason for assuming that it is not type shifting per se that causes difficulty is that the interpretation of metonymic expressions like *A lot of Americans protested during Vietnam* does not appear to engender analogous processing difficulty, despite the fact that their interpretation also involves shifting the interpretation of an NP from an entity (a place) to an event (Frisson & Pickering, 1999). As noted, Frisson & Pickering (1999, 2001) argued that readers were able to adopt an *underspecified* interpretation, and to choose between alternative senses during a later *homing-in* stage of processing. The difference between the literal and metonymic interpretation of *Vietnam* appears to be primarily, if not exclusively, one of reference (does the expression refer

to a place or an event?) Underspecification may be possible because reference can be delayed. The type shifting of *the book* from entity to event also entails a shift in reference for the complement, but the required event interpretation must be constructed rather than simply selected from the lexicon. However, more empirical work is needed to directly contrast under more tightly controlled situations the interpretation of metonymic expressions like *protested during Vietnam* and coerced expressions like *began the book*.

It would in principle be possible to store frequent coerced structures in the lexicon, as might be the case for the quite common "war" sense of *Vietnam*. But for this to hold with examples like *began the memo*, it would be necessary to store interpretations that are specific to particular verbs (*began*, *finished*, *enjoyed*, etc.), which seems uneconomical if not implausible. Pustejovsky (1995) takes the less extreme and more flexible position that properties often used in coercion operations (e.g., telic and agentive structures in the qualia representation) come to be stored in the lexical representation for the noun, but that specific interpretations required by a verb (as well as other elements in the discourse) are computed on-line. Our experiments provide support for the assumption of additional on-line computations, and are therefore compatible with a generative approach to semantic composition.

Although our results are fully consistent with the computational machinery outlined in Pustejovsky (1995), they do not, of course, exclude other approaches. However, we believe that alternative approaches will need to draw analogous distinctions to those outlined above. For example, Fodor & Lepore (1998) reject the generative mechanisms that Pustejovsky (1995) localizes in the lexicon (but see Pustejovsky, 1998). They argue for an approach in which lexical representations are atomistic and lexical meaning is identified with denotation exclusively. In this approach, the semantic content that a lexical item contributes to a phrasal interpretation is argued to be context invariant. The differences documented here, which we attribute to generative lexical operations, must be viewed as arising from other aspects of interpretive processing. The obvious candidate for the Fodor & Lepore (1998) approach is a difference in the difficulty of finding references for coerced versus simple expressions. However, we remain skeptical of whether such an approach is fully compatible with our results and with results such as those found with metonymic expressions (Frisson and Picketing, 1999). Such an approach would seem to predict that both types of expression should induce processing difficulties.

We believe that our results firmly establish that expressions like *began the book* require an enriched form of semantic composition, and that the most consistent account of that enriched processing is one that assumes the generation of additional semantic structure. What remains an open issue is the degree to which this type of enriched processing can be viewed as a constrained set of composition operations acting on highly structured lexical representations, as proposed by Pustejovsky (1991, 1995) and others like Jackendoff (1997). A more flexible approach might be to assume a general purpose reasoning system drawing on world knowledge of the form typically thought to mediate inferences in discourse processing. Formal arguments against subsuming the former under the latter are that composition and discourse constraints appear to interact in a principled manner (see Asher & Pustejovsky, 2001; Lascarides & Copestake, 1998). It remains to be determined whether these formally distinct mechanisms can be dissociated on behavioral grounds.

Appendix: Stimuli from Experiments 1–3

Experiment 1

The first version of each sentence represents the coerced condition, the second represents the preferred condition, and the third represents the non-preferred condition.

The engineer started/read/wrote the memo to send to the employees.
The girl tried/ate/tasted the soup while she was visiting friends.
The secretary began/wrote/typed the memo about the new office policy.
The editor finished/read/wrote the article before going home to dinner.
The architect finished/designed/drew the house and met with the contractor.
The stylist started/braided/made the braid after brushing the girl's hair.
The interior designer began/designed/decorated the kitchen before the family moved in.
The editor finished/edited/read the newspaper first thing in the morning.
The publisher began/read/published the novel written by Mark Twain's son.
The student tried/wrote/read the papers that were assigned for class.

The critic started/criticized/analyzed the portrait that was modeled after Picasso.
The guard finished/closed/locked the gates before going home at night.
The woman started/planted/weeded the garden after the last winter frost.
The farmer started/plowed/planted the fields in the early spring months.
The waitress started/made/poured the coffee when the customers walked in.
The director started/read/wrote the script after hearing about the story.
The banker started/drank/made the coffee since he was getting sleepy.
The teacher started/recorded/corrected the grades before report cards went out.
The professor finished/wrote/typed the syllabus before writing up his lectures.
The lawyer preferred/drove/parked the convertible with the fine leather seats.
The publisher started/read/wrote the manuscript then gave it to Tom.
The lawyer endured/defended/questioned the defendant but thought he was guilty
The doctor began/wrote/filled the prescription for the new cold medicine.
The auditor began/audited/did the taxes and finished by early April.
The surfer endured/wore/bought the tuxedo but did not like it.
The nurse preferred/wore/felt the velvet that was made in India.
The child began/wrote/read the letter for Santa Claus before Christmas.
The pilot preferred/flew/landed the biplane because it was very sturdy.
The journalist began/wrote/read the article about the floods and hurricanes.
The builder started/built/designed the house that his family lived in.
The mechanic finished/repaired/waxed the truck and then fixed the car.
The dieter resisted/ate/tasted the cake and had some baby carrots.
The teenager began/read/wrote the novel about the vampires and ghosts.
The student finished/read/wrote the book about learning how to sail.
The robber attempted/stole/took the necklace while visiting the art museum.
The pilot mastered/flew/landed the plane but nearly crashed at take-off.

Experiments 2 and 3

For each item, the verb to the left of the first "/" mark is the event verb. The verb to the right of the "/" mark is the neutral verb. The noun to the left of the second "/" mark is the noun used in the event conditions. The noun to the right of the second "/" mark is the noun used in the entity conditions.

The boy started/saw the fight/puzzle after school today.
The soldier began/observed the battle/trench early in the morning.
The catcher enjoyed/cursed the game/coach, the reporter told us.
The builder completed/recalled the contest/mansion, David said to Joan.
The track star endured/watched the marathon/trainer despite the intense pain.
The audience member preferred/praised the speech/dress by the Italian designer.
The debater started/forgot the argument/article about health care spending.
The banker expected/remembered the audit/check after the police investigation.
The singer began/praised the concert/letter in the auditorium near the university.
The pianist ended/ignored the recital/letter in the classroom on the fourth floor.
The minister finished/prepared the funeral/sandwich before returning to the chapel.
The psychologist enjoyed/described the session/dessert with the patient.
The victim endured/reported the robbery/driver after the bar closed.
The cheating husband began/recalled the affair/novel only last week.
The housewife enjoyed/watched the show/kids on her birthday.
The student ended/outlined the visit/paper and returned to her dorm room.
The schoolboy started/saw the fight/puzzle with his little brother.
The knights began/observed the battle/trench outside the castle.
The athlete enjoyed/cursed the game/coach yesterday at the stadium.
The designer completed/recalled the contest/mansion during the summer.
The inexperienced athlete endured/watched the marathon/trainer for an hour.
The teacher preferred/praised the speech/dress that the girl made.
The politician started/forgot the argument/article that upset the voters.
The accountant expected/remembered the audit/check before the end of the month.
The soprano began/praised the concert/letter in the theater near the train station.
The lazy guy ended/ignored the recital/letter because he was tired.
The pastor finished/prepared the funeral/sandwich before the family left.
The schizophrenic patient enjoyed/described the session/dessert at the hospital.

The banker endured/reported the robbery/driver just last week.
The writer began/recollected the affair/novel at the tropical resort.
The stressed businessman enjoyed/watched the show/kids on Saturday afternoon.
The teacher ended/approved the visit/paper before recording the grades.

References

Asher, N., & Pustejovsky, J. (2001). The metaphysics of words in context. Submitted.

Bach, E. (1986). The algebra of events. *Linguistics and Philosophy, 9*, 5–16.

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language, 27*, 429–446.

Clifton, C., Jr. (1993). Thematic roles in sentence parsing. *Canadian Journal of Experimental Psychology, 47*, 222–246.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word recognition and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior, 20*, 641–655.

Fodor, J. A., & Lepore, E. (1998). The emptiness of the lexicon: Reflections on James Pustejovsky's "The Generative Lexicon". *Linguistic Inquiry, 29*, 269–288.

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

Frazier, L., & Rayner, K. (1990). Taking on semantic commitments: Processing multiple meanings vs. multiple senses. *Journal of Memory and Language, 29*, 181–200.

Frisson, S., & Pickering, M. J. (1999). The processing of metonymy: Evidence from eye-movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1366–1383.

Frisson, S., & Pickering, M. J. (2001). Obtaining a figurative interpretation of a word: Support for underspecification. *Metaphor & Symbol, 16*, 149–171.

Grimshaw, J. (1979). Complement selection and the lexicon. *Linguistic Inquiry, 10*, 279–326.

Hyönä, J. (1993). Effects of thematic and lexical priming on readers' eye movements. *Scandinavian Journal of Psychology, 34*, 293–304.

Jackendoff, R. (1997). *The architecture of the language faculty*. Cambridge, MA: MIT Press.

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General, 111*, 228–238.

Lascarides, A., & Copestake, A. (1998). Pragmatics and word meaning. *Journal of Linguistics, 34*, 387–414.

Levin, B. (1993). *English verb classes and alternations*. Chicago, IL: University of Chicago Press.

Marslen-Wilson, W. D. (1975). Sentence perception as an interactive parallel process. *Science, 189*, 226–228.

Marslen-Wilson, W. D., & Tyler, L. (1980). The temporal structure of spoken language comprehension. *Cognition*, 1–72.

McConkie, G. W., & Zola, D. (1981). Language constraints and the functional stimulus in reading. In A. M. Lesgold, & C. A. Perfetti (Eds.), *Interactive processes in reading*. Hillsdale, NJ: Erlbaum.

McElree, B. D., Traxler, M. J., Picketing, M. J., Jackendoff, R., & Seely, R. E. (2001). Coercion in on-line semantic processing. *Cognition, 78*, B17–B25.

Mitchell, D. (1994). Sentence parsing. In M. A. Gemsbacher (Ed.), *Handbook of Psycholinguistics*. San Diego: Academic Press.

Partee, B. (1992). Syntactic categories and semantic type. In M. Rosner, & R. Johnson (Eds.), *Computational linguistics and formal semantics*. Cambridge: Cambridge University Press.

Pickering, M. J., & Frisson, S. P. (2001). Processing ambiguous verbs: Evidence from eye-movements.. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 556–573.

Pickering, M. J., & Traxler, M. J. (1998). Plausibility and recovery from garden paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 940–961.

Pinango, M. M., Zurif, E., & Jackendoff, R. (1999). Real-time processing implications of enriched composition at the syntax-semantics interface. *Journal of Psycholinguistic Research, 28*, 395–414.

Pustejovsky, J. (1991). The syntax of event structure. *Cognition, 41*, 47–81.

Pustejovsky, J. (1995). *The generative lexicon*. Cambridge, MA: MIT Press.

Pustejovsky, J. (1998). Generativity and explanation in semantics. A reply to Fodor and Lepore. *Linguistic Inquiry, 29*, 289–311.

Pustejovsky, J., & Bouillon, P. (1995). Aspectual coercion and logical polysemy. *Journal of Semantics, 12*, 133–162.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. Essex, England: Longman Press.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372–422.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, 14,191–201.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ: Prentice Hall.

Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review, 3*, 504–509.

Schwanenflugel, P. J., & LaCount, K. L. (1988). Semantic relatedness and scope of facilitation for

upcoming words in sentences. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 14*, 344–354.

Schwanenflugel, P. J., & Shoben, E. J. (1985). The influence of sentence constraint on the scope of facilitation for upcoming words. *Journal of Memory and Language, 24*, 232–252.

Speer, S. R., & Clifton, C., Jr. (1998). Plausibility and argument structure in sentence comprehension. *Memory & Cognition, 26*, 965–978.

Todorova, M., Straub, K., Badecker, W., & Frank, R. (2000). Aspectual coercion and the on-line computation of sentential aspect. In *Proceedings of the twenty-second annual conference of the Cognitive Science Society, Philadelphia, PA*.

Traxler, M. J., Bybee, M. D., & Picketing, M. J. (1997). Influence of connectives on language comprehension: Eye-tracking evidence for incremental interpretation. *Quarterly Journal of Experimental Psychology A, 50*, 481–497.

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

Zola, D. (1984). Redundancy and word perception during reading. *Perception & Psychophysics, 36*, 277–284.