

The difficulty of coercion: A response to de Almeida

Martin J. Pickering^{a,*}, Brian McElree^b, Matthew J. Traxler^c

^a *University of Edinburgh, UK*

^b *New York University, USA*

^c *University of California, Davis, USA*

Accepted 15 July 2004

Available online 11 September 2004

Abstract

The sentence *The secretary began the memo* requires specifying what event the secretary began, because *the memo* does not refer to an event. McElree, Traxler, Pickering, Seely, and Jackendoff (2001) and Traxler, Pickering, and McElree (2002) found evidence from both self-paced reading and eye-tracking that such sentences caused processing difficulty, and thus argued that people “coerced” the object to refer to an event (e.g., writing the memo). de Almeida (2004) reports two self-paced reading experiments that failed to replicate some aspects of previous studies, and thereby argued against coercion during comprehension. A new experiment demonstrates coercion costs using new items, and provides evidence of coercion cost with de Almeida’s stimuli. We conclude that coercion does cause processing difficulty.

© 2004 Elsevier Inc. All rights reserved.

Keywords: Coercion; Semantics; Sentence processing

1. Introduction

A sentence like *John began the book* requires the reader to determine what John began doing to the book. Formally, the verb *began* “coerces” the interpretation of *the book* into an event related meaning, such as “reading the book” or “writing the book” (Jackendoff, 1997; Pustejovsky, 1995). In two papers, we documented a “coercion” effect, viz., that such coercion engenders processing difficulty during reading (McElree, Traxler, Pickering, Seely, & Jackendoff, 2001; Traxler, Pickering, & McElree, 2002). In a recent paper published in this journal, de Almeida (2004) reports little evidence that coercion produced difficulty with comparable sentences. These null findings are used to question the robustness of the coercion effects reported by McElree et al. (2001) and Traxler et al. (2002), and as a vehicle to chal-

lenge decompositional approaches to lexical semantics (e.g., Pustejovsky, 1995) and to champion atomistic approaches that seek to interpret coercion effects as an instance of a meaning postulate or rule of inference (Fodor & Lepore, 1998; but see Pustejovsky, 1998).

Our primary intention in this paper is to firmly establish the reliability of the coercion effect. To this end, we report an eye-tracking experiment using the stimuli from de Almeida’s (2004) Experiment 1 and new set of stimuli that demonstrate clear and reliable coercion effects, comparable to our other published findings. Collectively, these findings are consistent with aspects of the decompositional framework proposed by Pustejovsky (1995), but, as noted in McElree et al. (2001), they “do not, of course, motivate an approach like Pustejovsky (1995) to the exclusion of other approaches” (p. B23). However, in contrast to de Almeida, we do not believe that time-course data can be straightforwardly used to distinguish decompositional and atomistic accounts of coercion phenomena.

* Corresponding author. Fax: +44 131 650 3461.

E-mail address: martin.pickering@ed.ac.uk (M.J. Pickering).

1.1. Enriched composition and processing difficulty

Verbs like *begin*, *finish*, and *enjoy* appear to require a complement that refers to an event (or activity), as in *began to swim* or *enjoyed playing tennis*. Interestingly, they do not need a verbal complement but can take a noun phrase object. This object can refer to an event, as in *began a fight* or *enjoyed dinner*, but it need not. When the object does not refer to an event, as in *began the book* or *enjoyed the cupcakes*, there appears to be “clash” between the requirements of the verb and the properties of the noun. But rather than regard the expression as anomalous, people interpret the complements as events. For instance, *began the book* can mean “began to read the book” or “began to write the book,” and *enjoyed the cupcakes* can mean “enjoyed eating the cupcakes.”

How does this process come about? When the verb requires a complement that refers to an event, and the noun phrase refers to an entity, its semantic type is coerced into an interpretation that changes from entity to event, via a process often termed “type-shifting” (e.g., Bach, 1986). In *began the book*, the complement is shifted from referring to an entity to referring to some activity associated with that entity. The event is typically one that is strongly associated with the object (e.g., reading the book or writing the book), though it need not be (e.g., *the goat began the book* can mean the goat began eating the book). Unlike standard composition between verbs and their arguments, which involve the combination of lexical representations according to their positions in syntactic structure, this process is sometimes called “enriched composition,” as it involves the interpolation of new semantic material that is not contained in the expression (Jackendoff, 1997; Lascarides & Copestake, 1998; Pustejovsky, 1991, 1995). Hence, when the semantic requirements of the verb and the semantic properties of the noun phrase clash, the noun phrase has to be type-shifted to refer to an event, and the specific event not contained in the expression has to be selected.

Given the complexity of enriched composition, we might expect it to incur a processing cost. To investigate this, McElree et al. (2001) presented participants with sentences like (1a–c):

- (1a) The secretary began the memo before the annual sales conference. (coerced)
- (1b) The secretary typed the memo before the annual sales conference. (preferred)
- (1c) The secretary read the memo before the annual sales conference. (non-preferred)

Sentences like (1b) used a verb (*typed*) that explicitly expressed the event interpretation that raters typically (most often) ascribed to sentences like (1a). Sentences like (1c) used a verb phrase (*read*) that raters less typically ascribed to sentences like (1a). However, all three sentences were

rated as equally plausible. Using word-by-word self-paced reading, McElree et al. found that participants experienced more difficulty with coerced sentences than the non-coerced controls. Specifically, participants took longer reading *memo* in (1a) and (1c) than (1b), and longer reading *before* in (1a) than either (1b) or (1c). Therefore there was some localized difficulty with non-preferred sentences in comparison to preferred sentences, but a longer-lasting difficulty with coerced sentences.

Traxler et al. (2002) reported three studies that confirmed the difficulty with coerced sentences. Experiment 1 replicated McElree et al.’s (2001) experimental design using an eye-tracking procedure. This time, no differences emerged between the preferred and non-preferred conditions. In contrast, the coerced sentences were more difficult than their non-coerced controls. Eye-tracking allowed us to investigate the difficulty reading particular regions of the sentence when they are first encountered and during subsequent processing (in terms of time spent reading the region or likelihood of regressing from that region). The first reliable evidence of difficulty emerged when participants first read the two words following the noun phrase (*before the*). However, difficulty also emerged on measures of subsequent processing for various regions. Collectively, measures of total time spent on a region demonstrated substantial and reliable effects of coercion, not, as de Almeida (2004) suggests “only marginally longer reading times” (p. 252).

Experiment 2 considered the processing of event and neutral verbs in the context of noun phrases that refer either to an entity [as in (1)] or an event (e.g., *the fight*). If the source of difficulty is the act of combining an event verb with an entity noun phrase, then *started the puzzle* should cause difficulty like (1a), but *started the fight* should not, because *the fight* refers to an event. Intuitively, there is no need to “insert” an event in *started the fight*, because the fight is the event. Likewise, there should be no difficulty with either *saw the puzzle* or *saw the fight*. In this experiment, the critical prediction is an interaction, with *started the puzzle* causing specific difficulty. The interaction first emerged in the two words following the noun phrase, measured in terms of the likelihood of regressing from that region before moving on. de Almeida (2004) downplayed this result, stating that effects occurred on “analyses that reflected later interpretative processes” (p. 252). However, he apparently overlooked the fact that this effect occurred on first-pass regressions (see Traxler et al., 2002, p. 540). A similar interaction emerged on measures of subsequent processing on the noun phrase. These results demonstrate that the cause of difficulty in (1a) could not be the verb *started*, but rather had to be the process of combining *started* with its object. Experiment 3 replicated Experiment 2 using self-paced reading, and found the same interaction on the word following *puzzle*. (de Almeida does not mention this experiment.)

de Almeida (2004) assumes that effects are generally delayed beyond what would be expected if they were associated with “a mandatory process of type-shifting” and instead argues that they might be due to late inferential processes. But the precise locus of effects often varies between eye-tracking experiments for comparable sentences. Furthermore, there is much controversy about the extent to which eye-tracking measures map onto particular stages of processing (e.g., Rayner, 1998), so his inferences from time-course to underlying cause are unfounded and dangerous. Therefore, although there are slight differences in our four experiments relating to the precise point at which the cost emerges, the important point is that they demonstrate a very consistent pattern: Sentences requiring coercion are harder than comparable sentences that do not require coercion.

2. de Almeida’s findings

de Almeida (2004) challenged our account of enriched composition. His challenge is largely based on two self-paced reading experiments, which is very curious given the importance of the time-course of effects to his arguments. Experiment 1 used three conditions similar to (1), but with items modified from McElree et al. (2001). It produced largely null results, with the only reliable difference among the conditions being a late tendency for the coerced condition to be especially easy. Experiment 2 employed the same sentences preceded by a short context (typically two sentences). It yielded results that were similar to those reported in McElree et al. (2001), in that the coerced condition showed evidence of processing difficulty in the coerced as compared to the control condition. The results differed from the pattern in McElree et al. (2001) in that the level of difficulty for non-preferred condition was intermediate, a pattern which, as we point out in Section 4, is uninterpretable. On the basis of these results he proposes that coercion does not cause processing difficulty.

Our main concern is with Experiment 1, because it produced results that were at odds with all four experiments reported in McElree et al. (2001) and Traxler et al. (2002). Why the discrepancy? It is possible that our results were erroneous, either because of a flaw in our items or procedure, or because of simple Type I error. The latter is extremely unlikely, because our results were replicated four times (with only minor differences between the patterns of results). The experiments used standard eye-tracking and (word-by-word, moving window) self-paced reading, which are the principal techniques employed in sentence-comprehension research. The items might have been flawed in some way, but we carefully matched or controlled for length and frequency of the verb, preferred interpretation of

the coerced sentences, and plausibility of the sentences in the different conditions. Furthermore, we used substantially different items in different experiments.

Alternatively, de Almeida’s (2004) null results may have been erroneous. A likely explanation is a simple Type II Error, caused partly by the comparatively small number of items. de Almeida’s Experiment 1 had 21 items (though 69 participants). In contrast, McElree et al. (2001) had 33 items (with 51 participants), and Traxler et al. (2002) had 36 items (with 33 participants) in Experiment 1, and 32 items (with 36 participants) in Experiments 2 and 3. Even if the large number of participants countered the small number of items, de Almeida may simply have been “unlucky.” It is worth noting that in de Almeida’s Experiment 1 the noun phrase (e.g., *the memo*) was 22 ms longer in the type-shifted condition than the non-preferred condition, and 27 ms longer than the preferred condition. These differences are actually numerically larger than the significant effects in McElree et al. found on the noun (and the next word). We cannot explain de Almeida’s reverse effect downstream, but neither can he, and it is most likely a Type I Error.

It is also possible that his results reflect a problem with his items or procedure. Obviously a small number of items increases the chances of idiosyncrasies affecting the results. Perhaps more importantly, de Almeida (2004) modified McElree et al.’s (2001) items but did not norm them. The coercing verbs in those experiments were about a character longer on average than the other conditions [coerced: 6.92 vs. preferred: 5.62 vs. non-preferred: 6.05, $F(2,40) = 4.72$, $p = .01$], but they also tended to be more frequent in the Francis and Kucera (1982) norms [coerced: lemma frequency 319 vs. preferred: 228 vs. non-preferred: 174, $F(2,40) = 4.57$, $p = .02$]. It is possible that the high frequency of the coercing verbs reduced the difficulty of the coercing condition in both experiments. Additionally, de Almeida did not norm for plausibility. Although self-paced reading is a fairly standard measure, some of his presentation segments were words and some were phrases, and it is conceivable that the fact that the critical noun phrase was presented as a whole led to some surprise (the previous segment was a single word). More worryingly, he trimmed or removed 17.2% of observations, which is extremely high for self-paced reading.

The discrepancies might also be the result of the modifications de Almeida (2004) made to the items. First, he attempted “to obtain a more semantically congruent relation between the agent ... and the object of the verb” (p. 254). It is hard to determine how successful this was, but it may have removed some difficulty with the sentences. Why this should remove the difficulty with the coerced condition in relation to its controls is not clear, but it is a possibility. Additionally, he removed some possible priming between words and produced grammatical consistency in the sentence continuations

between items, although we do not see why this should have been a concern. However, his modifications may even have changed the preferred interpretation of some items (Lapata, Keller, & Scheepers, 2003). To investigate these possibilities, we ran a new eye-tracking experiment using both de Almeida's (2004) stimuli and a new set of our own normed items. If the coerced condition is no harder than the other conditions in both sets of stimuli, then the claim that coercion produces processing difficulty is much weakened. If the coerced condition is harder than the other conditions in both sets of stimuli, then de Almeida's Experiment 1 appears to have produced a Type II Error, and the fact that coercion causes processing difficulty would be firmly established. If the sets of stimuli produce different patterns, then we should need to assess the results in light of the characteristics of the stimuli discussed above.

Additionally, we introduce a new control condition for assessing the level of difficulty associated with processing expressions like *began the memo*. In prior work, we contrasted these expressions with simpler versions containing the verb implicit in the coerced interpretation (e.g., *type the memo*) or with an expression using the same coercing verb but with a complement noun phrase that denotes an event (e.g., *began the meeting*). Both types of controls have their strengths but they differ in meaning from the coercing expression. For example, *The secretary typed the memo* is not semantically equivalent to "The secretary began to type the memo," the preferred interpretation of *The secretary began the memo*, because the former is telic (i.e., the described activity has an endpoint). And, although *The secretary began the meeting* and *The secretary began the memo* are matched in telicity, the former is arguably semantically simpler than latter in that does not involve an implicit verb phrase (viz., it does not mean "The secretary began to verb the meeting"). To match more precisely both semantic properties, we introduce a new control consisting of the full verb phrase inherent in the preferred interpretation of *The secretary began the memo*, specifically, *The secretary began typing the memo* or *The secretary began to type the memo*. Additionally, the tight coupling of meaning precisely controls for potential differences in plausibility, as the two sentence forms describe the same event.

3. Experiment

3.1. Method

3.1.1. Participants

Twenty undergraduates from the University of California, Davis participated in return for course credit. All of the participants were native English speakers with normal uncorrected vision and hearing.

3.1.2. Stimuli

We tested two sets of experimental sentences. The first 21 sets were de Almeida's (2004), which were modifications of a subset of the original McElree et al. (2001) stimuli. We made one change to the de Almeida stimuli. We changed *alpinist* to *backpacker* in one set of sentences because American English speakers are not likely to know the meaning of *alpinist*. The other test sentences (the DAVIS items) were developed and normed for American English speakers. For the de Almeida items, each sentence appeared in one of three versions, as in (1a–c); for the DAVIS items, each sentence appeared in one of four versions, as in (1a–d):

- (1a) The carpenter began the table during the morning break. (coerced)
- (1b) The carpenter built the table during the morning break. (preferred)
- (1c) The carpenter sanded the table during the morning break. (non-preferred)
- (1d) The carpenter began building the table during the morning break. (full-VP preferred)

In the *coerced* condition (1a), an event-denoting verb, such as *began*, was followed by a noun denoting an entity, such as *a table*. In the *preferred* condition (1b), the event-denoting verb was replaced by a verb specifying the action that readers infer when they read the coerced version. In the *non-preferred* condition (1c), the event-denoting verb was replaced by a verb specifying a plausible action connecting the subject and object of the verb, but one that readers rarely infer when reading the coerced version of the sentence. (Note that all verbs were in the simple past tense, as progressives like *was beginning* may be somewhat infelicitous.) In the *full-VP preferred* condition (1d), as with the preferred condition, we used the verb specifying the action that readers frequently infer when they read the coerced version, but the verb was embedded within the verb phrase headed by the coerced verb. (We did not run a full-VP condition using the non-preferred event.)

The only difference between the coerced, preferred, and non-preferred conditions was that the verbs changed, and so we matched the verbs for length and frequency (on the Brown corpus, Francis & Kucera, 1982). The average length of the verb for the coerced, preferred, and non-preferred conditions was 6.9, 6.5, and 6.9 characters, respectively [$F(2, 66) = 2.00$, NS, $MSe = .705$]. The average lemma frequency was 288, 223, and 153 for the three conditions, respectively [$F(2, 66) = 2.59$, $p = .08$, $MSe = 59,645$]. This marginal effect of frequency occurred because the non-preferred verbs were, on average, somewhat less frequent than the coerced verbs [$F(1, 33) = 5.19$, $p < .05$, $MSe = 59,645$]. Note, however, that this pattern of frequency differences cannot account for the reading time differences described in the following results section.

Preferred and non-preferred verbs were selected on the basis of completion norms administered to a group of 48 further participants. They were asked to provide fill-in-the-blank responses to coerced sentences like “The carpenter began _____ the table” indicating how they would interpret the string. The set of verbs used in the preferred conditions were selected 47% of the time, while the set of verbs used in the non-preferred verbs were selected 3% of the time. Forty-seven percent is somewhat lower than is typical of our past studies, but we felt that it was more important to match verbs on length and frequency, and this required using variants of the dominant verb given by subjects in 13 of the 33 constructions. However, when a different preferred verb was selected, we attempted to select one that expressed the same sense as the dominant verb. For example, we used *The gymnast crossed the beam during the morning competition* rather than *The gymnast walked the beam during the morning competition*.

Given that additional compositional operations are required in the coerced condition, as indicated by previous results from experiments testing similar stimuli, we predicted slower reading times in the coerced condition than in the two preferred conditions. If readers found the non-preferred sentences semantically odd, that condition might also be elevated relative to the preferred condition. But if the non-preferred interpretation was as plausible as the preferred interpretation, we expect readers to process it about as quickly as they process the preferred condition.

3.1.3. Procedure

A Fourward Technologies Dual-Purkinje Image eye-tracker monitored participants' eye movements while they read sentences like (1a)–(1d). The tracker has angular resolution of 10' of arc. The tracker monitored only the right eye's gaze location, but viewing was binocular. A PC displayed materials on a VDU 70cm from participants' eyes. The display consisted of Borland C default font with approximately four characters per degree of visual angle. 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 passage, the participant pressed a key. After 10 of the filler passages, the participant responded to comprehension questions. Participants did not receive 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 indi-

cated the participant's current gaze location. If the tracker was misaligned, the experimenter recalibrated it before proceeding with the next trial.

3.2. Results

We identified three scoring regions for analysis. The first is the *main verb* region [i.e., for de Almeida stimuli, using (1a–c) as examples, *began* vs. *built* vs. *sanded*; for the DAVIS stimuli, *began* vs. *built* vs. *sanded* vs. *building*]. This region included auxiliary verbs in items that had one. The second is the *noun* region, which included the determiner and the head noun (e.g., *the table*). The third is the *post-noun* region, which included the two words immediately following the noun region. We report three standard eye-movement measures. *First-pass time* includes all of the fixations within a scoring region from when the reader's gaze first enters the region until the reader leaves the region, either to the left or to the right. *First-pass regressions* occur when the reader's gaze crosses the region's left hand boundary immediately following a first-pass fixation. *Total time* includes all of the fixations within the scoring region. Mean values for the three dependent measures by stimulus set, region, and condition appear in Table 1. Figs. 1 and 2 represent first-pass and total time for the two stimulus sets by region and condition. We analyzed eye-movement patterns for the two stimulus sets separately. The data were initially subjected to a series of one-way ANOVAs treating Condition (coerced vs. preferred vs. non-preferred, for the de Almeida stimuli; coerced vs. preferred vs. non-preferred vs. full-VP preferred, for the DAVIS stimuli) as a within-participants and -items factor.

3.3. de Almeida stimuli

3.3.1. Verb region

There was a trend toward a main effect of condition in the total time data [$F_1(2, 38) = 2.51$, $p < .10$, $MSe = 14,562$; $F_2(2, 40) = 2.25$, $p = .12$, $MSe = 16,715$]. Although the preferred and non-preferred conditions did not differ, and the coerced condition did not differ from the non-preferred condition, the coerced condition had greater total time than the preferred condition [$F_1(1, 19) = 4.78$, $p < .05$; $F_2(1, 20) = 4.41$, $p < .05$].

3.3.2. Noun region

There was a main effect of condition in the first-pass time data [$F_1(2, 38) = 3.45$, $p < .05$, $MSe = 11,866$; $F_2(2, 40) = 4.11$, $p < .05$, $MSe = 8032$]. This main effect occurred because first-pass times in the non-preferred condition were longer than in the preferred condition [$F_1(1, 19) = 6.75$, $p = .01$; $F_2(1, 20) = 7.76$, $p < .01$]. There also was some indication that the coerced condition had longer first-pass time than the preferred condition. Specifically, tests for simple effects produced a significant result

Table 1
Mean first pass, first-pass regressions, and total time by stimulus set, region, and condition

Condition	Region		
	Verb	Noun	Post-noun
de Almeida stimuli			
<i>First-pass time (ms)</i>			
Coerced	340	415	367
Preferred	317	381	319
Non-Preferred	316	471	387
<i>First-pass regressions (%)</i>			
Coerced	8.8	18.4	14.2
Preferred	8.5	6.7	8.2
Non-Preferred	9.2	11.2	6.5
<i>Total time (ms)</i>			
Coerced	505	616	468
Preferred	422	520	476
Non-Preferred	447	605	505
DAVIS stimuli			
<i>First-pass time (ms)</i>			
Coerced	344	463	340
Preferred	345	443	356
Non-Preferred	355	427	347
Full-VP preferred	337 (387)	435	350
<i>First-pass regressions (%)</i>			
Coerced	12.2	19.0	14.5
Preferred	7.43	9.5	8.4
Non-Preferred	15.4	11.1	7.8
Full-VP preferred	10.7 (10.1)	16.0	7.0
<i>Total time (ms)</i>			
Coerced	527	684	471
Preferred	483	551	517
Non-Preferred	524	556	505
Full-VP preferred	513 (529)	557	483

Note. Material in parentheses reflects values for the complement of the main verb in the full-VP preferred condition: The verb region was broken in two, with the first part being the same as the coerced condition (e.g., *began*) and the rest reflecting the additional material that instantiated the preferred interpretation (e.g., *building*).

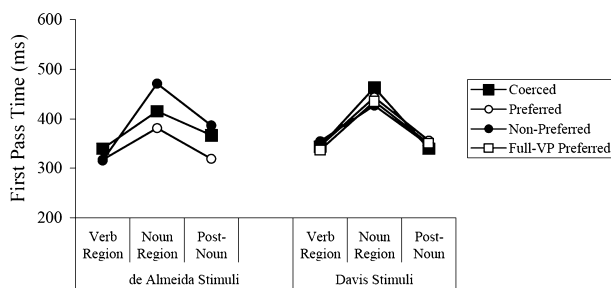


Fig. 1. First-pass time by region, condition, and stimulus set.

in the by-items analysis [$F_2(1, 20) = 3.91, p = .05$], but the by-participants analysis did not produce a significant result [$F_1(1, 19) = 2.67, p = .11$]. There was also a significant main effect of condition in the first-pass regressions data [$F_1(2, 38) = 4.32, p < .05, MSe = 160; F_2(2, 40) = 3.33, p < .05, MSe = 235$]. This effect occurred because the

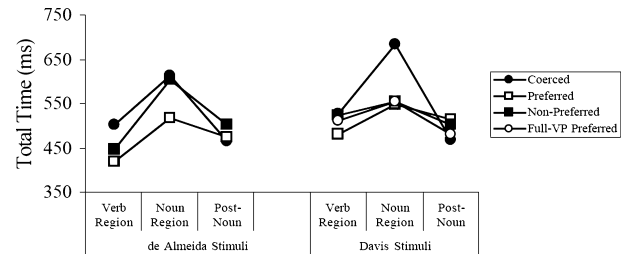


Fig. 2. Total time by region, condition, and stimulus set.

coerced condition produced more regressions from this region than the preferred condition [$F_1(1, 19) = 8.59, p < .01; F_2(1, 20) = 6.55, p < .01$]. The total time data produced a main effect of condition in the by-participants analysis [$F_1(2, 38) = 3.06, p = .06, MSe = 17,788$], but not in the by-items analysis [$F_2(2, 40) = 2.26, p = .12, MSe = 24,247$]. Tests for simple effects showed that the coerced condition had longer total time than the preferred condition [$F_1(1, 19) = 5.11, p < .05; F_2(1, 20) = 3.77, p = .06$]. The non-preferred condition also had greater total time than the preferred condition [significant by participants, $F_1(1, 19) = 4.01, p = .05$; and nearly significant by items, $F_2(1, 20) = 2.96, p < .10$]. The data indicate that processing the noun phrase was more difficult in both the coerced and non-preferred conditions.

3.3.3. Post-noun region

There was only a trend toward a main effect of condition in the first-pass regressions data [$F_1(2, 38) = 2.81, p = .07, MSe = 135; F_2(2, 66) = 2.30, p = .11, MSe = 227$], where tests for simple effects indicated that more regressions occurred in the coerced condition than the non-preferred condition [$F_1(1, 19) = 4.36, p < .05, MSe = 135; F_2(1, 20) = 3.32, p = .08, MSe = 183$].

3.4. DAVIS stimuli

Data from the DAVIS stimuli were subjected to 1×4 Condition (coerced vs. preferred vs. non-preferred vs. full-VP preferred) repeated measures ANOVAs, with condition treated as a within-participants and -items random factor.

3.4.1. Verb region

No effects attained statistical significance.

3.4.2. Noun region

The first-pass data from the noun region did not produce any effects on first-pass ($F_1, F_2 < 1, NS$). However, the first-pass regressions data produced a strong trend toward an effect of condition [$F_1(3, 57) = 2.52, p = .07, MSe = 150; F_2(3, 99) = 2.75, p < .05, MSe = 225$], suggesting that there were fairly rapid effects of the experimental manipulation. Tests for simple effects indicated that the coerced condition had more first-pass regres-

sions than either the preferred [$F(1, 19) = 6.05, p < .05; F(2, 33) = 6.68, p = .01$] and the non-preferred condition [$F(1, 19) = 4.11, p < .05; F(2, 33) = 4.53, p < .05$], and there was a trend toward a difference between the coerced and the full-VP preferred conditions [$F(1, 19) = 2.67, p = .11; F(2, 33) = 2.78, p = .10$]. The preferred, non-preferred, and full-VP preferred conditions did not differ (all $F < 1.48$, NS).

The total time data also produced a main effect of condition [$F(3, 57) = 5.44, p < .01, MSe = 15,324; F(2, 99) = 4.25, p < .01, MSe = 29,399$]. This effect was driven by the coerced condition, which had longer total time than any of the other three conditions, which did not differ from one another [coerced vs. preferred: $F(1, 19) = 11.4, p < .01, F(2, 33) = 9.31, p < .01$; coerced vs. non-preferred: $F(1, 19) = 10.6, p < .01; F(2, 33) = 7.58, p < .01$; coerced vs. full-VP preferred: $F(1, 19) = 10.5, p < .01; F(2, 33) = 8.49, p < .01$; (all other $F < 1$, NS)].

The first-pass regressions results from the noun region indicate that readers began to experience difficulty in the coerced condition relative to the other three conditions fairly quickly after fixating the critical noun. The total time data indicate very clearly that processing difficulty was isolated to just the coerced condition.

3.4.3. Post-noun region

First-pass and total-time in the post-noun region did not differ by condition (all $F < 1$, NS). However, there was a trend toward an effect in the first-pass regressions data [$F(3, 57) = 2.28, p = .09, MSe = 104; F(2, 99) = 2.17, p = .10, MSe = 199$]. The pattern was the same as the previous region. The coerced condition had more first-pass regressions than any of the other conditions, which did not differ from one another [coerced vs. preferred: $F(1, 19) = 3.60, p = .06; F(2, 33) = 3.58, p = .06$; coerced vs. non-preferred: $F(1, 19) = 4.38, p < .05; F(2, 33) = 4.26, p < .05$; coerced vs. full-VP preferred: $F(1, 19) = 5.43, p < .05; F(2, 33) = 4.99, p < .05$, all other $F < 1$, NS]. These results also indicate fairly rapid effects of coercion.

4. Discussion

In summary, coercion costs appeared for both sets of stimuli. For the DAVIS stimuli, the results are straightforward, with the coerced condition causing more first-pass regressions and greater total time in the noun region than the other conditions. The preferred, full-VP preferred, and non-preferred conditions did not differ. The first-pass regressions data for the post-noun region produced a tendency toward the same pattern. For the de Almeida (2004) stimuli, there was also evidence of increased first-pass regressions for the coerced condition in the noun region, though it only differed significantly from the non-preferred condition, and

a strong tendency toward greater total time in the noun region than the other two conditions. There was weak evidence of coercion cost on first-pass regressions from the post-noun region, and on total time on the verb region. Finally, first-pass time on the noun region suggested some difficulty with the non-preferred condition at this point. Collectively, these data are less-clear cut than the results with the DAVIS stimuli, but the first-pass regressions from the noun region in particular clearly indicate coercion cost. Although de Almeida's stimuli produce more variable results than our own, the overall patterns are similar. We therefore conclude that coercion produces processing difficulty. Moreover, this difficulty emerged during very early processing, as indicated most clearly by the first-pass regressions effect on the noun region for the DAVIS stimuli.

It therefore seems likely that de Almeida's (2004) Experiment 1 produced a Type II Error. The fact that our replication of his experiment produced less stable results than the DAVIS stimuli suggests that this may partly reflect some characteristics of his stimuli. These may include the length and frequency differences noted above, as well as other unknown properties.

Our more fine-grained eye-tracking measures on both sets of stimuli produced results that were fairly comparable to McElree et al. (2001) and Traxler et al. (2002). However, there are a couple of points to note. First, effects of coercion for both sets emerged on the noun region itself. This accords with McElree et al. but none of Traxler et al.'s experiments, where the critical effects occurred on the following words, sometimes on re-reading of the verb region. The location of effects in eye-tracking (and self-paced reading) experiments varies slightly (see Pickering, Frisson, McElree, & Traxler, 2004), and it appears that all we can conclude is that coercion costs can occur on the noun itself. Second, the DAVIS stimuli showed no sign of any difference between the preferred and non-preferred condition. This is in keeping with Traxler et al.'s Experiment 1 but not with McElree et al. We suspect that preferred and non-preferred conditions do not differ if plausibility is matched. If this is correct, the effect in McElree et al. was erroneous (note that it only occurred on the noun, unlike the coercion effect), and that effects with de Almeida's materials may have reflected uncontrolled stimulus differences. But whether this is correct or not, differences in preference cannot explain coercion costs, which occur whether the coerced condition is compared with the preferred condition, with the non-preferred condition, or, as additionally documented here, with the condition using a full VP to express the event sense. Additionally, the fact that the coercion difficulty emerged during very early processing shows that this cost cannot just reflect "later interpretative processes" (de Almeida, 2004, p. 252).

The findings from de Almeida's (2004) Experiment 2, where the sentences from his Experiment 1 were pre-

ceded with a short context (typically two sentences), are in many respects consistent with the experiment reported here (although it is possible that the contexts may have altered the plausibility of the different conditions, or even changed the preferred interpretation, as no pre-tests were conducted). The purpose of the context was “to constrain the range of possible events to be described by the target sentence.” Like us, he reports a reliable coercion effect, with the coerced condition showing significantly longer reading times than the preferred control. This effect occurs at his noun phrase region (presented as a single region) and on the next word. However, the non-preferred condition behaves more like the pattern observed in McElree et al. (2001). Numerically, this condition is comparable to the coerced condition at the noun phrase and then drops to a middling position between the coerced and preferred condition at the next word. At the noun phrase, the non-preferred condition, like the coerced condition, was found to be significantly slower than the preferred condition. However, at the next word, the non-preferred condition was not significantly different from either the preferred or the coerced condition.

Intermediate results like these are difficult to interpret, but de Almeida’s (2004) interpretation is certainly unwarranted. He infers that “type-shifting constructions behave similarly to non-preferred constructions” (p. 258). However, there is simply no basis for this conclusion. If condition A (preferred) is significantly faster than condition B (coerced), and condition C is not significantly different from either conditions A or B, one cannot draw any conclusion about whether condition C is like condition A or condition B. On theoretical grounds, crucially, even if the power of the experiment were increased and the non-preferred condition was found to pattern like the coerced condition, it would not license de Almeida’s conclusion that the non-preferred condition was processed in a similar fashion to the coerced condition. To do so would be to assume that numerically similar reading time differences arise from identical underlying processes. This of course need not be true. In this case, there is no reason to assume that processing an expression where the reader is explicitly given a non-preferred event involves the same type of processing as when the reader must infer an unexpressed event. The assumption is all the more questionable when McElree et al. (2001) and Traxler et al. (2002, Experiment 1), along with the current study, showed that non-preferred constructions do significantly differ from coerced constructions.

The effect of context on coercion is an important issue. Traxler, McElree, Williams, and Pickering (submitted) systematically examined the affect of varying contextual information on the cost of coercion. They found that the coercion effect was not eliminated or attenuated when the immediate context provided the activity implicit in the complement coercion. Consequently, retrieving or infer-

ring the activity does not appear to be responsible for the difficulty, and difficulty remains when any ambiguity about the implicit event is all but removed. de Almeida’s (2004) Experiment 2 appear to be compatible with this: Difficulty with the coerced condition versus the preferred condition occurs even in the context of an event schema that is likely compatible with the preferred interpretation of the coerced sentence. On this account, the unreliable difference between the coerced and the non-preferred condition either is a Type II error or reflects plausibility differences.

Finally, the additional control condition employed in our study—sentences employing an embedded VP expressing the preferred sense implicit the coerced expression—provides further evidence for a cost associated with complement coercion. That we found reliable differences when measured against this type of control alleviates any concern that other documented differences might be due to uncontrolled semantic differences between conditions.

4.1. Accounting for the difficulty of coercion

As with prior published findings (McElree et al., 2001; Traxler et al., 2002), the current study does not enable us to determine what specific operations distinguish simple from enriched composition. However, collectively these studies demonstrate unequivocally that expressions argued to require an enriched composition are indeed costly to process. Further experimental work will be needed to precisely determine what specific operations are involved. Below we briefly outline our current hypothesis and how this hypothesis differs from the perspective sketched by de Almeida (2004).

Pustejovsky’s (1995) treatment of coercion views it as an operation that converts an expression, α , to the semantic type expected by a governing function, β . In *began the book*, the verb *began*, β , selects for an eventive function, and a coercion operation is applied to convert the expression *the book*, α , from its default semantic type entity to the type event. This is accomplished by first selecting an activity compatible with the agent and commonly associated with the noun, and then incorporating it into the interpretation of the expression by building an event structure analogous to the structure in (2):

(2) [β began[α reading the book]].

Pustejovsky (1995) argues that the activity incorporated into the event structure (e.g., *reading* in 2) is often derived from the *Qualia* structure stored in the noun’s lexical representation (which contains the information that books are read). We have reservations about the strong representational assumptions in Pustejovsky’s framework (see also Jackendoff, 2002). However, this aspect of the framework is not central to our account

because we do not attribute the increased processing costs to the processes that are responsible for retrieving or inferring the activity implicit in the event structure (Traxler et al., submitted). Rather, we argue that the observed costs are engendered by the operations that readers undertake to build the representation for the event sense of the complement in (2). That is, our emphasis is on the compositional operations needed to generate a representation for the expression.

Importantly, our hypothesis is compatible with either the decompositional or the atomistic view of lexical representation. Under the latter, the activity in the event sense can be retrieved from complex lexical representations; in the former, the activity can be inferred from some meaning postulate or more general inference mechanism. However, in both cases, comprehenders must make use of a representation that realizes an eventive interpretation of the complement, an operation that is apparently costly. Consequently, we believe that de Almeida's (2004) use of time-course measures of coercion phenomena to motivate an alternative atomistic account of lexical representations is largely misguided. In any case, the data do not support his conclusions.

References

- Bach, E. (1986). The algebra of events. *Linguistics and Philosophy*, 9, 5–16.
- de Almeida, R. G. (2004). The effect of context on the processing of type-shifting verbs. *Brain and Language*, 90, 249–261.
- 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: Lexicon and grammar*. Boston: Houghton Mifflin.
- Jackendoff, R. (1997). *The architecture of the language faculty*. Cambridge, MA: MIT Press.
- Jackendoff, R. (2002). *Foundations of language*. New York: Oxford University Press.
- Lapata, M., Keller, F., & Scheepers, C. (2003). Inter-sentential context effects on the interpretation of logical metonymy. *Cognitive Science*, 27, 649–668.
- Lascarides, A., & Copestake, A. (1998). Pragmatics and word meaning. *Journal of Linguistics*, 34, 387–414.
- McElree, B., Traxler, M. J., Pickering, M. J., Seely, R. E., & Jackendoff, R. (2001). Reading time evidence for enriched semantic composition. *Cognition*, 78, B15–B25.
- Pickering, M. J., Frisson, S., McElree, B., & Traxler, M. J. (2004). Eye movements and semantic composition. In M. Carreiras, Jr. & C. Clifton (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERPs, and beyond*. Hove: Psychology Press.
- 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.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372–422.
- Traxler, M. J., McElree, B., Williams, R. S., & Pickering, M. J. (submitted). Context effects in coercion: Evidence from eye-movements. Submitted.
- Traxler, M. J., Pickering, M. J., & McElree, B. (2002). Coercion in sentence processing: Evidence from eye-movements and self-paced reading. *Journal of Memory and Language*, 47, 530–547.