Available online at www.sciencedirect.com
ScienceDirect
Brain & Language 107 (2008) 44–61
Brain & Language
www.elsevier.com/locate/b&l

The cost of question concealment: Eye-tracking and MEG evidence

Jesse Harris [a,b], Liina Pylkkänen [a,b,*], Brian McElree [a], Steven Frisson [c]

[a] *Department of Psychology, New York University, 6 Washington Place, Room 870, New York, NY 10003, USA*
[b] *Department of Linguistics, New York University, USA*
[c] *School of Psychology, University of Birmingham, UK*

Accepted 14 September 2007
Available online 29 October 2007

Abstract

Although natural language appears to be largely compositional, the meanings of certain expressions cannot be straightforwardly recovered from the meanings of their parts. This study examined the online processing of one such class of expressions: *concealed questions*, in which the meaning of a complex noun phrase (*the proof of the theorem*) shifts to a covert question (*what the proof of the theorem is*) when mandated by a sub-class of question-selecting verbs (e.g., *guess*). Previous behavioral and magnetoencephalographic (MEG) studies have reported a cost associated with converting an entity denotation to an event. Our study tested whether both types of meaning-shift affect the same computational resources by examining the effects elicited by concealed questions in eye-tracking and MEG. Experiment 1 found evidence from eye-movements that verbs requiring the concealed question interpretation require more processing time than verbs that do not support a shift in meaning. Experiment 2 localized the cost of the concealed question interpretation in the left posterior temporal region, an area distinct from that affected by complement coercion. Experiment 3 presented the critical verbs in isolation and found no posterior temporal effect, confirming that the effect of Experiment 2 reflected sentential, and not lexical-level, processing.

Keywords: Semantic composition; Coercion; Concealed questions; N400; AMF; MEG; Eye-tracking

1. Introduction

Linguists and philosophers often assert that natural language is by and large *compositional*; the meaning of an expression is constructed from the meanings of the smaller elements within it (Fodor & Lepore, 2002; Heim & Kratzer, 1998; Montague, 1970). A novel phrase can be uttered and understood precisely because its meaning can be calculated from the words that comprise it, even if they have never before been combined together in such a way.

Sometimes, however, the meaning of an expression does not appear to be directly composed from its constituent parts. Consequently, semantic theories commonly include non-syntactic operations that *enrich* meaning by contributing extrasyntactic content that allows constituents to compose when they are otherwise semantically incompatible (Groenendijk & Stokhof, 1989; Hendriks, 1988; Partee, 1986, 1995; Partee & Rooth, 1983). Such 'type-shifting' operations violate strong compositionality because they contribute to the meaning of an expression without overtly manifesting within the expression itself.

Enriched meanings and associated processing costs have received much attention in recent psycholinguistic literature (for a review, see Pylkkänen & McElree, 2006). One such process of semantic enrichment is known as *coercion*, in which the meaning of an incompatible linguistic element is *coerced* to another meaning that better coheres with the semantic context (Jackendoff, 1997; Pustejovsky, 1995; Pustejovsky & Bouillon, 1995). Although natural language exhibits numerous types of coercions, only two variants of coercion have been investigated in psycholinguistic studies

* Corresponding author. Address: Department of Psychology, New York University, 6 Washington Place, Room 870, New York, NY 10003, USA.
E-mail address: liina.pylkkanen@nyu.edu (L. Pylkkänen).

0093-934X/$ - see front matter
doi:10.1016/j.bandl.2007.09.001

so far. One is so-called *aspectual coercion* (Pinango & Zurif, 2001; Piñango, Zurif, & Jackendoff, 1999; Todorova, Straub, Badecker, & Frank, 2000; Pickering, McElree, Frisson, Chin, & Traxler, 2006), in which a punctual action (e.g., *hop*) requires an iterative interpretation in durative contexts (e.g., *The robin hopped all afternoon*). The other is *complement coercion* (McElree, Traxler, Pickering, Seely, & Jackendoff, 2001; McElree, Frisson, & Pickering, 2006a; McElree, Pylkkänen, Pickering, & Traxler, 2006b; Traxler, Pickering, & McElree, 2002; Traxler, McElree, Williams, & Pickering, 2005), in which an object-denoting noun (e.g., *book*) is interpreted as an event to combine with an event-selecting verb (e.g., *The author began the book*).

In this research we investigated the processing of *concealed questions* (Grimshaw, 1979; Heim, 1979), a noncompositional construction that has not yet been studied experimentally. Like complement coercion, concealed questions (CQs) involve a semantic type-mismatch between a verb and its noun phrase (NP) argument. But unlike complement coercion, concealed questions involve shifting the meaning of the NP argument to a question-like meaning, as opposed to an event. The verb *guess* is a typical question-selecting predicate which allows as its complement both overt and concealed questions, as illustrated in (1).

(1) a. The announcer guessed [who the winner of the contest was] (Overt question)
b. The announcer guessed [the winner of the contest] (Concealed question)

Concealed questions are syntactically similar to expressions such as *the announcer praised the winner of the contest*, where the direct object receives a straightforward interpretation as the individual who won the contest. But the question-like interpretation of concealed questions can be diagnosed in various ways. For example, different types of anaphora refer to CQ interpretations than to regular nouns (Romero, 2005). In English, a noun denoting a person is typically referred to by a pronoun agreeing with the noun in gender and number (*he*, *she*), as shown in (2a). However, CQs are referred to by the neuter pronoun, (2b), which is also used for sentential complements, (2c). Thus, with respect to anaphora, CQs behave like sentences and not like individuals. The sentences in (3) and (4) further illustrate this property with examples from Nathan (2006) and Romero (2005), respectively.

(2) a. Individual complement:
The announcer praised [the winner of the contest] and the judges praised her/*it too.
b. CQ complement:
The announcer guessed [the winner of the contest] and the judges guessed *her/it too.
c. Sentential complement:
The announcer guessed [who the winner of the contest was] and the judges guessed *her/it too.

(3) a. Individual complement:
Kim introduced Sam to [the governor of California], so now Sam knows him/*it too.
b. CQ complement:
Kim told Sam [the governor of California], so now Sam knows it/*him too. (Nathan, 2006)

(4) CQ complement:
John guessed the winner of the Oscar for best actress before I guessed it/*her. (Romero, 2005)

What kind of "sentential" meanings do CQs convey? Nathan (2006) has argued that the interpretations of CQs are limited to identity questions. In other words, a sentence such as *I told Leslie the capital of Vermont* requires a situation where the speaker told Leslie what the capital of Vermont is (i.e., the speaker revealed to her the identity of that city), and is inappropriate if the speaker, for example, told Leslie *where* the capital of Vermont is.

Further, although CQs can intuitively be paraphrased as questions, they do not technically behave as questions either with respect to their syntax or semantics. Specifically, both the distribution and the truth conditions of CQs differ from those of overt question complements (Frana, 2006; Greenberg, 1977; Heim, 1979; Nathan, 2006). Thus, instead of a question meaning, recent theories have analyzed CQs as involving a propositional meaning (Nathan, 2006; Romero, 2006). Under this hypothesis, the meaning of a sentence such as *the doctor guessed the name of the child* would be better paraphrased as *the doctor guessed that the name of the child was x* than by an overt question complement. The formal details of various contemporary analyses differ (see in particular Romero, 2006, for a systematic comparison), but importantly for our purposes, they all involve an operation that shifts either the meaning of the complement NP (Frana, 2006; Romero, 2006) or the head noun of that NP (Nathan, 2006). In this article, we will use the term "question-like" to describe the shifted meaning of CQs.

In sum, concealed questions and complement coercion both involve shifting an NP-denotation to a more complex type. But in the case of complement coercion, the output of the shift is an event predicate, whereas CQs involve a shift to a question-like meaning. Complement coercion has been shown to be costly across numerous experimental methods, including self-paced reading (McElree et al., 2001; Traxler et al., 2002), eye tracking (Traxler et al., 2002), speed-accuracy tradeoff measures (McElree et al., 2006b) and magnetoencephalography (MEG) (Pylkkänen & McElree, 2007). In eye-tracking, coerced constructions (*The author began the book*), have been shown to engender longer gaze fixations on the noun (McElree et al., 2001) or spillover regions directly following the noun (Traxler et al., 2002) than controls (*The author wrote the book*). In MEG, coerced nouns have elicited reliably larger amplitudes in the Anterior Midline Field (AMF), generated in the ventromedial prefrontal cortex at 350–450 ms (Pylkkänen & McElree, 2007).

Although there is a detailed body of results pertaining to the processing of complement coercion, it is at present unclear whether the effects associated with this construction generalize to other cases of type-mismatch. The one other coercion type that has been studied experimentally, aspectual coercion, elicits a rather different pattern of behavioral effects than complement coercion (Pickering et al., 2006), but these two constructions are representationally so different (Pylkkänen & McElree, 2006) that there are many possible reasons for their distinct effects on processing. Concealed questions, on the other hand, are not obviously different from complement coercion: It is perfectly plausible that the language processor would have a general device for converting NP-denotations to more complex types. In fact, most semantic theories that involve NP type-shifting rules do not explicitly classify them into qualitatively distinct operations, which is compatible with the notion of a generic processing mechanism. Under this type of theory, question concealment should affect processing in ways similar to complement coercion.

However, it is also plausible that complement coercion and concealed question formation have no common mechanisms. In other words, converting entity-denotations, such as *the book*, to activities associated with those entities (*reading/writing the book*) could easily engage entirely different resources and representations from the operations by which question-like meanings are generated in concealed questions. If so, concealed questions should elicit different effects from complement coercion, at least at the neural level.

Whether type-shifting operations divide into empirically distinguishable classes is a question that formal semantic theories have been relatively silent on (although see Winter, 2007). While it is sometimes acknowledged that the term "type-shifting" is used for a deceptively wide variety of phenomena, differing in many properties (e.g., Dowty, 2007), there has been little empirical research directly investigating whether type-shifting is simply an umbrella term for a host of qualitatively different phenomena or whether all type-mismatch resolution shares some common mechanism(s). The goal of the present work is to contribute to a general research program where processing measures are used to address this question (Pylkkänen & McElree, 2006).

The processing of concealed questions was investigated both with eye-tracking and MEG. With eye-tracking, we first aimed to establish whether question concealment is associated with delayed reading times behaviorally. The subsequent MEG experiment investigated the neural bases of concealed question interpretation. Specifically, we tested whether the AMF response, found to be modulated by complement coercion in our previous study (Pylkkänen & McElree, 2007), is also sensitive to question concealment.

2. Experiment 1

As an initial means of determining whether the processing of concealed questions engenders a cost similar to complement coercion, we monitored participants' eye-movements while they processed concealed questions and control expressions. Eye-tracking studies on complement coercion have generally shown effects on later measures such as first-pass regression, second pass time and total time (McElree et al., 2006a; Pickering, McElree, & Traxler, 2005; Traxler et al., 2005; Traxler et al., 2002) with the exception of one study, where an early effect on first-pass reading time effect was obtained (Frisson & McElree, in press). Our aim was to assess how the processing of concealed questions compares with this reading time profile.

The eytracking and MEG studies on complement coercion that formed the background for our study (Traxler et al., 2002, 2005; Pylkkänen & McElree, 2007) have contrasted expressions such as *the author began the book* with control expressions such as *the author wrote the book*, where the verb directly selects for an entity-denoting argument. A similar comparison with concealed questions would contrast a CQ selecting verb with an entity selecting verb and hold the object NP constant. The NPs that have CQ interpretations, however, differ from those allowing complement coercion in a way that matters for the manipulation. Specifically, CQs are typically built on *relational* nouns, such as *price*, *distance*, *age* and so forth (Caponigro & Heller, 2007; Nathan, 2006). This is evidenced by the contrast between well-formed expressions such as *the contestant guessed the population of California* and ill-formed ones such as *the contestant guessed the coast of California*, in that *population* is a relational noun, while *coast* is a non-relational noun. Without a supporting context, relational nouns can sound awkward if neither of their arguments is mentioned, as illustrated by the contrast between (5a), where *population* occurs without either of its arguments, and the examples in (5b) and (5c), where either one or both of the arguments of the relation are mentioned.

(5) a. ??The contestant guessed the **population**. (no explicit arguments)
b. The contestant guessed the **population** of *California*. (one explicit argument)
c. The contestant guessed the **population** of *California* to be *34 million*. (two explicit arguments)

Consequently, it was necessary for the test items to involve a complex NP, such as the one in (5b); (5c) involves a full propositional argument and hence no type-mismatch. However, for the purposes of our neuromagnetic study, i.e., Experiment 2, (5b) has the unfortunate property that the element eliciting the hypothesized cost is complex. This could spread a potential effect over several words, complicating the analysis. For this reason, we used passive versions of sentences such as (5b), which should make the type-shifting cost maximally local to the verb. Thus we contrasted passivized CQ sentences with expressions involving an entity selecting verb and no type-mismatch, as illustrated in (6).

(6) a. The design of the movie set was *explained* by the artist. (CQ-verb)
b. The design of the movie set was *sketched* by the artist. (Entity selecting verb)

A downside of using the passive voice was that it made our materials less parallel to previous studies on complement coercion, which have so far only used the active voice. In the active, type-shifting is triggered by the presentation of the argument that needs to be shifted. In the passive, on the other hand, the presentation of the passivized verb triggers type-shifting of a previously presented argument. Thus a question relevant to the comparison between the present CQ study and previously reported complement coercion results is whether type-shifting of an argument that is currently being processed and type-shifting of a previously presented argument employ the same mechanism. There are no theories of type-shifting where this type of difference would alter the nature of the type-shifting mechanism. Empirical support for this conclusion is also provided by two recent studies on expressions such as *the article was finishable* (vs. *the editing was finishable*), which are similar to passives in that the predicate *finishable* triggers type-shifting of the previously presented entity-denoting NP. Behaviorally, this type of expressions were found to elicit the same size processing delay as active expressions involving complement coercion (*the professor finished the article*) (McElree et al., 2006b). Further, a subsequent MEG study localized the effect in anterior midline regions, consistent with previous findings for active complement coercion sentences where the measurement was performed on the direct object noun (Pylkkänen & McElree, 2007). These results strongly suggest that the same mechanisms are used whether or not type-shifting is performed on the target item or on a previously presented item.

Another factor affecting the construction of the materials was that CQ expressions are often ambiguous between CQ and non-CQ readings. The sentences in (7), from Frana (2006), exemplify this ambiguity.

(7) a. Meg has forgotten the capital of Italy.
b. Kim knows the governor of California.
c. They revealed the winner of the contest.
(Frana, 2006, example 1)

In addition to the CQ-meaning that Meg has forgotten that the capital of Italy is Rome, (7a) can also mean that Meg has lost her memory of Rome; imagine for example that she is compiling a list of all European capitals and fails to include Rome on the list. Similarly, (7b) can mean that Kim personally knows the individual who is the governor of California, i.e., Arnold Schwarzenegger. Finally, in addition to the CQ reading 'they revealed who the winner of the contest was', (7c) would also be true in a situation where everyone knows who the winner of the contest is but that individual is physically revealed from under a cover (Frana, 2006). That the sentences in (7) have non-CQ readings is also witnessed by their well-formedness with non-relational object NPs. The examples in (8) replicate those in (7) except that the object NPs are clearly non-relational proper names, which do not, in general, support a question paraphrase (see Romero, 2006 for discussion).

(8) a. Meg has forgotten Sam.
b. Kim knows Sam.
c. They revealed Sam.

The ambiguity of the examples in (7) has to do with the semantics of the verbs: one can forget, know and reveal both facts as well as individuals. It was important that our materials did not involve this ambiguity and therefore verbs that allowed for individuals as their semantic argument were avoided (see Appendices). Verbs used in our study were predicates such as *decide* and *estimate*, which disallow a non-relational proper name as their object, as shown in (9):

(9) a. They decided/estimated the size of the conference room.
b. #They decided/estimated John Smith.

Finally, we needed to assess whether any effect elicited by the contrast in (6) could be due not to the type-mismatch present in the CQ expressions, but simply to the contrast in the meanings of the verbs. Consequently, in addition to maximally matching the critical verbs for lexical properties, we also conducted a lexical decision study on the verbs themselves, with a simultaneous MEG recording (Experiment 3). Crucially, the aim of our research was to identify processing effects of CQ-interpretation that were *not* explainable as differences in the comprehension of the verbs themselves. As the results of Experiment 3 show, the CQ and control verbs did not differ in lexical decision times, suggesting a lack of a behavioral effect based on verb meaning alone. Further, the MEG effects obtained for the sentential stimuli of Experiment 2 were absent in the lexical decision experiment (Experiment 3), where the verbs were presented in isolation.

2.1. Methods

2.1.1. Participants

Forty native English speakers from New York University participated in our eye-tracking study. They were paid $10 for one 45-min session. All participants had normal or corrected-to-normal vision.

2.1.2. Materials

As described above, the critical contrast consisted of 28 pairs of passive sentences such as (10)–(11). The full set of sentences can be found in Appendix A.

(10) The proof of the theorem was guessed by the mathematician. (CQ)
(11) The proof of the theorem was erased by the mathematician. (Control)

We attempted to match the average length and frequency of the verbs in the critical comparison, but this was difficult given the limited number of appropriate CQ verbs. The CQ verbs used in Experiment 1 were slightly longer than the control verbs (8.1 vs. 7.3 characters, respectively, $t(27) = 2.87$, $p < .05$), but also slightly more frequent (74.8 vs. 53.5 occurrences per million, based on Francis & Kučera (1982), ($t(27) = 1.92$, $p = .07$)).

In addition, due to the scarcity of CQ verbs, they were repeated 1.3 times on average (separated by 49 filler sentences on average), which might have speeded up their reading times slightly. Hence, the putative disadvantage CQ verbs have with respect to length is likely being offset by an advantage in frequency and a possible repetition advantage. In any case, we will address this putative length and frequency confound in Experiment 3, where the critical verbs are presented in isolation.

The experimental stimuli were mixed with 216 filler sentences of varying syntactic constructions and length, testing other hypotheses which will not be discussed here. The sentences were presented in a fixed random order. For counterbalancing purposes, it was necessary that each experiment within the entire set of materials contain four conditions. To achieve this, a condition that manipulated the meaning of the subject was added as a pilot (e.g., *the success of the theorem was guessed...*) as well as a condition involving an overt question complement although not syntactically or semantically an identity question (*how to solve the theorem was guessed...*). The latter could have potentially provided a better control condition than the entity-denoting predicate condition, but unfortunately these overt question stimuli proved excessively difficult to interpret. Reading times were extremely slow and a number of participants found these sentences grammatically incorrect and/or non-sensical. The experimental stimuli were divided into 8 lists (4 lists were the reverse of the other lists), with 5 participants randomly assigned to each list. This way, each participant saw 7 CQ sentences of the form in (11) and 7 control items of the form in (11).

2.1.3. Procedure

Participants were run individually on a SensoriMotor Instruments EyeLink I head-mounted eye-tracker apparatus and presentation software. The eye cameras recorded eye movements and fixations every 4 ms. Viewing was binocular, but only data from the eye that was calibrated best was used in the analyses. Screen resolution was 1600×1200 pixels and sentences were presented in fixed font, each letter being 18 pixels wide and 33 pixels high, with a maximum of 80 characters per line. Participants were seated 71 cm from the display monitor. With this setup, 1° of visual angle corresponded to 2.7 characters. A chin rest was used to reduce head movements.

Participants were encouraged to read for understanding and to read at a normal pace. Once they finished reading a trial, they pressed a button on a button box to make the sentence disappear from the screen. Comprehension questions were asked after 50% of the trials (counterbalanced across conditions). Half of the questions required a yes response, half a no response. Participants answered the questions by means of two buttons on the game console. Accuracy was high at 90%.

A calibration procedure was performed at the beginning of the experiment and was repeated whenever the experimenter felt necessary. Before each trial, a fixation box coinciding with the position of the first letter of the upcoming sentence was presented on the screen. This box served as a trigger, with the sentence only being displayed if the fixation was close enough to the middle of the fixation box. After each trial, a drift correction was performed for a fixation point in the middle of the screen. The entire experiment lasted about 45 min.

2.1.4. Eye-tracking analyses

We report analyses on 2 regions: the past participle (e.g., *guessed*) and a spill-over region consisting of the following two words (*by the*). The following standard measures are discussed: *first-pass duration* (the summed fixation times on a region before leaving that region to the left or the right), *first-pass regressions* (percentage of regressions out of a region following a first-pass fixation), and *total reading time* (the sum of all fixations in a region). Fixations less than 100 ms and over 1200 ms were excluded from the analyses. We set the maximum cut-off at 800 ms for first-pass duration and at 1200 ms for total time duration. Analyses with higher cut-offs did not change the pattern of results. The reported means and analyses for first-pass duration and first-pass regressions are based on reading times excluding trials on which a region was skipped on first pass.

2.2. Results and discussion

Prior to all analyses, we eliminated sentences with major track losses, due to head movements or blinks, and sentences for which the first part (i.e., the subject plus the auxiliary verb) was skipped. This resulted in the elimination of less than 2% of the data. For each measure and each region, we subjected the data to separate one-way Analyses of Variance (ANOVAs), treating participants (F_1) and items (F_2) as random effects. All analyses are within-partic-

Table 1
Experiment 1: mean reading time durations and percentage of first-pass regressions

Measure	Verb region	Spill-over
First-pass reading time		
Control	278 (9.9)	283 (12.8)
CQ	304 (11.7)	287 (15.3)
First-pass regression		
Control	7.9 (1.9)	11.0 (2.4)
CQ	9.8 (2.7)	15.7 (2.5)
Total reading time		
Control	371 (13.5)	371 (20.1)
CQ	420 (17.1)	396 (19.6)

Note. CQ, concealed question condition. Reading times are in milliseconds, first-pass regressions in percentages. Standard errors are in parentheses.

ipants and items. Table 1 presents the averages, using participants' means.

First-pass duration analyses revealed a significant difference between the CQ and the control condition on the verb region, with the CQ verb taking on average 26 ms longer to process than the control verb [$F_1(1,39) = 9.91$, $p < .01$, $MSE = 1383$; $F_2(1,27) = 5.29$, $p < .05$, $MSE = 1498$]. The 4 ms difference in the spill-over region was not significant. First-pass regression analyses did not show a significant difference in the percentage of first-pass regressions out of the verb region. For the spill-over region, a higher percentage of first-pass regressions was observed for the CQ sentences, but this difference was only marginally significant in the participants' analysis [$F_1(1,39) = 3.65$, $p < .07$, $MSE = 124$; $F_2(1,27) = 1.44$, $p > .24$, $MSE = 199$]. Finally, the total time analyses showed a significant effect on the verb region, with reading times for the CQ verbs 49 ms longer than the control verbs [$F_1(1,39) = 10.41$, $p < .01$, $MSE = 4583$; $F_2(1,27) = 4.77$, $p < .05$, $MSE = 6562$]. The 25 ms difference on the spill-over region was not significant.

The results clearly indicate that, similarly to complement coercion, concealed questions require additional processing time compared to control sentences. The effect emerged immediately during first-pass processing on the question-selecting verb itself, which is earlier than what most studies have reported for complement coercion (McElree et al., 2005, 2006; Pickering et al., 2005; Traxler et al., 2002, 2005). This could be indicative of distinct mechanisms for the two different types of type-mismatch. In Experiment 2, we used MEG to determine to what extent the concealed question processing cost can be localized in specific areas of the brain and to what extent these areas overlap with the ventromedial prefrontal areas implicated for complement coercion.

3. Experiment 2

MEG measures fluctuations in the magnetic fields generated by neural activity with an excellent combination of spatial and temporal resolution (Hämäläinen, Hari, Ilmoniemi, Knuutila, & Lounasmaa, 1993). As described above, MEG has previously been used to investigate the neural bases of complement coercion, findings indicating that expressions such as *begin the book* elicit larger amplitudes than expressions such as *write the book* in the Anterior Midline Field (AMF) at 350–450 ms after the onset of the entity-denoting noun (Pylkkänen & McElree, 2007). The same study found that anomalous verb-noun combinations (*delight the book*) elicited no AMF effect but instead an effect in left posterior temporal regions, suggesting that the operations required by complement coercion are distinct from the processing of contextually implausible words. Plausibility effects have been heavily studied in ERPs, where the amplitude of the N400 deflection inversely correlates with semantic expectancy (Kutas & Hillyard, 1984). N400 effects have consistently localized to the left superior temporal gyrus and surrounding areas in MEG (Halgren et al., 2002; Helenius, Salmelin, Service, & Connolly, 1998; Simos, Basile, & Papanicolaou, 1997).

We hypothesized that if concealed questions are resolved by a similar mechanism as complement coercion, the AMF should show increased amplitudes to question-selecting passive verbs such as *guessed* when preceded by a simple non-question NP. The materials of our MEG study were near-identical to those of Experiment 1 (see Section 2.1.2 for details). However, like in the previous complement coercion study, we also included an anomalous control condition, in order to assess the relationship between CQ interpretation and the interpretation of semantically ill-formed expressions. Participants judged the stimuli for sensicality off-line, at the end of each sentence.

Although our main aim was to investigate the contribution of the AMF to CQ interpretation, we also performed a complete analysis of the neural generators of all the major response components elicited by the critical verbs. Since Experiment 1 already found that CQs are associated with a different pattern of eye-tracking effects than complement coercion, it was reasonable to assume that the two constructions might also dissociate at the neural level. Since relatively few previous studies have investigated the neural bases of type-shifting phenomena, there is little prior literature that can be used to generate hypotheses about the possible localizations of the CQ effect other than the AMF. However, there is one deficit-lesion study suggesting a possible contribution of the left posterior temporal regions in the processing of aspectual coercion (Pinango & Zurif, 2001). In this study, Wernicke's aphasics with left posterior damage were found to be impaired on expressions requiring aspectual coercion, while Broca's aphasics (with no posterior damage) showed no evidence of impairment (see Section 5 and Pylkkänen & McElree, 2006, for a more detailed discussion). Although these deficit-lesion data do not discriminate between the hypothesis that aspectual coercion occurs in Wernicke's area and the hypothesis that Wernicke's area provides crucial input for aspectual coercion, this study does raise the possibility that left posterior regions play a role in sentence level semantic processing even outside classic "N400" plausibility manipulations. If question concealment modulated activity in similar regions, this would further corroborate such a hypothesis.

3.1. Methods

3.1.1. Participants

Nineteen monolingual native English speakers with normal or corrected-to-normal eyesight participated. They were paid $40 each for one 2-h session.

3.1.2. Materials

The materials of Experiment 2 were identical to those of Experiment 1, except for two minor changes. First, in order to achieve a consistent placement of triggers across our critical materials and the filler items (testing other hypotheses), the Control and CQ stimuli from the eye-tracking experiment were modified to include additional auxiliary verbs (such as *must be*). Second, one item from Experiment 1 was excluded from Experiments 2 and 3 because it contained the verb *asked*, which has recently been argued not to take a CQ complement (Nathan, 2006).[1] In addition, its exclusion allowed the exact balance of repetition of verb presentation; there were 9 verbs in the CQ condition and each verb was presented 3 times. An anomalous control condition was added to the design. The experimental task was an off-line sensicality judgment. The materials of Experiment 2 are listed in full in Appendix B.

(12) The proof of the classic theorem must be guessed by the brilliant mathematician (CQ)
(13) The proof of the classic theorem must be erased by the brilliant mathematician (Control)
(14) The proof of the classic theorem must be docked by the brilliant mathematician (Anom.)

As with the previous materials, CQ verbs were longer (8.2 characters) than either Control (7.4 characters) or Anomalous verbs (7.5 characters) [$F(2,26) = 6.15$; $p < .01$]. Frequency differed between conditions overall [$F(2,26) = 3.26$; $p = .05$], CQ verbs being more frequent in pair-wise comparisons than Anomalous (Scheffe test, $p < .05$), but not Control (Scheffe test, $p = .22$) verbs. Again, we expected that the lower frequency (Francis & Kučera, 1982) of Control (45.52 occurrences per million) and Anomalous (38.1 occurrences per million) verbs over CQ verbs (63.8 occurrences per million) would compensate for the shorter letter string length. Further, there was no difference in cumulative root frequency between any of the conditions [$F(2,26) = .02$; $p = .98$].

Additionally, we examined whether there were differences in plausibility between the CQ and Control sentences. Thirty monolingual native speakers of English participated in a separate study that rated the plausibility of the materials on a scale of 1 (implausible)–7 (very plausible). Materials were counterbalanced into 3 randomized lists containing equal numbers of items per condition. Analysis revealed a reliable effect of condition [$F(2,28) = 56.38$; $p < .0001$]. Crucially, pair-wise comparisons confirmed that Control (M = 5.3) and CQ (M = 5.2) sentences were rated as equally plausible (Scheffe, $p = .99$). Not surprisingly, anomalous (M = 2.2) sentences were rated less plausible than Control (Scheffe, $p < .0001$) and CQ (Scheffe, $p < .0001$) sentences. Thus, a difference in plausibility could not explain an effect between CQ and Control stimuli.

Finally, we also assessed the cloze probabilities of the critical verbs in their sentential context. Cloze probability is known to be one the main factors explaining variance in the amplitude of the N400 ERP (Kutas & Hillyard, 1984). Twenty participants wrote sentence completions for each sentence type used in the experiment. The participants were presented with the experimental sentences up to the point of the critical verb (e.g., *the name of the furry animal had been*...). Participants' completions were compared with the actual experimental sentences to assess how predictable the critical verbs were. Unsurprisingly, the mean cloze probability of our anomalous items was 0, but the cloze probabilities of the CQ and Control verbs were also extremely low, 0.004 and 0.002, respectively. Consequently, there was no overall effect of condition on cloze probability ($F < 1$). Thus our critical verbs were in general unpredictable, ruling out cloze probability based explanations of any measured MEG effects.

3.1.3. Procedure

Subjects lay in supine position in a dimly lit magnetically shielded room and viewed the stimuli through fiberoptic goggles (Avotec, FL), which were adjusted to their vision. The sentences were preceded by a fixation cross in the middle of the screen which was presented until the participant initiated the trial by pressing a button. Sentence stimuli were presented word by word (300 ms on, 300 ms off) in non-proportional Courier font (size = 70). A question mark appeared in the center of the screen after each sentence, at which point the participant had 4 s to indicate whether the sentence made sense or not. Subjects were instructed not to move their head or blink while reading and were encouraged to rest as often as necessary.

Neuromagnetic activity was recorded with a whole-head, 148-channel neuromagnetometer array (4-D Neuroimaging, Magnes WH 2500). The sampling rate was set to 679 Hz in a band between 0.1 and 200 Hz. The recording lasted approximately 45 min per participant. Afterwards, a short auditory baseline test was conducted, in which one hundred 1 kHz tones were presented through earpieces. Collecting these data permitted us to establish the source

[1] Taking this verb out from the eye-tracking data resulted in slightly higher levels of significance as the reading time patterns for this verb were the reverse of the standard pattern found.

location of the auditory M100 response, which was used as a functional landmark in further data analysis.

3.1.4. MEG data analysis

Prior to source modeling, MEG data were cleaned of artifacts and trials with incorrect responses. The data were averaged by condition using an epoch length of 900 ms, with 100 ms of prestimulus interval. The data were high-pass filtered at 1 Hz and low-pass filtered at 40 Hz prior to source analysis.

The generators of the magnetic fields were modeled as equivalent current dipoles. A multiple source model (BESA) was created for each participant on the basis of that individual's grand-averaged data across conditions. Sources were first fit at the peaks of all prominent response components between 0 and 500 ms after the onset of the critical verb and then combined together into a multidipole model. This multidipole model was then kept constant across conditions. All sensors were used in localization.

Source localizations and orientations were evaluated on the basis of two visual representations of magnetic activity: the *magnetic flux pattern*, representing the surface distribution of electromagnetic energy around the head, and the *minimum norm*, estimating current density. Only those sources consistent with the location and orientation of the magnetic field pattern and the minimum norm visualizations were accepted for analysis.

Goodness of fit (GOF) did not vary significantly between conditions [$F(2,16) = 1.3$; $p = .29$]; on average, the multi-dipole solutions explained approximately the same percentage of activity for the interval of 0–500 ms for the Control (84%), CQ (83%), and Anomalous (83%) conditions. Multi-dipole solutions explained approximately 87% of the activity from 0 to 500 ms after stimulus-onset in the grand average. On average, stimulus categories contained the same percent of trials (78%) per participant for all conditions.

3.2. Results and discussion

3.2.1. Sensicality judgment data

For inclusion in the statistical analysis, we required that participants perform with a minimum of 75% sensicality accuracy on the main two experimental conditions (CQ and Control). Two participants were excluded on these grounds from Experiment 2. Due to a button-box failure, response times from one participant were not recorded.

Sensicality judgments were recorded at the end of each trial at the question mark and are summarized in Table 2. Analysis of sensicality judgment response time revealed a main effect of condition [$F(2,15) = 4.06$; $p < .05$]. The speed of sensicality judgments did not differ significantly between Control and CQ conditions (Scheffe test, $p = .97$), but was faster in the Anomalous condition over controls (Scheffe test, $p = .05$). Crucially, Control and CQ conditions did not differ in sensicality judgment accuracy ($\sim$86%). This suggests that participants did not find the CQ condition less well-formed than the Control condition, confirming the previous plausibility ratings. As expected, anomalous conditions were judged as sensical in only a fraction of the trials (7%).

Table 2
Experiment 2: mean sensicality judgment times and percent materials judged sensical

Condition	Sensicality judgment time (ms)	% Sensical
Control	747	86.8
CQ	725	85.6
Anomalous	656	7.6

3.2.2. Multiple source models

As described above, a multiple source model was created for each participant's data based on the grandaverage of that individual's responses to all critical verbs. We first describe the characteristics of these multiple source models, before reporting on the effects of condition on source strengths and latencies.

Fig. 1 displays all individual source localizations (grey dipoles) and average localizations (black dipole). Auditory M50/M100 dipoles are included for reference. Fig. 2 illustrates for a single individual the typical field patterns associated with all the major response components as well the distributed source solutions of each component, as obtained by minimum norm estimates, which were used to guide the number of dipoles that were entered into the source modeling.

The MEG signals showed a pattern of response components largely familiar from previous MEG studies of language comprehension in the visual modality (Embick, Hackl, Schaeffer, Kelepir, & Marantz, 2001; Fiorentino & Poeppel, 2004; Helenius et al., 1998; Pylkkänen, Stringfellow, Flagg, & Marantz, 2001; Pylkkänen, Stringfellow, & Marantz, 2002; Pylkkänen, Feintuch, Hopkins, & Marantz, 2004; Pylkkänen, Llinas, & Murphy, 2006; Stockall, Stringfellow, & Marantz, 2004; Tarkiainen, Helenius, Hansen, Cornelissen, & Salmelin, 1999.). The only difference to previously reported activation patterns was the spatial distribution of left hemisphere activity at 300–400 ms. In this time window, previous MEG studies have consistently found a source cluster in the posterior temporal lobe, the so-called M350 (or N400m) (Embick et al., 2001; Helenius et al., 1998; Pylkkänen & Marantz, 2003; Pylkkänen et al., 2004, 2006), which has been hypothesized to index aspects of lexical access (Pylkkänen & Marantz, 2003). In addition to the posterior M350, several of our participants exhibited more anterior left lateral activity at 300–400 ms. This finding is consistent with a large body of neuroimaging studies on language comprehension, including syntactic (Caplan, Hildebrandt, & Makris, 1996; Friederici, Rüschemeyer, Hahne, & Fiebach, 2003; Just, Carpenter, Keller, Eddy, & Thulborn, 1996; Noppeney & Price, 2004; Stowe et al., 1998), sentence-level semantic (Baumgärtner, Weiller, & Büchel, 2002; Friederici et al., 2003; Kuperberg et al.,

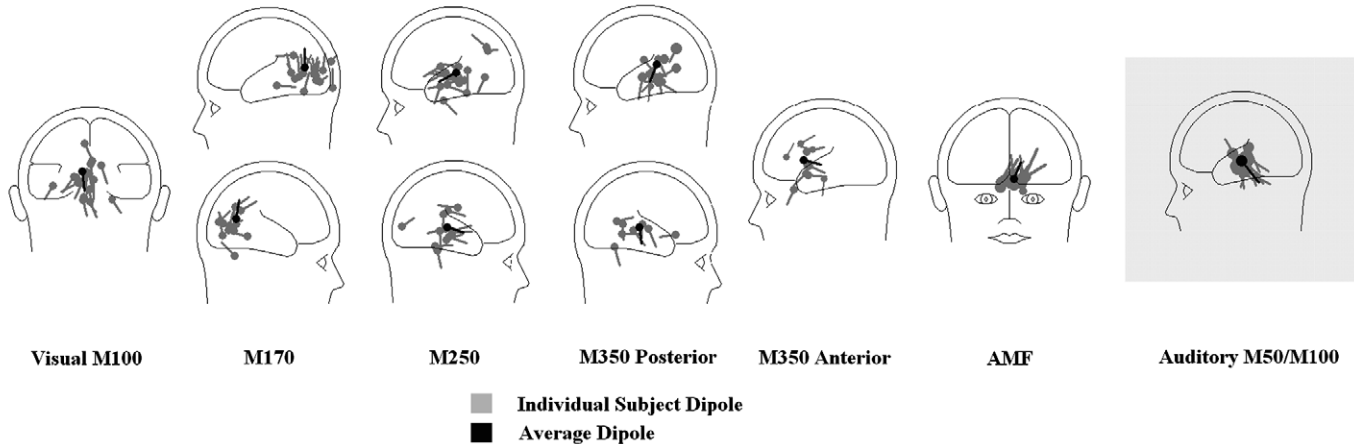

Fig. 1. Experiment 2: individual dipole (grey) and averaged dipole (black) locations of major MEG response components. Auditory M50/M100 dipoles were localized from a separate auditory baseline test. All sources are plotted inside an average-size head model.

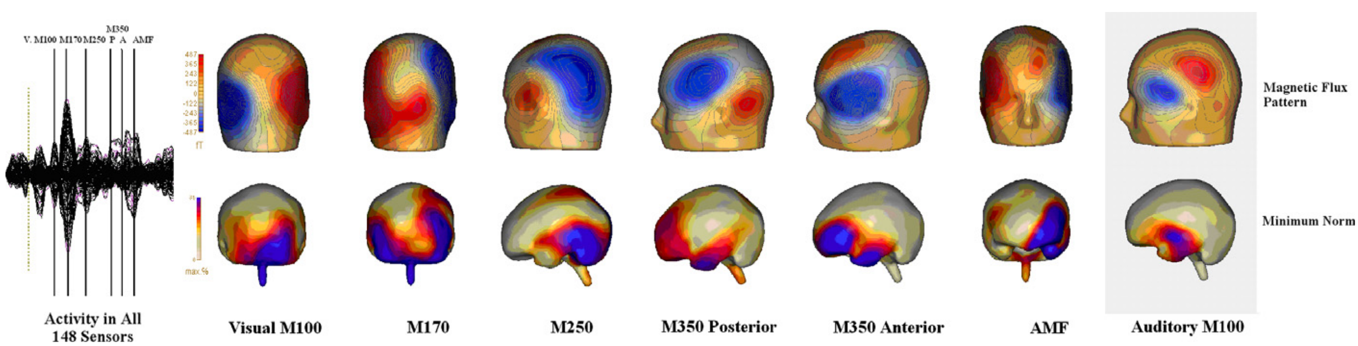

Fig. 2. Experiment 2: major MEG response components to visually presented in a sentential context from a single participant. Estimates of magnetic flux distributions are shown on top and minimum norm activations on the bottom. A typical auditory M100 response is shown separately.

2003), and lexical-level manipulations (Kotz, Cappa, von Cramon, & Friederici, 2002; Noppeney & Price, 2002), which have also reported activity in left frontotemporal regions.

3.2.2.1. Activity at 0–300 ms. MEG data from all participants ($N = 17$) exhibited a magnetic flux distribution typical to the visual M100 field pattern, characterized by an out-going magnetic field over the right-occipital sensors and an in-going field over the left-occipital sensors. This activity was best explained by a single midline dipole over occipital sensors in all participants.

After the visual M100 peak, activation spread to occipito-temporal areas bilaterally, peaking at ~150–200 ms (M170). In general, the M170 peak was associated with an out-going field over the left-occipital sensors, re-entering over the right-occipital sensors. Bilateral sources in the left and right occipital areas best explained the M170 activity for most participants ($N = 9$), and in only left but not right occipital areas for others ($N = 5$).

Activity then spread to posterior temporal areas, peaking between ~200–300 ms (M250). Bilateral two-dipole solutions best accounted for activity within the M250 time window for most participants ($N = 9$). Some participants displayed activity only in the left ($N = 2$) or right hemisphere ($N = 2$). Additional posterior sources were fit in the left ($N = 5$) and/or right ($N = 2$) hemisphere to create physiologically accurate multi-dipole solutions.

3.2.2.2. Midlatency left hemisphere activity. Previous research indicates that areas clustered around the left hemisphere auditory cortex are sensitive to semantic congruence (Helenius et al., 1998; Helenius et al., 2002). We used the localization of the auditory cortex as a functional landmark to classify midlatency left hemisphere activity as either 'posterior' or 'anterior'. Left hemisphere sources at 300–400 ms was labeled as 'anterior M350s' (M350-A), if they localized anterior to the auditory cortex and as 'posterior M350s', if they localized posterior to the auditory cortex. The auditory cortex was located on the basis of the neural response to 100 1 kHz tones in an auditory baseline test. Localization was performed either at the auditory M100 component ($N = 11$) or at the M50 component ($N = 1$), whichever field pattern was the clearest (for evidence that the two components are generated by similar cortical regions, see Mäkelä, Hämäläinen, Hari, & McEvoy, 1994). Auditory baseline data were not available for 5 participants; for those participants we used the mean auditory cortex localization obtained from the other 12 participants.

In total, 14 participants were modeled with left hemisphere sources in the M350 time window. The data from 6 participants were most accurately explained by both an M350-P and M350-A source in the left hemisphere, whereas 5 participants showed exclusively posterior and 3 participants exclusively anterior activity. As in previous studies, the M350-P was associated with an out-going field over left posterior areas, re-entering over sensors covering or anterior to the auditory cortex. For seven participants, accurate localization of the M350-P field pattern required an additional right hemisphere dipole (cf., Pylkkänen et al., 2006). In some participants, it was necessary to model other sources as co-active with the M350-P source: a parietal midline ($N = 2$), occipital ($N = 4$), or a frontal right lateral ($N = 2$) source. The M350-A, which closely followed the M350-P, was typically associated with an ingoing magnetic field in areas superior to the left anterior temporal lobe and an out-going field over inferior left-hemisphere sensors.

3.2.2.3. Midlatency right hemisphere activity. Eight participants showed right hemisphere activity at 300–400 ms. This activity was most often coactive with the M350-P left temporal source, consistent with the findings of Pylkkänen et al. (2006). Right hemisphere activity ranged over the entire right temporal and frontal lobes and was generally characterized by an outgoing field over sensors covering the lateral sulcus and an ingoing field in areas superior to this, although the precise orientation varied between participants.

3.2.2.4. Anterior midline field (AMF). Fourteen participants showed an Anterior Midline Field (AMF) at ~350–500 ms, which was characterized by an out-going field over frontal right hemisphere sensors, and a re-entering field over frontal left hemisphere sensors. As discussed above, this component has been implicated in the cost of complement coercion (Pylkkänen & McElree, 2007).

3.2.3. No effect of condition for sources from 0–300 ms

No reliable effect of question concealment or semantic anomaly was obtained either in peak latency or amplitude for sources at 0–300 ms. This is consistent with previous research where effects of semantic manipulations have only been found within later time windows (Helenius et al., 1998; Pylkkänen & McElree, 2007). The only early source that showed any effect of the stimulus manipulation was the left hemisphere M250 source, where pair-wise analyses revealed a marginal delay for concealed questions as compared to controls (Scheffe test, $p = .11$), although the main effect of condition was not significant [$F(2,12) = .82; p = .45$].

3.2.4. Instead of the AMF, left lateral effect of question concealment in the extended M350 time window

Contrary to the hypothesis that complement coercion and question concealment involve the same semantic repair mechanisms, the AMF showed no effect of condition in either latency or amplitude (both F's < 1). Instead, a main effect of condition was found in the peak latencies of both the M350-P [$F(2,10) = 5.99$; $p < .01$] and the M350-A [$F(2,8) = 4.67$; $p < .05$]. Pair-wise analyses revealed that the M350-P component peaked later for both CQ (Scheffe test, $p < .01$) and Anomalous (Scheffe test, $p < .05$) conditions than controls. The M350-A component, on the other hand, peaked reliably earlier for the CQ than for the Control condition (Scheffe test, $p < .05$). The Anomalous condition did not differ reliably from controls (Scheffe test, $p = .16$). Fig. 3 summarizes the M350 left hemisphere latency results. In addition, there was a trend towards diminished M350-A amplitudes [$F(2,8) = 3.23$; $p = .07$] in the Anomalous condition as compared to sensical conditions, which showed no difference in a pair-wise comparison (Scheffe, $p = .99$). The stimulus manipulation had no reliable effect on the amplitudes of the M350-P source, and no effect on either peak latencies or amplitude of right hemisphere sources. Table 3 shows the means for peak latency and amplitude for all sources in all conditions.

In sum, Experiment 2 aimed to identify the specific neural components underlying the processing cost elicited in Experiment 1. We found a reliable delay at ~350 ms in the left posterior temporal lobe. This effect did not extend to anterior temporal regions, which in fact showed the opposite effect for CQs. As shown in Table 3, in the control and anomalous conditions, the left anterior temporal sources peaked somewhat later than the left posterior temporal sources. For CQs, however, the anterior and posterior sources peaked approximately at the same time. It is possible that the simultaneity of the anterior and posterior activity for the CQ's reflects the involvement of both regions in CQ-type-shifting. However, at this point this hypothesis remains a speculation only, in particular since only 6 of our subjects exhibited both an anterior as well as a posterior left temporal source. Importantly, the AMF showed no sensitivity to question concealment. Thus, the overall pattern of our results suggests distinct processing mechanisms for concealed questions compared to complement coercion.

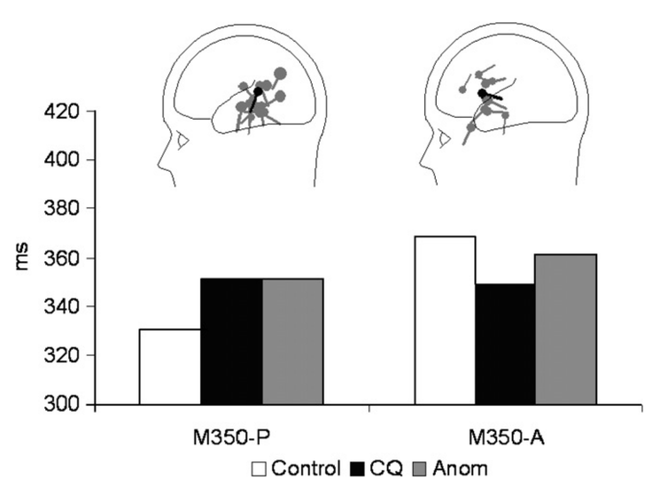

Fig. 3. Experiment 2: Mean peak latencies of M350-P and M350-A components. The grey dipoles represent individual and the black averaged dipole locations.

Table 3
Mean source latencies and amplitudes in all conditions for all components

Source dipole	Latency (ms)			Amplitude (nAm)		
	Control	CQ	Anomalous	Control	CQ	Anomalous
Visual M100	117	118	119	38.11	35.21	32.94
M170 LH	176	175	176	16.71	20.77	18.32
M170 RH	173	175	184	19.88	15.32	19.52
M250 LH	250	257	255	21.7	22.54	21.4
M250 RH	265	267	270	21.05	24.67	20.83
M350-P LH	331	351	351	10.49	12.45	13.05
M350-P RH	355	357	354	20.44	19.14	11.34
M350-A LH	369	349	361	12.57	12.85	9
M350-A RH	373	365	390	9.25	13.53	8.61
AMF	405	404	398	27.97	24.56	26.74

Statistically reliable effects are shaded grey.

Given that our manipulation varied the target verb, Experiment 2 cannot, however, conclusively determine whether the latency effect in the posterior M350 reflects an increased processing cost associated with type-mismatch or difficulty accessing relevant lexical properties of the question-selecting verbs. We addressed this question in Experiment 3.

4. Experiment 3

Since the CQ and Control verbs of Experiment 2 had different lexical semantics and selectional properties, the M350 delay associated with the CQ verbs could be due to longer access times as opposed to increased composition effort. In fact, the latencies and/or amplitudes of the posterior left hemisphere M350 source are modulated by a variety of lexical-level factors outside a sentential context (Beretta, Fiorentino, & Poeppel, 2005; Embick et al., 2001; Fiorentino & Poeppel, 2004; Pylkkänen et al., 2001, 2002, 2004, 2006; Pylkkänen & Marantz, 2003; Stockall et al., 2004), suggesting that this activity is involved in aspects of lexical access (Pylkkänen & Marantz, 2003).

If our CQ verbs invoke delayed processing independent of semantic composition, the M350 latency effect should be elicited even when the CQ verbs are viewed in isolation. In contrast, if the latency effect reflects increased difficulty in semantic composition, it should not be elicited in the absence of a sentential context. To test this, we presented the critical verbs without a sentential context in a lexical decision task.

4.1. Methods

4.1.1. Participants

The same 19 participants who participated in Experiment 2 also participated in Experiment 3.

4.1.2. Materials and procedure

The lexical decision study was initiated approximately 30 min after the sentence-level study had been completed. The two experiments were separated by the auditory baseline test as well as a significant rest. Consequently, repetition priming from Experiment 2 to Experiment 3 was extremely unlikely.

The verbs of Experiments 1 and 2 served as the word stimuli and the nonwords were generated using the English Lexicon Project (Balota et al., 2002). Nonwords were pairwise matched to the lexical level factors of each verb. The stimuli were presented after a centered fixation cross was displayed for 500 ms. Items were visible until the button press response. Subjects were given four rest breaks. The recording lasted approximately 20 min per participant.

4.1.3. MEG data analysis

The same procedure for filtering, cleaning, and averaging the MEG data was followed as in Experiment 2, except that data were averaged on a smaller epoch interval, ranging from 100 to 500 ms. As before, only trials with correct responses were included in analyses.

We focused our analyses on the source that showed a processing delay in the previous experiment, i.e., the M350-P. Since there is no standard method for comparing activity between a sentential context and isolation, we analyzed the data in the following two ways. First, we imported the multi-dipole models from the M350 time window in Experiment 2 into each averaged condition of Experiment 3. By importing the sources from Experiment 2, we kept source location and orientation identical between the experiments, allowing us to assess to what extent the M350-P, as localized in a sentential context, showed the same effect in isolation. However, while importing sources permitted direct comparison of how context affected the M350-P, the imported solutions were modeled from different data and thus were not necessarily accurate models of activity measured in Experiment 3. Therefore, we also created multi-dipole solutions for the M350 time-window by using the data measured in the lexical decision task. As is shown in Fig. 4, posterior M350 sources localized very similarly whether the lexical decision or the sentential data were used, suggesting that similar activity

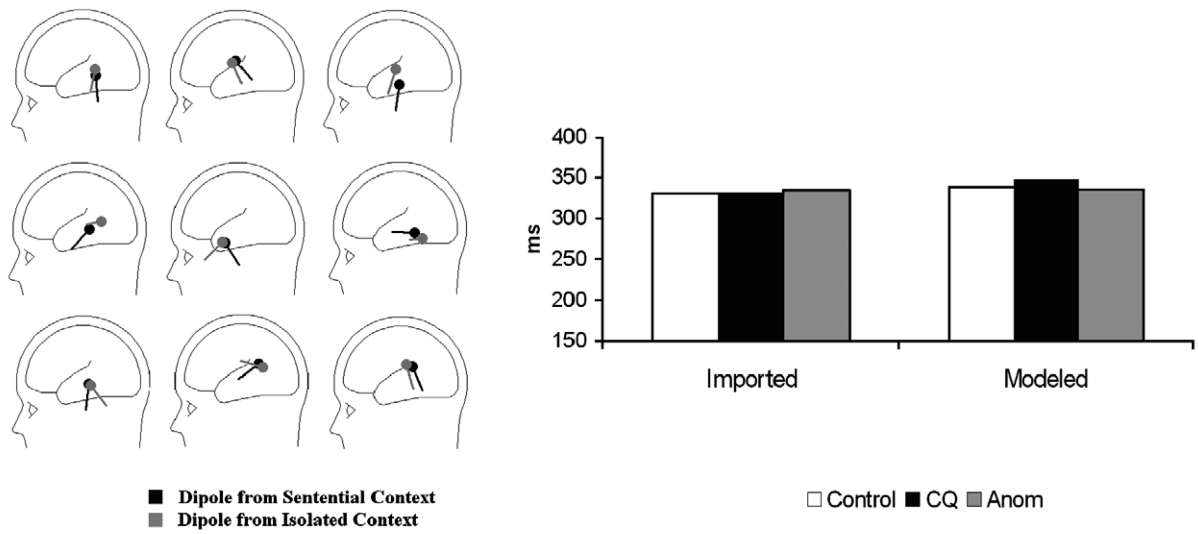

Fig. 4. Dipoles obtained from Experiment 2 (grey) versus Experiment 3 (black) are shown for each individual participant with M350-P activity in both experiments. Mean peak latencies for dipoles modeled from Experiment 2 (imported) and Experiment 3 (modeled) are shown on the right. No effect of verb type on peak latency for any condition within Experiment 3 was obtained. All sources are plotted within each individual's head model (contra Fig. 1).

was indeed elicited in the two different contexts. Having established that the presence of the M350-P was not context dependent, we proceeded to examine whether the effect it showed in Experiment 2 was.

4.2. Results

4.2.1. Lexical decision data

The behavioral lexical decision data are summarized in Table 4. There were no differences in lexical decision time between conditions overall [$F(2,17) = 1.53$; $p = .23$], and the accuracy difference were less than 2% between the Control conditions and the others (97% accuracy on both CQ and Anomalous conditions versus 99% accuracy on Control conditions). Both measures suggest that participants found the two verb classes equally easy to process. As expected, nonwords were significantly slower than all real word conditions ($M = 756$ ms) and were judged less accurately on average (86% correct).

4.2.2. MEG data

As already noted above, modeling M350-P sources directly from the lexical decision data revealed a close correspondence between patterns of activation and source locations in the two experiments. All but two M350-P sources fit in the sensicality experiment had corresponding activity in the lexical decision task. M350-P dipoles did not differ in location or orientation between types of models.

Table 4
Experiment 3: mean lexical decision times and percent correct

Condition	Decision response time	% Correct
Control	629	99.0
CQ	653	97.4
Anomalous	646	97.1

There was no difference in goodness of fit recorded at the M350-P peak [$F(1,9) = .56$; $p = .58$] or between the M350-P time window (300–400 ms), [$F(1,9) = 1.86$; $p = .20$] between imported dipoles and dipoles modeled directly from the lexical decision data.

Fig. 4 shows the mean latencies and individual locations of both the imported dipoles and the modeled dipoles. Our results clearly show that the latency effect obtained in Experiment 2 was absent when the sentential context was removed: There was no significant effect of condition on either M350 peak latency or amplitude in either analysis.

To further assess the dependence of the latency effect on the sentential context, M350-P latencies from Experiments 2 and 3 were entered into a 2×3 ANOVA testing for the interaction between context (sentential vs. isolation) and verb type (CQ, Anomalous, Control) (Fig. 5). When imported multi-dipole models were used for the M350 values of Experiment 3, the interaction was reliable [$F(2,10) = 4.01$; $p < .05$]. When the latencies of the M350s

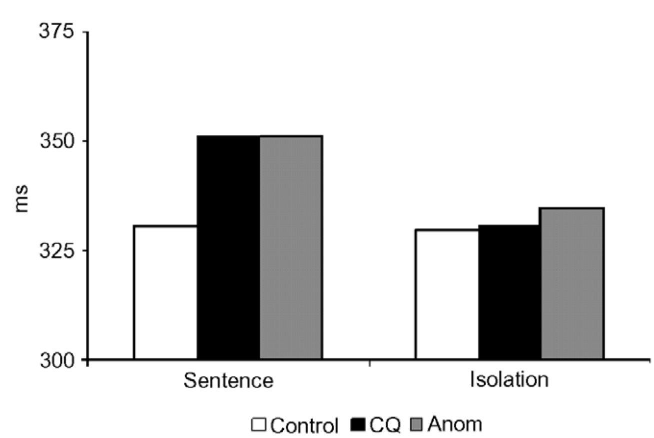

Fig. 5. Interaction between context and verb type (Experiments 2 and 3).

modeled on the basis of the lexical decision data were used instead, the interaction did not quite reach significance [$F(2, 14) = 2.93$; $p = .09$].

In conclusion, the results of Experiment 3 show that when the critical verbs of Experiments 1 and 2 are presented in isolation, they elicit no effect in the latency of the posterior M350 source. This suggests that the effect obtained in Experiment 2 was due to the type-mismatch between the CQ-verbs and their object, and not to the meaning difference between the CQ-verbs and the control verbs themselves.

5. General discussion

5.1. Concealed questions vs. complement coercion

This study was the first experimental investigation of concealed questions, a non-compositional construction involving a type-mismatch between a verb and its direct object. We found that the type-mismatch elicits a cost in both eye-tracking measurements and MEG signals, thus adding to the inventory of known processing effects incurred when comprehenders cope with non-compositionality.

Question concealment did not affect the same brain area as complement coercion, another construction involving a type-mismatch between a verb and its object. Concealed questions elicited a delay in a posterior left hemisphere source (M350-P), whereas complement coercion has been reported to elicit larger peak amplitudes in the AMF generated in the ventromedial prefrontal cortex (Pylkkänen & McElree, 2007). The distinct localizations indicate that different types of semantic repair call upon distinct resources for resolving type-mismatch.

Thus on the basis of these data we can reject the hypothesis that the resolution of all predicate-argument type-mismatches occurs via some common processing mechanism. Since this study mainly served to collect basic processing data on a so far uninvestigated construction, further research is obviously needed in order to fully understand the contrast between concealed questions and complement coercion. In what follows we outline several possible explanations and evaluate their plausibility.

First, it is possible that "type-shifting" is simply a descriptive term for a set of phenomena that do not in fact constitute any type of natural class at the neural level. Under this hypothesis, each construction would elicit a different pattern of activation depending on the details of the input and the output of the operation that serves to repair the type-mismatch. At present, our data are entirely consistent with this account.

Another possibility is that type-mismatch phenomena divide into cases of lexical ambiguity and cases of extra composition, the idea being that in the case of concealed questions, the basic and shifted senses of the noun are both listed in the noun's lexical entry, whereas for complement coercion, the shifted sense is derived via an operation. This hypothesis essentially equates CQs with polysemy, i.e., a situation where a single word has multiple related meanings. This account has some initial plausibility given that we indeed found CQs to modulate activity in an area associated with lexical access. Under this hypothesis, the M350 delay associated with CQs would result from competition between the two related senses of the CQ noun. However, this account is challenged by recent MEG evidence that competition effects between two related senses of a polysemous word localize to the *right* temporal lobe (Pylkkänen et al., 2006), whereas the CQ effect identified here was limited to the left hemisphere only.

A second variant of the polysemy hypothesis would be one where CQs are qualitatively similar not to cases of multiple listed senses but to cases of productive polysemy, where the derived sense may not necessarily be lexically listed. For example, many languages have a productive way to use the name of a producer for a product, as in *the gentleman read Dickens*. In most theories this would not be a case of semantic type-mismatch since the verb *read* selects for an entity and Dickens does denote an entity. However, in its person-sense Dickens denotes the wrong sort of entity to be a readable object, and thus the sense of Dickens is shifted to the 'product by Dickens' sense. McElree et al. (2006a) investigated this construction in eye-tracking, finding no evidence for a cost for expressions such as *the gentleman read Dickens* as compared to control stimuli such as *the gentleman spotted Dickens*. Since for concealed questions we did find an effect in eye-tracking, our evidence so far suggests that CQs do not pattern with productive polysemy, or at least not with the 'producer for product' case tested by McElree et al. (2006a).

Finally, it is possible that the relevant dimension on which complement coercion and concealed questions contrast is in the complexity of the operation required to fix the type-mismatch. Although these two cases of mismatch both involve NPs that are shifted to higher types, it can be argued that the input to CQ type-shifting is more complex than the input to complement coercion and that less semantic material is *added* in CQ-shifting than in complement coercion. The input to complement coercion is a simple entity, such as *the book*, that has no argument structure and can therefore only function as an argument, not as a predicate. Complement coercion converts *the book* into an event predicate, i.e., into a meaning roughly paraphrasable as *an activity involving the book*. In order to achieve this, the language processor must introduce an activity (or more technically, an event variable) into the semantic structure, construe the subject of the overt event-selecting verb (e.g., *begin*) as the agent of that activity, and create a thematic relation between the overt object, *the book*, and the covert activity. Concealed questions, on the other hand, are typically built on *relational* nouns, such as *price*, *distance*, *population* and so forth (Caponigro & Heller, in press; Nathan, 2006), as already discussed in the introduction to Experiment 1. Relational nouns do not denote entities but rather relations between entities. The noun *population*, for example, denotes a mapping from areas to the number of people living in an area, *price* a mapping from entities to the amount of money that must be paid in order to buy that entity, and so forth. Thus relational nouns are two-argument predicates. When they figure as concealed questions, their second argument is always implicit: *the population of California* mentions the area California (the first argument) but not the number of people living in that area (the second argument). In order to derive a question-like meaning from this NP, all that is intuitively needed is saturation of the implicit argument: To guess the population of California is to guess that the population of California is some value x. In other words, with concealed questions, all the semantic "ingredients" of the derived question-like meaning are already in present in the relational noun that serves as the input to type-shifting. In complement coercion, on the other hand, at least two thematic relations and one argument (the event argument) need to be added. Thus one hypothesis of the contrast in processing between complement coercion and concealed questions is that both involve type-shifting but complement coercion adds semantic content whereas CQ-shifting does not. Indeed, we found that CQs elicit faster effects than complement coercion both in eye-tracking and MEG. Under the complexity hypothesis, the longer effect latencies of complement coercion would be associated with the introduction of new variables and relations.

Interestingly, our previous MEG results suggest that the processes by which complement coercion is resolved localize in a frontal region not traditionally associated with language processing but rather with social cognition and theory of mind (Pylkkänen & McElree, 2007). The concealed question effect, on the other hand, affected activity in the left posterior temporal lobe, which is part of the "language network" in every model of the neurocognition of language. This raises the possibility that purely logical type-shifts engage traditional language areas whereas the addition of novel semantic content requires the recruitment of non-linguistic cognitive systems.

5.2. Left posterior temporal lobe as a locus of semantic processing

In this study we found evidence that converting a complex noun phrase such as *the winner of the contest* into a question-like meaning increases the peak latency of the M350 response, generated in the posterior parts of the left temporal lobe. This response is also sensitive to a variety of lexical-level factors, but the M350 effect of this study crucially depended on a sentential context, as it was not obtained in simple lexical decision. This suggests that the M350 may reflect not only the activation of lexical representations but also operations on them.

As already mentioned in our introduction, posterior parts of the left temporal lobe have been linked to type-mismatch in previous research as well. Pinango and Zurif (2001) found that Wernicke's aphasics with posterior left hemisphere damage showed significant difficulty when interpreting expressions requiring aspectual coercion, whereas Broca's aphasics, with no posterior damage, behaved like healthy controls. These results parallel our study, where the delay associated with type-mismatch occurred solely on posterior sites, anterior fronto-temporal regions showing no evidence for a cost. Piñango and Zurif also investigated complement coercion in the two aphasic groups, but the relationship between their complement coercion data and our MEG results on this construction is more complicated. Piñango and Zurif found that both Broca's and Wernicke's patients showed impaired comprehension of complement coercion (although the effect only reached significance for Wernicke's aphasics), suggesting that the comprehension of complement coercion may depend on both posterior and anterior left hemisphere regions. However, when we measured the neural activity associated with complement coercion directly in MEG, the effect did not localize to either of the areas implicated in Piñango and Zurif's study, but rather to ventromedial prefrontal regions. One way to reconcile these findings is the hypothesis that both the posterior temporal lobe and anterior fronto-temporal regions provide input for complement coercion. Consequently, damage to either area would result in problems with this construction, without either area being primarily responsible for computing the type-shift (Pylkkänen & McElree, 2006). Overall, then, the combined MEG and deficit-lesion data are currently compatible with the notion that concealed questions and aspectual coercion involve some similar processing mechanisms, located in posterior regions of the left temporal lobe, whereas complement coercion engages a more frontal midline region.

In addition to the effect of question concealment, the left posterior M350 was also modulated by semantic anomaly, traditionally associated with the N400 response in ERPs (Kutas & Hillyard, 1980). The primary generator of the N400 has been localized to the posterior parts of the left superior temporal gyrus in previous MEG studies as well (Halgren et al., 2002; Helenius et al., 1998). This finding is further corroborated by deficit-lesion studies (for review see Van Petten & Luka, 2006), which have found evidence of a dependency of the N400 effect on the left hemisphere in "split-brain" epileptics (Kutas, Hillyard, & Gazzaniga, 1988). Also, several studies have found that damage to the right hemisphere reduces but does not eliminate the N400 (Hagoort, Brown, & Swaab, 1996; Kotz & Friederici, 2003; Kotz, Friederici, & von Cramon, 1999; Swaab, Brown, & Hagoort, 1997). Finally, Friederici, von Cramon, and Kotz (1999) reported that damage to anterior or frontal areas does not eliminate the N400.

5.3. Timing of eye-tracking vs. MEG effects

In this study we investigated the processing of type-mismatch with two techniques, both of which yield more detailed temporal information than most traditional behavioral measures, such as self-paced reading. In our results, question concealment elicited an effect in eye-tracking about 50 ms earlier than in MEG. Although this seems surprising, it exemplifies a general trend: eyetracking effects often precede electrophysiological effects. As recently discussed by Bornkessel and Schlesewsky (2006), gender agreement conflicts, for example, have shown a correlation between first pass reading times and the much later P600 (Deutsch & Bentin, 2001). Similarly, lexical factors such as semantic priming and frequency most commonly affect the N400 in ERPs (Kutas & Federmeier, 2000), even though in eyetracking these effects can be observed already in first fixation times (Staub & Rayner, in press).

Although a detailed theory about the relationship between the two measurements is beyond the scope of our study, there are several possible explanations for the discrepancy between the timing of our eyetracking and MEG findings. First, the earlier eye-tracking effect of Experiment 1 may correspond to the left lateral M250 delay that did not reach significance. Second, it could reflect a more gradual ascending slope of the M350-P signal in CQ conditions and not the peak latency. Third, the method of presentation could have affected the time course of the processing costs. In the eye-tracking experiment, entire sentences were presented all at once. Comprehension could have been aided by parafoveal extraction of linguistic information, postulated to become available 70–140 ms after first viewing the pretarget word (Inhoff, Eiter, & Radach, 2005). As sentences were presented word by word in the MEG experiment, participants could not have made use of additional information downstream to speed up comprehension. Finally, Bornkessel and Schlesewsky (2006) sketch a general hypothesis about the timing discrepancy between eyetracking and electrophysiological effects, building on the idea that ERP components reflect the phase resetting of underlying oscillatory activity in different frequency bands (Basar, 1998, 1999; Makeig et al., 2002). Bornkessel and Schlesewsky hypothesize that eye-tracking effects may be a direct reflection of the phase resetting, with the corresponding ERP effect occurring later, especially when the critical information is encoded in slower frequency ranges, as is commonly the case for language-related ERPs. Clearly, more direct comparisons between eye-tracking and electrophysiological data are needed to clearly elucidate their relationship. What is important for the purposes of the present research, though, is that the timing discrepancy observed here is not specific to our study, but rather represents a generalization which deserves independent investigation in its own right.

6. Conclusion

In conclusion, eye-tracking and MEG measurements revealed that concealed questions are costly to interpret, in a way distinct from complement coercion. A separate MEG experiment presenting the critical words in isolation revealed no processing difficulty, showing that the cost indeed reflected the type-mismatch present in the sentential context and not the meaning difference between the critical verbs. The cost of question concealment localized in the posterior left temporal region, in an area known to be sensitive to both lexical and sentence-level semantic manipulations.

Acknowledgments

This research was supported by the National Science Foundation Grants BCS-0545186 (to L.P.) and BCS-0236732 (to B.M.) as well as the New York University Challenge Fund Award (to L.P.). We also thank Dr. Rodolfo Llinás for generously allowing us to perform the MEG recordings in his facility.

Appendix A. Experiment 1: Items

For each item, the verb before the "/" symbol is the CQ verb, the verb following this symbol is the control verb.

1. The name of the animal was learned/written by the students.
2. The title of the book was disclosed/listed by the author.
3. The flag of the nation was determined/carried by the king.
4. The age of the infant was guessed/confirmed by the doctor.
5. The alphabet of the language was predicted/printed by the linguist.
6. The contract of the company was decided/signed by the president.
7. The size of the herd was estimated/regulated by the inspector.
8. The variable in the equation was explained/omitted by the physicist.
9. The length of the operation was queried/recorded by the patient.
10. The anthem of the team was learned/shouted by the player.
11. The script of the show was disclosed/destroyed by the writer.
12. The costume of the actor was determined/delivered by the stylist.
13. The alias of the criminal was guessed/reported by the detective.
14. The paint of the portrait was predicted/stirred by the apprentice.
15. The number of the restaurant was asked/seen by the student.
16. The label of the shirt was decided/removed by the designer.
17. The volume of the music was estimated/noted by the engineer.

18. The design of the set was explained/sketched by the artist.
19. The distance of the flight was queried/stated by the pilot.
20. The speech of the senator was learned/published by the aide.
21. The recording of the band was disclosed/preserved by the singer.
22. The price of the phone was determined/reduced by the salesman.
23. The proof of the theorem was guessed/erased by the mathematician.
24. The peak of the stock was predicted/plotted by the broker.
25. The glaze of the cake was decided/eaten by the cook.
26. The value of the diamond was estimated/appraised by the jeweler.
27. The blueprint of the house was explained/modified by the architect.
28. The speed of the rocket was queried/computed by the observer.

Appendix B. Experiment 2: Items

For each item, the verb before the "/" separates the CQ verb, from the Control verb, from the Anomalous verb.

1. The name of the furry animal had been learned/written/climbed by the elementary students.
2. The title of the popular book had been disclosed/listed/poured by the popular author.
3. The flag of the young nation had been determined/carried/married by the despotic king.
4. The age of the sick infant had been guessed/confirmed/strangled by the family doctor.
5. The alphabet of the exotic language had been predicted/printed/dressed by the field linguist.
6. The contract of the bankrupt company had been decided/signed/parked by the sickly president.
7. The size of the wild herd had been estimated/regulated/delighted by the wildlife inspector.
8. The variable in the stochastic equation had been explained/omitted/accused by the nuclear physicist.
9. The length of the risky operation had been queried/recorded/bothered by the nervous patient.
10. The anthem of the local team should be learned/shouted/advised by the baseball player.
11. The script of the primetime show should be disclosed/destroyed/convinced by the television writer.
12. The costume of the famous actor should be determined/delivered/surprised by the fashion stylist.
13. The alias of the illusive criminal should be guessed/reported/interested by the private detective.
14. The paint of the infamous portrait should be predicted/stirred/wounded by the young apprentice.
15. The logo on the company shirt should be decided/removed/trained by the new designer.
16. The volume of the rock music should be estimated/noted/drunk by the sound engineer.
17. The design of the movie set should be explained/sketched/silenced by the movie artist.
18. The distance of the international flight should be queried/stated/ground by the tired pilot.
19. The speech of the senior senator must be learned/published/satisfied by the faithful aide.
20. The recording of the jazz band must be disclosed/preserved/persuaded by the star singer.
21. The price of the digital phone must be determined/reduced/pleased by the electronics salesman.
22. The proof of the classic theorem must be guessed/erased/docked by the brilliant mathematician.
23. The peak of the volatile stock must be predicted/plotted/spilled by the stock broker.
24. The glaze of the wedding cake must be decided/eaten/wired by the pastry cook.
25. The value of the genuine diamond must be estimated/appraised/smothered by the discount jeweler.
26. The blueprint of the modern house must be explained/modified/reminded by the successful architect.
27. The speed of the distant rocket must be queried/computed/astute by the astute observer.

References

Balota, D. A., Cortese, M. J., Hutchison, K. A., Neely, J. H., Nelson, D., Simpson, G. B., & Treiman, R. (2002). The English Lexicon project: A web-based repository of descriptive and behavioral measures for 40,481 English words and nonwords. http://elexicon.wustl.edu/, Washington University.

Basar, E. (1998). *Brain function and oscillations: Principles and approaches*. Berlin, Germany: Springer.

Basar, E. (1999). *Brain function and oscillations II: Integrative brain function. Neurophysiology and cognitive processes*. Berlin: Springer.

Baumgärtner, A., Weiller, C., & Büchel, C. (2002). Event-related fMRI reveals cortical sites involved in contextual sentence integration. *Neuroimage, 16*, 736–745.

Beretta, A., Fiorentino, R., & Poeppel, D. (2005). The effects of homonymy and polysemy on lexical access: An MEG study. *Cognitive Brain Research, 24*, 57–65.

Bornkessel, I., & Schlesewsky, M. (2006). The extended argument dependency model: A neurocognitive approach to sentence comprehension across languages. *Psychological Review, 113*(4), 787–821.

Caplan, D., Hildebrandt, N., & Makris, N. (1996). Location of lesions in stroke patients with deficits in syntactic processing in sentence comprehension. *Brain, 119*, 933–949.

Caponigro, I., & Heller, D. (2007). The non concealed nature of free relatives: Implications for connectivity crosslinguistically. In C. Barker & P. Jacobson (Eds.), *Direct Compositionality* (pp. 237–263). Oxford, UK: Oxford University Press.

Deutsch, A., & Bentin, S. (2001). Syntactic and semantic factors in processing gender agreement in Hebrew: Evidence from ERPs and eye movements. *Journal of Memory and Language, 45*, 200–224.

Dowty, D. (2007). Compositionality as an empirical problem. In C. Barker & P. Jacobson (Eds.), *Direct Compositionality*. Oxford.

Embick, D., Hackl, M., Schaeffer, J., Kelepir, M., & Marantz, A. (2001). A magnetoencephalographic component whose latency reflects lexical frequency. *Cognitive Brain Research, 10*, 345–348.

Fiorentino, R., & Poeppel, D. (2004). Decomposition of compound words: An MEG measure early access to constituents. In R. Alterman & D. Kirsch (Eds.), *Proceedings of the 25th annual conference of the cognitive science society* (pp. 13–42). Mahwah, NJ: Erlbaum.

Fodor, J. A., & Lepore, E. (2002). *The compositionality papers*. Oxford: Clarendon Press.

Frana, I. (2006). The de re analysis of concealed questions: A unified approach to definite and indefinite concealed questions. In *Proceedings of semantics and linguistic theory XVI* (pp. 17–34). Ithaca, NY: Cornell University.

Francis, W. N., & Kučera, N. (1982). *Frequency analysis of English usage*. Boston: Houghton-Mifflin.

Friederici, A. D., Rüschemeyer, S., Hahne, A., & Fiebach, C. J. (2003). The role of the left inferior frontal and superior temporal cortex in sentence comprehension: Localizing syntactic and semantic processes. *Cerebral Cortex, 13*, 170–177.

Friederici, A. D., von Cramon, D. Y., & Kotz, S. A. (1999). Language related brain potentials in patients with cortical and subcortical left hemisphere lesions. *Brain, 122*, 1033–1047.

Frisson, S., & McElree, B. (in press). Complement coercion is not ambiguity resolution: Evidence from eye movement. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Greenberg, B. (1977). A semantic account of relative clauses with embedded question Interpretations. Ms. UCLA.

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

Groenendijk, J., & Stokhof, M. (1989). Type-shifting rules and the semantics of interrogatives. In *Properties, types and meanings*. In Chierchia et al. (Eds.). *Semantics issues* (Vol. 2, pp. 21–68). Kluwer Academic Publishers.

Hagoort, P., Brown, C. M., & Swaab, T. Y. (1996). Lexical-semantic event related potential effects in patients with left hemisphere lesions and aphasia and patients with right hemisphere lesions without aphasia. *Brain, 119*, 627–649.

Halgren, E., Dhond, R. P., Christensen, N., VanPetten, C., Marinkovic, K., Lewine, J. D., et al. (2002). N400-like MEG responses modulated by semantic context, word frequency, and lexical class in sentences. *NeuroImage, 17*, 1101–1116.

Hämäläinen, M., Hari, R., Ilmoniemi, R., Knuutila, J., & Lounasmaa, O. (1993). Magnetoecncephalography–theory instrumentation, and applications to non-invasive studies of the working human brain. *Review of Modern Physics, 65*(2), 413–497.

Heim, I. (1979). Concealed questions. In Rainer Bäuerle, Urs Egli & Arnim von Stechow (Eds.), *Semantics from different points of view* (pp. 51–60).

Heim, I., & Kratzer, A. (1998). *Semantics in generative grammar*. Oxford: Blackwell Publishers.

Helenius, P., Salmelin, R., Service, E., & Connolly, J. F. (1998). Distinct time courses of word and context comprehension in the left temporal cortex. *Brain, 121*, 1133–1142.

Helenius, P., Salmelin, R., Service, E., Connolly, J. F., Leinonen, S., & Lyytinen, H. (2002). Cortical activation during spoken-word segmentation in nonreading-impaired and dyslexic adults. *Journal of Neuroscience, 22*, 2936–2944.

Hendriks, H. (1988). Type change in semantics: The scope of quantification and coordination. In E. Klein & J. van Benthem (Eds.), *Categories, polymorphism and unification* (pp. 96–119). Amsterdam: ITLI.

Inhoff, A. W., Eiter, B. M., & Radach, R. (2005). Time course of linguistic information extraction from consecutive words during eye fixations in reading. *Journal of Experimental Psychology: Human Perception and Performance, 31*, 979–995.

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

Just, M. A., Carpenter, P. A., Keller, T. A., Eddy, W. F., & Thulborn, K. R. (1996). Brain activation modulated by sentence comprehension. *Science, 274*, 114–116.

Kotz, S. A., Cappa, S. F., von Cramon, D. Y., & Friederici, A. D. (2002). Modulation of the lexical-semantic network by auditory semantic priming: An event-related functional MRI study. *NeuroImage, 17*, 1761–1771.

Kotz, S. A., & Friederici, A. D. (2003). Electrophysiology of normal and pathological language processing. *Journal of Neurolinguistics, 16*, 43–58.

Kotz, S. A., Friederici, A. D., & von Cramon, D. Y. (1999). Auditory word list priming in left and right temporal lobe lesion patients. *Brain and Language, 69*, 294–296.

Kuperberg, G. R., Holcomb, P. J., Sitnikova, T., Greve, D., Dale, A. M., & Caplan, D. (2003). Distinct patterns of neural modulation during the processing of conceptual and syntactic anomalies. *Journal of Cognitive Neuroscience, 15*, 272–293.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*, 463–469.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 207*, 203–205.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*(5947), 161–163.

Kutas, M., Hillyard, S. A., & Gazzaniga, M. (1988). Processing of semantic anomaly by right and left hemispheres of commissurotomy patients. *Brain, 111*, 553–576.

Makeig, S., Westerfield, M., Jung, T. P., Enghoff, S., Townsend, J., Courchesne, E., et al. (2002). Dynamic brain sources of visual evoked responses. *Science, 295*, 690–694.

Mäkelä, J. P., Hämäläinen, M., Hari, R., & McEvoy, L. (1994). Whole-head mapping of middle-latency auditory evoked magnetic fields. *Electroencephalography and Clinical Neurophysiology, 92*, 414–421.

McElree, B., Frisson, S., & Pickering, M. J. (2006a). Deferred interpretations: Why starting Dickens is taxing but reading Dickens isn't. *Cognitive Science, 30*, 113–124.

McElree, B., Pylkkänen, L., Pickering, M., & Traxler, M. (2006b). The time course of enriched composition. *Psychonomic Bulletin and Review, 13*, 53–59.

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.

Montague, R. (1970). Universal grammar. *Theoria, 36*, 373–398.

Nathan, L. (2006). The interpretation of concealed questions. PhD Thesis, MIT, MA.

Noppeney, U., & Price, C. J. (2002). A PET study of stimulus and task induced semantic processing. *Neuroimage, 15*, 927–935.

Noppeney, U., & Price, C. J. (2004). An FMRI study of syntactic adaptation. *Journal of Cognitive Neuroscience, 16*, 702–713.

Partee, B. (1986). Noun phrase interpretation and type-shifting principles. In J. Groenendijk et al. (Eds.), *Studies of discourse representation theory and the theory of generalized quantifiers* (pp. 115–143). Foris Publications.

Partee, B. H., & Rooth, M. (1983). Generalized conjunction and type ambiguity. In R. Bäuerle, C. Schwarze, & A. von Stechow (Eds.), *Meaning, use, and interpretation of language* (pp. 361–383). Berlin: Walter de Gruyter.

Partee, B. (1995). Lexical semantics and compositionality. In *An invitation to cognitive science*. In L. R. Gleitman & M. Liberman (Eds.). *Language* (Vol. 1, pp. 311–360). MIT Press.

Pickering, M. J., McElree, B., Frisson, S., Chin, L., & Traxler, M. (2006). Aspectual coercion and underspecification. *Discourse Processes, 42*, 131–155.

Pickering, M. J., McElree, B., & Traxler, M. (2005). The difficulty of coercion: A response to de Almeida. *Brain and Language, 93*, 1–9.

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

Pinango, M. M., & Zurif, E. (2001). Semantic operations in aphasic comprehension: Implications for the cortical organization of language. *Brain and Language, 79*, 297–308.

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

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

Pylkkänen, L., Feintuch, S., Hopkins, E., & Marantz, A. (2004). Neural correlates of the effects of morphological family frequency and family size: An MEG study. *Cognition, 91*, B35–B45.

Pylkkänen, L., Llinas, R., & Murphy, G. (2006). Representation of polysemy: MEG evidence. *Journal of Cognitive Neuroscience, 18*, 1–13.

Pylkkänen, L., & Marantz, A. (2003). Tracking the time course of word recognition with MEG. *Trends in Cognitive Sciences, 7*, 187–189.

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., pp. 537–577). NY: Elsevier.

Pylkkänen, L., & McElree, B. (2007). An MEG study of silent meaning. *Journal of Cognitive Neuroscience, 19*, 11.

Pylkkänen, L., Stringfellow, A., & Marantz, A. (2002). Neuromagnetic evidence for the timing of lexical activation: An MEG component sensitive to phonotactic probability but not to neighborhood density. *Brain and Language, 81*, 666–678.

Pylkkänen, L., Stringfellow, A., Flagg, E., & Marantz, A. (2001). A neural response sensitive to repetition and phonotactic probability: MEG investigations of lexical access. In *Proceedings of Biomag 2000, 12th international conference on biomagnetism* (pp. 363–367). Espoo, Finland: Helsinki University of Technology.

Romero, M. (2005). Concealed questions and specificational subjects. *Linguistics and Philosophy, 28*, 687–737.

Romero, M. (2006). On concealed questions. In *Proceedings of semantics and linguistic theory XVI* (pp. 208–227). Ithaca, NY: Cornell University.

Simos, P. G., Basile, L. F., & Papanicolaou, A. C. (1997). Source localization of the N400 response in a sentence-reading paradigm using evoked magnetic fields and magnetic resonance imaging. *Brain Research, 11*, 29–39.

Staub, A., & Rayner, K. (2007). Eye movements and on-line comprehension processes. In M. G. Gaskell (Ed.), *The Oxford Handbook of Psycholinguistics*. Oxford, UK: Oxford University Press.

Stockall, L., Stringfellow, A., & Marantz, A. (2004). The precise time course of lexical activation: MEG measurements of the effects of frequency, probability, and density in lexical decision. *Brain and Language, 90*, 88–94.

Stowe, L. A., Broere, C. A. J., Paans, A. M. J., Wijers, A. A., Mulder, G., Vaalburg, W., et al. (1998). Localizing cognitive components of a complex task: Sentence processing and working memory. *NeuroReport, 9*, 2995–2999.

Swaab, T., Brown, C., & Hagoort, P. (1997). Spoken sentence comprehension in aphasia: Event-related potential evidence for a lexical integration deficit. *Journal of Cognitive Neuroscience, 9*, 39–66.

Tarkiainen, A., Helenius, P., Hansen, P. C., Cornelissen, P. L., & Salmelin, R. (1999). Dynamics of letter string perception in the human occipitotemporal cortex. *Brain, 122*, 2119–2132.

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, 47*, 530–547.

Todorova, M., Straub, K., Badecker, W., & Frank, R. (2000). Aspectual coercion and the online computation of sentential aspect. In *Proceedings of the twenty-second annual conference of the cognitive science society* (pp. 3–8).

Van Petten, C., & Luka, B. J. (2006). Neural bases of semantic context effects in electromagnetic and hemodynamic studies. *Brain and Language, 97*, 279–293.

Winter, Y. (2007). Type shifting with semantic features: A unified perspective. In C. Barker & P. Jacobson (Eds.), *Direct compositionality*. Oxford University Press.