

Complement Coercion Is Not Modulated by Competition: Evidence From Eye Movements

Steven Frisson
University of Birmingham

Brian McElree
New York University

An eye-movement study examined the processing of expressions requiring complement coercion (J. Pustejovsky, 1995), in which a noun phrase that does not denote an event (e.g., *the book*) appears as the complement of an event-selecting verb (e.g., *began the book*). Previous studies demonstrated that these expressions are more costly to process than are control expressions that can be processed with basic compositional operations (L. Pykkänen & B. McElree, 2006). Complement coercion is thought to be costly because comprehenders need to construct an event sense of the complement to satisfy the semantic restrictions of the verb (e.g., *began writing the book*). The reported experiment tests the alternative hypotheses that the cost arises from the need to select 1 interpretation from several or from competition between alternative interpretations. Expressions with weakly constrained interpretations (no dominant interpretation and several alternative interpretations) were not more costly to process than expressions with a strongly constrained interpretation (1 dominant interpretation and few alternative interpretations). These results are consistent with the hypothesis that the cost reflects the on-line construction of an event sense for the complement.

Keywords: complement coercion, competition, semantics, sentence processing, eye movements

The meaning of many expressions in natural language can be constructed by applying a small set of basic compositional operations (Jackendoff, 2002; Pykkänen & McElree, 2006). However, some expressions contain semantically mismatching elements that appear to block the application of these basic operations and which seem to require more elaborate operations to repair the mismatch. Coercion is thought to be one type of operation that comprehenders can use to repair semantically mismatching elements in an expression (see Pykkänen & McElree, 2006).

Complement coercion is argued to be required when a verb requiring an event-denoting complement is paired with a nonevent-denoting complement (Jackendoff, 1997, 2002; Pustejovsky, 1995). Sentences 1–3 illustrate the paradigm:

1. *The man began/finished the hike.*
2. *The man began/finished reading the book.*
3. *The man began/finished the book.*

Aspectual verbs such as *begin* and *finish* are prototypical examples of verbs that semantically require a complement that denotes an event, either a noun phrase, such as *the hike* in 1, or a verb phrase, such as *reading the book* in 2. The event-denoting complement semantically combines with the verb to form a predicate denoting the initial or final part of the denoted event. It is interesting to note, however, that a similar interpretation results if the verb is paired with a nonevent-denoting complement, such as the simple noun phrase *the book* in 3. Indeed, if asked to indicate the meaning of 3, comprehenders will indicate that it has the same interpretation as 2 (McElree, Traxler, Pickering, Seely, & Jackendoff, 2001). That is, they seem to have “coerced” the complement into the right semantic type by embedding it in a plausible but unexpressed event (e.g., *began the book* is construed as *began doing something to the book*).

There is now a considerable body of research on the processing of expressions requiring complement coercion (see Pykkänen & McElree, 2006, for an overview). Coerced sentences such as 3 have been compared to noncoerced control sentences such as 1 and 2, as well as to expressions such as 4:

4. *The man read the book.*

Different processing measures have shown that expressions such as (3) requiring complement coercion are more costly to interpret than controls; these measures include self-paced reading and eye-tracking measures (Lapata, Keller, & Scheepers, 2003; McElree, Frisson, & Pickering, 2006; McElree et al., 2001; Pickering, McElree, Frisson, Chin, & Traxler, 2006; Pickering, McElree, & Traxler, 2005; Traxler, McElree, Williams, & Pickering, 2005; Traxler, Pickering, & McElree, 2002), magnetoencephalographic (MEG) patterns (Pykkänen, Llinás, & McElree, in press), and speed–accuracy trade-off measures (McElree, Pykkänen,

Steven Frisson, School of Psychology, University of Birmingham, Birmingham, England; Brian McElree, Department of Psychology, New York University.

This research was supported by National Science Foundation Grant 0236732 awarded to Brian McElree. We wish to thank Lillian Chen for conducting the pretests on the materials used in this study, Andrea Martin for her help in running the experiment, and Andrew Olson for his help with the statistical analyses.

Correspondence concerning this article should be addressed to Steven Frisson, School of Psychology, University of Birmingham, Edgbaston, Birmingham, England B15 2TT, United Kingdom. E-mail: s.frisson@bham.ac.uk

Pickering, & Traxler, 2006). For example, eye-tracking experiments have consistently found that coerced expressions engender more regressions back from the complement and longer total reading times on the verb phrase and spillover regions than control expressions do.

Several alternative explanations for this effect have been explored and rejected. For example, the coercion cost cannot be explained with an assumption that eventive verbs are more semantically complex than the verbs used in control conditions such as 4 (cf. Gennari & Poeppel, 2003). Traxler et al. (2002) found that the cost is uniquely linked to the pairing of an eventive verb and a non-event-denoting complement (e.g., *started the puzzle*) and that no comparable effect is found when the eventive verb is paired with an event-denoting complement (e.g., *started the fight*). Additionally, the cost cannot be attributed to differences in cloze probability, overall acceptability or plausibility, or to fine-grain co-occurrence differences between the verb and its complement. For example, robust coercion effects have been found when the experimental and control constructions were closely matched in terms of rated acceptability (e.g., Traxler et al., 2005), rated plausibility (e.g., McElree, Pyllkkänen, et al., 2006), and cloze probability (e.g., Traxler et al., 2002). Pyllkkänen and McElree (2006) showed that co-occurrence frequencies of particular verb-complement constructions (based on latent semantic analysis; Landauer, Foltz, & Laham, 1998) do not account for the coercion cost; in fact, they found that coerced constructions that had a higher frequency of co-occurrence than control constructions still engendered a robust coercion effect (see also Frisson, Rayner, & Pickering, 2005, for evidence against co-occurrence probability effects).

Why Is Complement Coercion Costly?

Comprehenders must often go beyond conventional word senses to obtain the appropriate interpretation of an expression. For example, Nunberg (2004) argued that standard metonymies such as *read Dickens*, where Dickens is naturally taken to refer to the writings of Dickens rather than to the person, involve a type of *deferred interpretation*, in that the expression is “used to refer to something that isn’t explicitly included in the conventional denotation of that expression” (p. 344). Interestingly, however, a wide range of these types of metonymic expressions, including PERSON-FOR-PRODUCT (e.g., *read Dickens*), PLACE-FOR-EVENT (e.g., *protested during Vietnam*), and PLACE-FOR-INSTITUTION (e.g., *talked to the school*) metonyms are not more costly to interpret than control expressions involving the conventional denotations of the nouns (Frisson & Pickering, 1999, 2007). These results suggest that constructing an alternative sense per se is not costly. McElree, Frisson, and Pickering (2006) explicitly contrasted expressions with metonymic and coerced interpretations, comparing both to expressions involving conventional interpretations. Eye-movement measures during reading indicated that metonymies such as *the gentleman read Dickens* were not more costly to interpret than were conventional expressions such as *the gentleman met Dickens*, but expressions that required coercing Dickens into an event, such as *the gentleman started Dickens*, were more taxing to interpret than both.

McElree and colleagues have argued that expressions involving complement coercion are more costly to interpret because they

require comprehenders to do more than simply shift the denotation of the complement. Instead, they require comprehenders to construct an event sense of the complement, one that is not lexically stored or available in the immediate discourse (McElree, Frisson, & Pickering, 2006; McElree, Pyllkkänen, et al., 2006; McElree et al., 2001; Pyllkkänen & McElree, 2006; Traxler et al., 2002, 2005).

Formal approaches to complement coercion, such as Pustejovsky (1995), treat it as an operation that converts an expression, α , into the semantic type expected by the function, β , which governs α . In *began the book*, the verb *begin* selects for an eventual function, β , and a coercion operation is applied to convert the complement *the book*, α , from its default semantic type ENTITY to the type EVENT. Pustejovsky (1995) argued that this is accomplished by first selecting an activity compatible with the agent and commonly associated with the complement noun, assumed to be stored in what he refers to as a noun’s *Qualia* structure, and then incorporating this activity into the interpretation of the expression by building an event structure, such as [β *began*[α *reading the book*]]. We do not fully subscribe to the lexical representations assumed in Pustejovsky’s framework because, minimally, we assume that coerced senses are computed from a broader range of properties than the *Qualia* structure of the complement noun proposed by Pustejovsky (see also Jackendoff, 2002).¹ Nonetheless, we attribute the increased processing costs in coercion to operations analogous to those that Pustejovsky proposed are used to build a representation for the event sense of the complement.

Specifically, as described in Traxler et al. (2005), we propose that comprehenders interpret expressions like *began the book* with the following sequence of operations: Step 1. When processing the noun, comprehenders attempt to integrate various stored senses of the word into the developing semantic representation of the sentence, but the semantic mismatch between the requirements of the verb and the stored senses of the noun blocks the application of any simple composition operation (see Pyllkkänen & McElree, 2006). Step 2. Comprehenders then use available lexical and discourse information to infer a plausible event in which to embed the noun phrase. Step 3. Once identified, comprehenders incorporate the event sense into their semantic interpretation of the verb phrase by reconfiguring the semantic representation of the complement from [β *began*[α *the book*]] into [β *began*[α *reading the book*]].

We argue that available evidence indicates that it is this final step that engenders the observed cost. There are several reasons to doubt that the locus of the effect resides in Step 1, namely, the observed cost simply reflects the detection of semantic anomaly (see Traxler et al., 2005). The most compelling empirical evidence comes from a recent MEG study of complement coercion (Pyllkkänen et al., in press). Complement coercion does not modulate the same brain activity found in a clear case of mismatching semantic relations between a verb and its complement. Relative to control expressions such as *the author wrote the book*, anomalous expres-

¹ Comprehenders are readily able to use properties that do not appear to be exclusively derived from the complement noun but which instead are associated with the agent of the sentence or other discourse elements. For example, authors typically write but goats do not; hence, interpreting *the goat began the book* will rely more on what goats can do to books than on the typical properties of books.

sions such as *the author amused the book* increase the activity in a left temporal source at 300–400 ms (M350), which is the MEG analogue of an event-related potential N400 component. However, coerced expressions such as *the author began the book* generate the same activity levels in this source as control expressions. Instead, coerced expressions modulate a frontal source (an anterior midline field) in a later 350–500-ms time window, generating more activity in this source than in either the anomalous or control sentences. The distinct neural sources suggest that the coercion cost reflects something other than the detection of an anomaly.

A more plausible alternative to our hypothesis that the effect resides in Step 3 is that the locus actually resides in operations performed at Step 2. One possibility is that the cost reflects the time needed to retrieve or infer the activity implicit in the event interpretation of the coerced complement. For example, in interpreting *the man began the book*, the cost could reflect the time required to infer that the man likely initiated the activity of reading the book. Traxler et al. (2005) tested this hypothesis by placing the required activity in the preceding context, reasoning that this would eliminate the cost if the difficulty involved inferring an appropriate activity. They found that a context sentence such as *The contractor had been building in the suburbs* did not eliminate the cost associated with the coerced expression *That spring, he began a condominium. . . .* This finding speaks against attributing the cost to the time needed to retrieve or infer the activity implicit in the event sense. However, it is consistent with an account that attributes the cost to the operations outlined in Step 3, as building an event sense into the semantic interpretation of the verb phrase would be required whether or not context included the activity that was implicit in the eventive interpretation of the complement.²

That the coercion cost is not eliminated by placing the required activity in the preceding context also provides some evidence against attributing the cost to selecting an activity from a set of plausible actions. The interpretation of expressions involving complement coercion are inherently underdetermined: For example, a sentence such as *the man began the book* can be interpreted in several ways, including, among others, that the man might have started to read the book, to write the book, to translate the book, or even to pack the book (if he was moving). In this sense, these underspecified expressions may be viewed as being ambiguous: not in the more restricted sense of being ambiguous between two or more fixed alternatives but in the common sense that they can be understood in more than one way. In principle, the processing cost could reflect the time required for comprehenders to select an activity from several possible alternatives, or it could reflect competition among alternative interpretations.

Given that constraining contexts often eliminate effects due to ambiguity (cf. Altmann & Steedman, 1988; Binder & Morris, 1995; Hess, Foss, & Carroll, 1995; Pickering & Traxler, 1998), it is reasonable to expect that having a plausible activity in the immediate context would have reduced the inherent ambiguity of the verb phrase such as *began the table* and eliminated the coercion cost. Hence, Traxler et al.'s (2005) finding is not fully compatible with accounts that would attribute the cost to the time required to instantiate one of several possible interpretations. However, the evidence against this alternative is not as strong as one would like, as it is somewhat indirect and relies on null findings. It is always possible that another type of context might cue the activity better and thereby eliminate the cost.³ In addition, Traxler

et al. tested only complement coercions with one dominant interpretation and did not control for the number of alternative interpretations, which could have affected the coercion cost as well. In the reported experiment, we took a different approach to testing this alternative hypothesis. Rather than attempting to reduce the inherent ambiguity of the coerced expression by various contextual manipulations, we investigated whether the coercion effect itself is modulated by the degree to which the coerced expression is inherently ambiguous.

Is the Coercion Cost Modulated by Ambiguity?

If the coercion cost is due to inherent ambiguity, then the magnitude of the effect should vary with the degree to which an expression is likely to engender different interpretations. For example, consider the Sentences in 5 and 6.

5. *The student finished the essay.*

6. *The director started the script.*

Intuition suggests that 5 is very likely to be interpreted as *the student finished writing the essay*, whereas 6 could be interpreted in several ways. A director often shoots or films a script, but he or she can also write, read, edit, review, or even rehearse a script. One way to measure how coerced expressions are likely to be interpreted is to use a fill-in-the-blank completion task (McElree et al., 2001). Participants are given sentences containing a blank between the verb and the complement, for example, *The student finished _____ the essay* and *the director started _____ the script*, and are asked to fill in the blank with one or two words indicating how they would interpret the sentence. Completion norms (see below) on 5 indicate that the activity “writing” is the dominant response to *finished the essay*, being given 92% of the time. The remaining 8% of the responses were either “typing” or “revising.” The dominant response to 6 was “writing,” but this response was given only 35% of the time. The remaining completions were composed of eight responses: “reading” (26%), “directing” (17%), “filming” (4%), “drafting” (4%), “using” (4%), “analyzing” (4%), and “reviewing” (4%).

If the coercion cost is due to the inherent ambiguity of coerced expressions—either because the difficulty of the selection process scales with the number of alternatives or because there is more competition engendered when there are more options and no single clearly dominant interpretation—then we would expect that the magnitude of the effect would be greater in cases such as 6 than they would be in 5. One could argue that this type of prediction

² Traxler et al. (2005) found that the cost was virtually eliminated if the context sentence contained the same coercion (e.g., *the student started a book. . . Before he started the book/it. . .*) or explicitly stated the underlying event structure (e.g., *the student read a book. . . Before he started the book/it. . .*). Hence, the cost of building an event sense for the target expression can be circumvented if a relevant event sense is in the immediate discourse.

³ In a second experiment, Traxler et al. (2005) found that a slightly modified item set did in fact attenuate, although not eliminate, the coercion cost. These items used nouns in the complement that were given more frequently in a norming task, for example, having *he began the house* rather than *he began the condominium*.

receives some support from analogous effects found in research on lexical and syntactic ambiguities, even though there are salient differences between these types of ambiguity and the underspecification found in complement coercion. Inflated fixation times are found for ambiguous words when the difference in frequency of the two meanings is small (e.g., Rayner & Duffy, 1986) but not when one meaning is substantially more frequent (e.g., Duffy, Morris, & Rayner, 1988; Frazier & Rayner, 1990; Rayner, Binder, & Duffy, 1999). Likewise, with syntactic ambiguity, processing costs appear to be highest when alternative analyses receive a comparable amount of support and lowest when one analysis is strongly favored (e.g., MacDonald, 1994; McRae, Spivey-Knowlton, & Tanenhaus, 1998; Spivey & Tanenhaus, 1998; Tabor & Tanenhaus, 1999; though see also Green & Mitchell, 2006). Perhaps more relevant, Thompson-Schill, D'Esposito, Aguirre, and Farah (1997) found higher activation in the left inferior frontal gyrus when the alternative responses in a verb generation task were of comparable frequency (though see Martin & Cheng, 2006, who showed that association strength rather than competition could explain these results).

In contrast, our hypothesis that the coercion cost reflects the building of an extended event sense of the complement predicts that the cost should not vary with the number of alternative interpretations or with whether or not there is a dominant interpretation of the expression. In general, we assume that the knowledge of the activity implicitly associated with most complements is constructed in an automatic and cost-free manner, and the cost is incurred by building an event sense, which is assumed to be a common requirement for interpreting 5 and 6.

To test these two hypotheses, we contrasted expressions such as 5 and 6. Prior investigations (e.g., Traxler et al., 2002) of complement coercion have tended to use expressions that have a clearly dominant interpretation, such as 5. As these types of expressions tend to engender a coercion cost relative to control structures, we expected to find a measurable cost here for expressions such as 5. At issue was whether this effect would be less pronounced than the effect observed for expressions such as 6, where one interpretation is less dominant and more interpretations are plausible.

Method

Participants

Forty students from New York University who spoke American English as their native language participated in the experiment. They were paid \$10 for one 45-min session. None of them participated in the pretests (see below).

Materials

We constructed 32 sentence quartets following what is illustrated in Sentences 7a–d (see Appendix for the full list of items). Two versions (7a and 7c) used eventive verbs (*began*, *started*) that logically require a complement expressing an event. As their complement (*novel*, *coffee*) does not refer to an event, the type mismatch is solved by coercing the complement into the required semantic type so that the expression can be interpreted as an unstated event involving this object (*began reading the novel*, *started serving the coffee*). In the other two conditions, the two

control conditions (7b and 7d), the verb can readily be paired with an object, and no coercion is required, as a straightforward compositional interpretation can be achieved.

- 7a. *The teenager began the novel as soon as he got to his room upstairs.* (strongly preferred–coerced)
- 7b. *The teenager read the novel as soon as he got to his room upstairs.* (strongly preferred–control)
- 7c. *The waitress started the coffee as soon as she returned to the counter.* (weakly preferred–coerced)
- 7d. *The waitress served the coffee as soon as she returned to the counter.* (weakly preferred–control)

The 7a and 7b versions of each quartet were identical except for the verbs, as were the 7c and 7d versions. We named the 7a version “strongly preferred” because Pretest 1 (see below) indicated that there is one strongly preferred interpretation for the coercion. The control condition 7b expresses this interpretation using a noneventive verb. The 7c version, termed “weakly preferred,” exemplifies an instance where multiple possible coercive interpretations were given in the pretest, with the control verb in 7d expressing the most common interpretation. (For one strongly preferred control item, the second most frequent interpretation [*iced*] was used in order to avoid repetition, as in “the baker baked. . . .” Analyses with this item taken out changed neither the pattern of results nor the strength of the effects.) As Pretest 1 shows, the most frequent interpretation of the strongly preferred–coerced items is on average twice as common as the most frequent interpretation of the weakly preferred–coerced items.

We matched as closely as possible the length and the frequency of the complements for each quartet: For the 7a and 7b versions, length (in characters) and frequency (in occurrences per million, based on Francis & Kucera, 1982) were 6.5 and 22.4, respectively; for the 7c and 7d versions, these values were 6.4 and 24.6, respectively. The differences are nonsignificant (all t s < 1). A spill-over region, defined as the next word following the complement, or the next two words if the first word was fewer than five characters long, was kept the same across the four conditions (e.g., *as soon*).

The 32-item quartets were divided over two lists so that each participant saw one coerced and one control condition of each quartet, without repetition of the complement (e.g., 7a and 7d, or 7b and 7c). Each list was also presented in its reverse order, and item presentation followed a fixed random order. The resulting four lists were randomly assigned to participants, with 10 participants for each list. This way, we obtained 64 data points for each participant. The critical items were intermixed with 174 filler items from different experiments. Due to the relatively small number of coercing verbs, some verbs were repeated. We tried to keep them apart as much as possible (on average, there were 58 intervening sentences, not including questions, between the repetition of the same verb); though even if there was some small repetition effect, this should decrease reading times for the coerced conditions, making it harder to find a coercion effect.

Pretests

We carried out two pretests, one to establish the preferred interpretation of a coerced construction, as well as the number of different interpretations of that construction, and the other one to examine the plausibility of the constructions.

Pretest 1: Preference. We generated 120 subject–verb–complement sentences and distributed them over two lists. The two lists were presented to 23 and 24 participants, respectively, and they were asked to fill in the blanks with a word or two that best expressed how they would interpret each sentence. For example, when presented with *the teenager began _____ the novel*, most participants would fill in *reading* or *to read*. For each sentence, we calculated three values: the frequency with which a specific verb was used, irrespective of its grammatical form (i.e., *reading* and *to read* were combined); the number of different verbs that were generated; and the number of different interpretations that were given to the constructs (e.g., *drinking* and *sipping* were combined). For the 32 strongly preferred–coerced constructions that we selected for the eye-tracking experiment, the dominant interpretation was used on average 90.4% of the time (range = 83%–100%), 3.2 different verbs were generated on average (range = 1–6), which translated into 2.7 different interpretations (range = 1–5). The ratio of the most frequently given interpretation to the second most frequent interpretation was 14:1. For the 32 weakly preferred–coerced constructions, these values were 45.4% (range = 27%–58%) for the most frequent interpretation, 9.0 (range = 4–14) for the number of different verbs, and 7.7 (range = 4–11) for the number of different interpretations. The ratio of most frequent to second most frequent interpretation was 2:1. All comparisons between the two sets were highly significant (all $ps < .001$). Hence, the complement coercions in the strongly preferred condition tended to have one strongly preferred or dominant interpretation and generated a smaller number of different interpretations. In contrast, the complement coercions in the weakly preferred condition were interpreted off-line in more diverse ways, with the most common interpretation used only about half as often as the most common interpretation for the strongly preferred set.

Pretest 2: Plausibility. The 120 subject–verb–complement sentences, along with their control form (using the most frequent noneventive verb interpretation, as determined by the completion norms), were divided into two lists. An equal number of filler sentences were added to each list. These filler items consisted of many implausible sentences, to encourage participants to use the full range of the rating scale. The task was to indicate, for each construction, how plausible the sentence was, with 1 (*totally implausible*) and 7 (*perfectly plausible*). The two lists were completed by 15 and 19 participants, respectively. The plausibility ratings for the four conditions indicated that all conditions were judged to be highly plausible: 6.4 ($SD = 0.4$) for strongly preferred–coerced, 6.7 ($SD = 0.2$) for strongly preferred–control, 6.3 ($SD = 0.5$) for weakly preferred–coerced, and 6.6 ($SD = 0.5$) for weakly preferred–control. Although the coerced conditions were judged lower than their control conditions were ($ps < .001$), a common finding probably related to participants having to make more effort to arrive at an interpretation, the two coerced conditions did not differ ($t < 1$). Crucially, the difference scores between the strongly preferred condition and its control, on the one

hand (0.3 on average), and between the weakly preferred and its control, on the other hand (0.3 on average) did not differ ($t < 1$). Hence, if we were to find a difference between the coerced conditions and their respective controls, then it would be unlikely that this difference was related to a discrepancy in plausibility.

Procedure

Participants were run individually on a SensoriMotor Instruments EyeLink I head-mounted eye-tracker (SensoriMotoric Instruments GmbH, Teltow, Germany), using software developed at the University of Massachusetts, Amherst (University of Massachusetts Eyetracking Lab, 2007). Eye cameras were positioned under each eye, recording eye movements and fixations every 4 ms. Screen resolution was set at $1,600 \times 1,200$ pixels. Sentences were presented in fixed font, with each letter being 18 pixels wide and 33 pixels high. A maximum of 80 characters were presented on each line of text. Participants were seated 71 cm from the display monitor; with this setup, 1° of visual angle corresponded to 2.7 characters. Viewing was binocular, but only the data from the eye that was calibrated best were used in the analyses. A chin rest was used to reduce head movements.

After signing a consent form, participants were presented with a general explanation of the eye-tracking procedure. Participants were encouraged to read at a normal pace for understanding. A calibration procedure was carried out at the beginning of the experiment, and recalibration was performed whenever the experimenter felt it necessary. Before a new trial was presented, participants first looked at a fixation box in the middle of the screen, and a drift correction was performed. They were then presented with a fixation box coinciding with the position of the first letter of the upcoming sentence. This box served as a trigger, with the sentence being displayed only if the fixation was judged to be close enough to the center of the box. Participants read sentences at their own pace and pressed a button on a hand-held button box to make the sentence disappear. Comprehension questions were asked after 50% of all trials (critical as well as filler trials), counterbalanced across conditions, with an equal number of *yes* and *no* responses. Participants answered the questions by pressing one of two buttons on the button box. An example of a *yes* question is “Did the newborn enjoy the milk?” and an example of a *no* question is “Did the gentleman eat a pizza?” Accuracy was high at 94.1%.

Analyses and Predictions

We report analyses on two regions, the complement (e.g., *the novel, the coffee*) and the spillover region (e.g., *as soon*). The following standard measures are discussed: first-pass duration (i.e., the summed fixations in a region before the eyes leave the region either to the left or to the right), second-pass duration (i.e., the summed fixations on a region after the eyes have fixated a region to the right of the critical region; this is usually taken as a measure of rereading), and total reading time (i.e., the sum of all fixations in a region). Fixations less than 80 ms and over 1,200 ms were excluded from the analyses. Maximum cutoff values were 1,600 ms for first-pass and second-pass reading times and 3,000 ms for total reading time (resulting in less than 1% removals). Analyses with different cutoffs did not change the pattern of results.

On the basis of previous results (e.g., McElree et al., 2001; Traxler et al., 2002), we expected to find a main effect of coercion, with the coerced conditions taking longer to process than the control conditions. We note that a main effect of preference is noninformative as it would merely indicate that, for whatever reason, the relations expressed by the subject–verb–complement in one subset were processed differently from the expressions in the other subset (e.g., that it is harder to understand *the waitress started/served the coffee* than to understand *the teenager began/read the novel*). Crucially, the hypothesis that frequency of interpretation and/or number of different interpretations is responsible for the cost of complement coercion predicts that an interaction should be observed. Specifically, we expected that the difference between the weakly preferred–coerced condition and its control would be greater than the difference between the strongly preferred–coerced condition and its control.

Results

Prior to all analyses, sentences with major tracking loss (e.g., as a result of major head movements or blinks) and sentences for which the subject and verb were skipped were excluded (2.2% of the data). For each measure and each region, we subjected the data to separate 2 (coercion: coerced vs. control) \times 2 (preference: strongly preferred vs. weakly preferred) analyses of variance, treating participants (F_1) and items (F_2) as random effects. All analyses are within-participants and -items. Table 1 shows the averages, using participants' means.

First-pass duration analyses of the complement region revealed a significant effect of coercion, with the coerced expressions taking on average 16.5 ms longer to process, $F_1(1, 39) = 5.06, p = .03, MSE = 2,161; F_2(1, 31) = 5.68, p = .02, MSE = 1,359$. The coercion cost for the strongly preferred condition was 17 ms, as compared to 16 ms for the weakly coerced condition. There was also a suggestion in the participants' analysis that the strongly preferred conditions were easier to process than the weakly preferred ones, but this effect was absent in the items' analysis, $F_1(1,$

$39) = 8.99, p = .005, MSE = 1,412; F_2(1, 31) = 1.39, p > .24, MSE = 5,414$. The interaction was not significant, $F_s < .04$ ($M_{int} = -0.6$ ms, 95% confidence interval [CI] = ± 25.1). For the spillover region, first-pass analyses revealed an effect of coercion, with the coerced conditions taking 12.3 ms longer to read (11 ms for the strongly preferred, 13 ms for the weakly preferred condition), $F_1(1, 39) = 3.87, p < .06, MSE = 1,561; F_2(1, 31) = 5.12, p = .03, MSE = 1,016$. There was no effect of preference ($F_s < .51$), nor was there any indication of an interaction ($F_s < .06, M_{int} = 2.6$ ms, 95% CI = ± 21.8).

Second-pass reading time analyses for the complement region showed a significant coercion effect, 36 ms for the strongly preferred and 30 ms for the weakly preferred items, $F_1(1, 39) = 13.97, p = .001, MSE = 3,070; F_2(1, 31) = 14.88, p = .001, MSE = 2,221$. On average, rereading times for the coerced conditions were 32.7 ms longer than for the control conditions. There was a significant effect of preference, $F_1(1, 39) = 8.66, p = .005, MSE = 2,208; F_2(1, 31) = 4.40, p = .04, MSE = 3,156$, with the strongly preferred conditions being read 21.9 ms faster on average than the weakly preferred conditions. Crucially, however, there were no indications of an interaction between coercion and preference ($F_s < .17, M_{int} = -6.8$ ms, 95% CI = ± 34.0). For the spillover region, none of the effects were significant.

Total reading time analyses for the complement region revealed a significant effect of coercion, with the coerced conditions taking 51.9 ms longer than the control conditions did (53 ms for the strongly preferred, 51 ms for the weakly preferred), $F_1(1, 39) = 13.03, p = .001, MSE = 8,262; F_2(1, 31) = 16.98, p < .001, MSE = 4,580$. The participants' analysis also revealed an effect of preference, with the strongly preferred conditions being read 50.5 ms faster than the weakly preferred conditions, but this effect was not significant in the items' analysis, $F_1(1, 39) = 23.51, p < .001, MSE = 4,344; F_2(1, 31) = 3.01, p = .09, MSE = 23,782$. Again, the crucial interaction was not significant ($F_s < .01, M_{int} = -2.3$ ms, 95% CI = ± 49.8). The same pattern was found for the spillover region. There was a significant effect of coercion, with the coerced conditions being read 33.8 ms faster (38 ms for the strongly preferred and 30 ms for the weakly preferred), $F_1(1, 39) = 14.54, p < .001, MSE = 3,138; F_2(1, 31) = 14.32, p = .001, MSE = 3,020$. The effect of preference again did not reach significance in the items' analysis, $F_1(1, 39) = 5.00, p = .03, MSE = 2,418; F_2(1, 31) = 2.43, p = .13, MSE = 4,367$, and there was no interaction between coercion and preference ($F_s < .20, M_{int} = -8.0$ ms, 95% CI = ± 36.5).

The results clearly indicate that complement coercion is costly. The results are consistent with other eye-tracking studies on coercion, with one minor exception. Here, the coercion effect was apparent on what is often considered an "early" processing measure of the complement, first-pass time. Although trends have often been observed in earlier measures (e.g., McElree, Frisson, & Pickering, 2006; Traxler et al., 2002, 2005), reliable effects typically emerge in "later" processing measures (e.g., regression-path, second-pass, total time) on either the complement region or the spillover region. This is the first time that a coercion cost has been observed so early during the eye-movement record. The reason why a coercion effect was observed immediately in the present experiment is not entirely clear, though it might be related to the substantially larger number of data points per participant that we collected (e.g., Traxler et al., 2002, Experiment 1, tested 36 par-

Table 1
Mean Reading Time Durations

Measure	Complement (<i>the novel/the coffee</i>)	Spill-over (<i>as soon</i>)
First-pass duration		
Strongly preferred–coerced	356 (15.0)	299 (10.7)
Strongly preferred–control	339 (13.9)	288 (10.7)
Weakly preferred–coerced	373 (14.6)	303 (9.9)
Weakly preferred–control	357 (14.6)	290 (10.9)
Second-pass duration		
Strongly preferred–coerced	101 (16.0)	71 (9.1)
Strongly preferred–control	65 (11.6)	58 (10.8)
Weakly preferred–coerced	120 (13.5)	77 (8.5)
Weakly preferred–control	90 (14.3)	74 (10.8)
Total reading time		
Strongly preferred–coerced	486 (28.4)	391 (18.0)
Strongly preferred–control	433 (23.4)	353 (17.3)
Weakly preferred–coerced	535 (27.5)	404 (16.4)
Weakly preferred–control	484 (26.5)	374 (17.8)

Note. Reading times are in milliseconds. Standard errors are presented in parentheses.

ticipants and obtained 12 data points per participant for coerced constructions, whereas the present experiment tested 40 participants and collected 36 data points per participant for coerced expressions).

The main effect of preference, as argued before, is not very informative, as it might merely indicate that it is easier to, for example, form a mental model of a teenager reading a novel than of a waitress serving coffee. In any case, the relations expressed in these subsets are so different that direct comparisons between them are not warranted. More interesting, there was no evidence that frequency of interpretation and/or number of different interpretations modulated the coercion cost, as in all measures the differences between the strongly preferred–coerced condition and its control were comparable to the difference between the weakly preferred–coerced condition and its control. As indicated by the very low F values for the interaction term, there was not even a trend in that direction.

We performed a series of regression analyses as a further means to testing the ambiguity hypothesis. One might object to conclusions based on analyses of variance because they depend on artificial cutoff points to dichotomize particular expressions as strongly or weakly biased (whether there is a dominant interpretation or not) or as balanced or unbalanced (whether there is more than one probable interpretation or not). Possibly, we might have missed small effects of either factor by dividing the items into two sets. To address this issue, for each item in the experiment and for each eye-tracking measure, we correlated the observed differences between the coerced and the control conditions with two relevant measures derived from the completion pretest, the item's dominance score (the percentage of time the most dominant response was given) and the number of different responses that were given to an item. If the degree of preference for the dominant interpretation or the number of interpretations of a coerced expression affected reading times, we would expect to find that when the frequency of the dominant interpretation increases, the difference with its control decreases, and that the higher the number of alternative interpretations, the higher the differences in reading times. Crucially, however, Pearson correlations for all measures on both the complement and the spillover region indicated that this interdependence did not exist (all r s < .12, all p s > .34). Hence, the coercion cost does not correlate with the frequency of the dominant interpretation or the number of different interpretations of a coerced expression.

General Discussion

Consonant with other reading time studies (e.g., Lapata et al., 2003; McElree, Frisson, & Pickering, 2006; McElree et al., 2001; Pickering et al., 2005, 2006; Traxler et al., 2002, 2005), we found that expressions requiring complement coercion are more costly to process than control expressions, in which the verb can combine with the complement through basic compositional operations (viz., function application, see Pykkänen & McElree, 2006). Crucially, however, we found that the magnitude of the coercion cost does not depend on the number of different interpretations that comprehenders are likely to assign to an expression and whether there is a dominant interpretation of the expression: Weakly constrained coercions, such as *the director started the script* (Example 6 above), are not more costly to interpret than strongly constrained

coercions, such as *the student finished the essay* (Example 5 above). Consequently, complement coercion effects appear to be fundamentally different from lexical and syntactic ambiguity effects, as the frequency of interpretation has measurable effects in the latter but not in the former. In this respect, complement coercion resembles the processing of polysemous words (words with semantically related interpretations), which also do not show frequency effects in on-line processing (Frisson & Pickering, 1999).

Our results challenge accounts of coercion that would attribute the cost to the need to select an appropriate interpretation from a set of plausible ones that have been generated or to competition between different possible (generated) interpretations. Rather, our results suggest that generating a plausible interpretation, or a set of plausible interpretations, and choosing one do not seem to be costly, as long as some (any) interpretation can be attained that is compatible with the subject of the sentence and other available constraints.

In contrast to such accounts, we have argued that coercion costs reflect enriched compositional operations that are needed to construct an event sense for a complement that is of a different semantic type. That is, in cases such as . . . *finished the essay* or . . . *started the script*, the cost reflects the deployment of operations to construct a semantic representation that can be represented as [*finished/started*[*VERBING the essay/script*]].⁴ As we assume that this type of operation is needed whenever complement coercion is required, this account predicts that coercion costs should be constant across expressions that might vary in how dominant an interpretation is or in how many alternative interpretations are likely.

These findings accord well with the results of Traxler et al. (2005), where it was found that placing the intended activity in a sentence immediately preceding the coerced expression (e.g., *The contractor had been building in the suburbs. That spring, he began a condominium. . .*) did not eliminate the coercion cost. Traxler et al.'s results are directly at odds with accounts that would attribute the coercion cost to retrieving or inferring a plausible activity for the complement noun. However, they are also inconsistent with accounts that would attribute the cost to selecting an activity or competition between alternative activities, as prior processing of the activity implicit in the dominant interpretation of the coerced expression should have increased its salience in the selection process or in the competition with other activities and thereby eliminated the observed coercion cost.

However, our findings are at odds with the results of Martin and Cheng (2006), who found an effect of the strength of the most frequent association in a verb generation task. It is not entirely

⁴ Complement coercion is often viewed as a strictly semantic operation that converts the entity-denoting object into an event description that satisfies the selectional demands of the verb (Jackendoff, 1997; Pustejovsky, 1995). However, inasmuch as these constructions are formed around a verb that selects a verb phrase (VP), complement coercion might also involve more elaborate syntactic operations. For example, coercion might be based on the insertion of a syntactically silent VP node, which converts a syntactic structure from $v_P[V NP]$ into $v_P[V[v_P[V NP]]]$. Although possible, Pykkänen and McElree (2006) showed that this account makes incorrect predictions about the distributional properties of coercion. Hence, current evidence indicates that coercion is a strictly semantic operation.

clear why the two tasks show differential sensitivity to this type of frequency information. However, we note two salient differences between the tasks. First, the verb generation task requires overt production of a verb that is associated with a noun. Frequency information may impact on processes that are specifically involved in overtly selecting and producing a verb form. Second, contextual information is minimal in the verb generation task, and it is possible that without the additional information provided by the sentential context, subjects rely to a greater extent on frequency in formulating their response. Clearly, however, additional research is needed to fully understand the differences between the two tasks.

In attributing the coercion cost to compositional operations, we do not suggest that comprehenders do not have to infer an activity or even to select one from several generated possibilities. To construct a specific interpretation, comprehenders clearly must come up with an appropriate activity for the event sense of the complement, one that is compatible with both the object noun, the agent of the sentence, and with other possible discourse constraints. Our argument is simply that one cannot reduce the cost of coercion to the more general effects of ambiguity or competition between different interpretations, as has been found in lexical and syntactic processing. Put another way, we do not assume that ambiguity has no consequences for the processing of expressions with complement coercion. We believe that it would be possible to construct coercions that would be more effortful to process than the materials contrasted in our experiments, if the intended activities were too vague given available constraints. For example, a sentence such as *the artist started the rock* might be very costly to process without a constraining context (cf. *For his upcoming show, the artist decided to paint natural objects. Before he started the rock. . .*). Although these types of constructions might engender costs due to the difficulty of selecting a possible activity, or even from competition among possible activities, we assume that those costs would simply add to the basic (and perhaps mandatory) costs associated with constructing an eventive representation of the complement.

Practically speaking, we suspect that the materials explored here and in other coercion studies, where there appears to be no measurable costs associated with generating and selecting an activity for the event sense of the complement, are quite representative of the expressions that comprehenders are likely to encounter in natural language settings. Corpus analyses of complement coercion (Briscoe, Copestake, & Boguraev, 1990; Lapata & Lascarides, 2003; Lapata et al., 2003) indicate that when the activities performed on the complement noun are atypical—for example, translating rather than reading or writing a book—language users rarely elide the activity, but instead produce an expression with an overt event structure, for example, *the man began translating the book*. In contrast, when the intended activity is commonly associated with the complement noun, eliding the intended activity with an expression such as *began the book* is greater than 9:1 times more likely than overtly expressing the full structure such as *. . . began reading the book*. Hence, cases where comprehenders may need to engage in taxing inferential processes to derive the intended activity for the event sense of the complement may be quite rare. These corpus findings converge nicely with studies showing that the robust effects of complement coercion found with different materials and methods are not likely to reflect costs

associated with deriving an activity for the event sense, but they rather reflect taxing compositional operations that are needed to build an event sense for what is typically a non-event-denoting complement.

References

- Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, *30*, 191–238.
- Binder, K. S., & Morris, R. K. (1995). Eye movements and lexical ambiguity resolution: Effects of prior encounter and discourse topic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 1186–1196.
- Briscoe, E. J., Copestake, A., & Boguraev, B. (1990). Enjoy the paper: Lexical semantics via lexicology. *Proceedings of the 13th International Conference on Computational Linguistics*, *2*, 42–47.
- Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, *27*, 429–446.
- Francis, W. N., & Kucera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: 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*, 1347–1365.
- Frisson, S., & Pickering, M. J. (2007). The processing of familiar and novel senses of a word: Why reading Dickens is easy but reading Needham can be hard. *Language and Cognitive Processes*, *22*, 595–613.
- Frisson, S., Rayner, K., & Pickering, M. J. (2005). Effects of contextual predictability and transitional probability on eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 862–877.
- Gennari, S., & Poeppel, D. (2003). Processing correlates of lexical semantic complexity. *Cognition*, *89*(1), B27–41.
- Green, M. J., & Mitchell, D. C. (2006). Absence of real evidence against competition during syntactic ambiguity resolution. *Journal of Memory and Language*, *55*, 1–17.
- Hess, D. J., Foss, D. J., & Carroll, P. (1995). Effects of global and local context on lexical processing during language comprehension. *Journal of Experimental Psychology: General*, *124*, 62–82.
- Jackendoff, R. (1997). *The architecture of the language faculty*. Cambridge, MA: MIT Press.
- Jackendoff, R. (2002). *Foundations of language*. New York: Oxford University Press.
- Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*, 259–284.
- Lapata, M., Keller, F., & Scheepers, C. (2003). Intra-sentential context effects on the interpretation of logical metonymy. *Cognitive Science*, *27*, 649–668.
- Lapata, M., & Lascarides, A. (2003). A probabilistic account of logical metonymy. *Computational Linguistics*, *29*(2), 263–317.
- MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*, *9*, 157–201.
- Martin, R. C., & Cheng, Y. (2006). Selection demands versus association strength in the verb generation task. *Psychonomic Bulletin & Review*, *13*, 396–401.
- McElree, B., Frisson, S., & Pickering, M. J. (2006). Deferred interpretations: Why starting Dickens is taxing but reading Dickens isn't. *Cognitive Science*, *30*, 115–124.
- McElree, B., Pylkkänen, L., Pickering, M. J., & Traxler, M. (2006). The time course of enriched composition. *Psychonomic Bulletin & Review*, *13*, 53–59.

- McElree, B., Traxler, M., Pickering, M. J., Seely, R., & Jackendoff, R. (2001). Reading time evidence for enriched composition. *Cognition*, 78, B17–25.
- McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, 38, 283–312.
- Nunberg, G. (2004). The pragmatics of deferred interpretation. In L. Horn & G. Ward (Eds.), *Handbook of pragmatics* (pp. 344–364). Oxford, England: Blackwell.
- Pickering, M. J., McElree, B., Frisson, S., Chin, L., & Traxler, M. (2006). Underspecification and aspectual coercion. *Discourse Processes*, 42, 131–155.
- Pickering, M. J., McElree, B., & Traxler, M. (2005). The difficulty of coercion: A response to de Almeida. *Brain & Language*, 93, 1–9.
- Pickering, M. J., & Traxler, M. (1998). Plausibility and recovery from garden paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 940–961.
- Pustejovsky, J. (1995). *The generative lexicon*. Cambridge, MA: MIT Press.
- Pylkkänen, L., Llinás, R., & McElree, B. (in press). An MEG study of silent meaning. *Journal of Cognitive Neuroscience*.
- Pylkkänen, L., & McElree, B. (2006). The syntax-semantics interface: On-line composition of sentence meaning. In M. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (2nd ed.). New York: Elsevier.
- Rayner, K., Binder, K. S., & Duffy, S. A. (1999). Contextual strength and subordinate bias effect. *Quarterly Journal of Experimental Psychology*, 52A, 841–852.
- 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.
- Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic ambiguity resolution in discourse: Modeling the effects of referential context and lexical frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 1521–1543.
- Tabor, W., & Tanenhaus, M. K. (1999). Dynamical models of sentence processing. *Cognitive Science*, 23(4), 491–515.
- Thompson-Schill, S. L., D’Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences, USA*, 94, 14792–14797.
- Traxler, M., McElree, B., Williams, R. S., & Pickering, M. J. (2005). Context effects in coercion: Evidence from eye-movements. *Journal of Memory and Language*, 53, 1–25.
- Traxler, M., Pickering, M. J., & McElree, B. (2002). Coercion in sentence processing: Evidence from eye-movements and self-paced reading. *Journal of Memory and Language*, 4, 530–547.
- University of Massachusetts Eyetracking Lab. (2007). EyeTrack, EyeDoctor, and EyeDry [Software]. Available from <http://www.umass.edu/psychology/div2/eyelab/>

Appendix Experimental Items

The items are divided up into strongly preferred and weakly preferred subsets. For each sentence, the verb for the coerced condition appears first, followed by the control verb. After each sentence, the number of different interpretations for the coerced condition is given between brackets, followed by the most common interpretation, followed by the percentage that interpretation was provided in the pretests.

Strongly Preferred

1. The wife [preferred | wore] the hat that was made out of velvet. (4; wearing: 88%)
2. The sister [preferred | wore] the sandals because it was such a beautiful day. (2; wearing: 96%)
3. The sick woman [resisted | did not take] the painkiller although the doctor urged her to take it. (3; taking: 91%)
4. The kids [completed | ate] the appetizers that the woman had specially ordered. (5; eating: 83%)
5. The protester [resisted | did not wear] the handcuffs although the police attempted arrest. (4; wearing: 88%)
6. The newborn [enjoyed | drank] the milk wholeheartedly and then went back to sleep. (3; drinking: 87%)
7. The hiker [attempted | climbed] the mountain immediately after he had a power bar. (3; climbing: 83%)
8. The man [completed | read] the manual that came with the new DVD player. (2; reading: 96%)
9. The pastor [endured | wore] the collar although it was really itchy. (2; wearing: 96%)
10. The professor [started | read] the book after he came back from his office. (2; reading: 92%)
11. The chef [started | cooked] the entrée and she remembered to put salt on it. (6; cooking: 87%)
12. The teenager [began | read] the novel as soon as he got to his room upstairs. (3; reading: 87%)
13. The schoolboy [completed | read] the comic even though his mother wanted him to study. (2; reading: 96%)
14. The lady [enjoyed | drank] the tea that was made out of Chinese leaves. (3; drinking: 96%)

(Appendix continues)

15. The gentleman [finished | ate] the sandwich as soon as he came into the office. (3; eating: 91%)
16. The performer [endured | wore] the costume that he had bought during the sales. (1; wearing: 100%)
17. The uncle [preferred | drank] the scotch but the child was not allowed touch it. (3; drinking: 96%)
18. The lawyer [preferred | drove] the convertible because the weather was balmy. (3; driving: 96%)
19. The handyman [completed | fixed] the sink of the huge mansion on the hill. (4; fixing: 83%)
20. The comedian [preferred | used] the microphone because there was a lot of noise. (4; using: 87%)
21. The youth [endured | wore] the braces although he hated the feel of it. (4; wearing: 83%)
22. The quarterback [enjoyed | drank] the champagne but his coach told him to lay it off. (6; drinking: 83%)
23. The professional [endured | wore] the suit that was bought for him by his mom. (2; wearing: 96%)
24. The bridesmaid [endured | wore] the heels during the entire wedding of her friend. (3; wearing: 96%)
25. The student [finished | wrote] the essay while his girlfriend was already asleep. (3; writing: 92%)
26. The policeman [endured | wore] the uniform while the sun was beating down outside. (1; wearing: 100%)
27. The toddlers [enjoyed | ate] the cupcakes while the caretaker was having a coffee. (5; eating: 83%)
28. The baker [finished | iced] the cake as soon as he put on his tunic. (4; baking: 87%)
29. The grandpa [tried | smoked] the pipe but it didn't make him feel any happier. (5; smoking: 87%)
30. The passenger [completed | read] the magazine after taking his seat in business class. (3; reading: 91%)
31. The customer [enjoyed | ate] the nachos together with a bit of guacamole. (2; eating: 96%)
32. The Amish man [resisted | did not use] the telephone but his brother thought it was silly. (2; using: 91%)
- Weakly Preferred*
1. The workmen [finished | fixed] the roof that was made out of hay. (9; fixing: 36%)
2. The woman [started | weeded] the garden because it was something she loved to do. (13; weeding: 29%)
3. The adolescent [tried | used] the surfboard although his mother told him not to do it. (10; using: 39%)
4. The guests [tried | used] the comforter that the woman had laid out for them. (10; using: 46%)
5. The yachtsman [tried | drove] the catamaran although the owner felt nervous about it. (12; driving: 27%)
6. The houseguests [enjoyed | ate] the eggs wholeheartedly and thanked their host. (7; eating: 52%)
7. The scribe [began | wrote] the documents immediately after he found his fountain pen. (10; writing: 50%)
8. The banquet-goers [tried | ate] the venison that came with a cranberry sauce. (11; eating: 52%)
9. The director [started | wrote] the script although he was feeling sick. (8; writing: 35%)
10. The architect [started | designed] the house after he came back from his holidays. (4; designing: 46%)
11. The secretary [began | wrote] the memo and she remembered to call the boss. (7; writing: 50%)
12. The waitress [started | served] the coffee as soon as she returned to the counter. (8; serving: 29%)
13. The skateboarder [attempted | jumped] the ramp even though it looked extremely dangerous. (12; jumping: 23%)
14. The boss [enjoyed | read] the card that was made out of Japanese paper. (9; reading: 48%)
15. The manufacturer [completed | made] the luggage as soon as he found some extra time. (11; making: 38%)

16. The tattoo artist [continued | drew] the outline that he had seen on-line somewhere. (12; drawing: 46%)
17. The stylist [started | made] the braid but the child was crying incessantly. (11; making: 50%)
18. The pharmacist [finished | filled] the prescription because the customer was waiting. (7; filling: 58%)
19. The horticulturist [preferred | smelled] the roses of the huge garden in the back. (11; smelling: 39%)
20. The publisher [started | read] the manuscript because there was a deadline to make. (7; reading: 52%)
21. The celebrity [enjoyed | read] the email although he was not feeling well. (5; reading: 54%)
22. The fisherman [preferred | caught] the swordfish but his mates couldn't care less. (6; catching: 57%)
23. The dog [enjoyed | chewed] the bone that was thrown to him by his owner. (8; chewing: 57%)
24. The tycoon [enjoyed | drove] the yacht during the entire holidays in the Bahamas. (14; driving: 25%)
25. The Northerner [tried | ate] the grits while his friends were laughing at him. (8; eating: 55%)
26. The pianist [began | played] the symphony while the audience was listening attentively. (6; playing: 54%)
27. The landscape artist [finished | painted] the seashore while the girls were watching him. (10; painting: 58%)
28. The leather worker [began | made] the shoe as soon as he found some time. (14; making: 29%)
29. The restaurant patrons [tried | ate] the steak but it didn't taste very good. (6; eating: 52%)
30. The artist [began | painted] the portrait after taking a short nap on his leather couch. (6; painting: 58%)
31. The Olympian [enjoyed | won] the medal together with his tennis partner. (8; winning: 36%)
32. The rugrat [tried | sucked] the pacifier but his brother took it away from him. (9; sucking: 57%)

Received November 23, 2006

Revision received June 28, 2007

Accepted July 3, 2007 ■