Cognition 161 (2017) 46–59

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Original Articles

Teasing apart coercion and surprisal: Evidence from eye-movements and ERPs

Francesca Delogu *, Matthew W. Crocker, Heiner Drenhaus

Saarland University, Saarbrücken, Germany

ARTICLE INFO

Article history:
Received 28 May 2015
Revised 16 December 2016
Accepted 31 December 2016
Available online 18 January 2017

Keywords:
Sentence processing
Coercion
Surprisal
Eye-movements
ERPs

ABSTRACT

Previous behavioral and electrophysiological studies have presented evidence suggesting that coercion expressions (e.g., *began the book*) are more difficult to process than control expressions like *read the book*. While this processing cost has been attributed to a specific coercion operation for recovering an event-sense of the complement (e.g., *began reading the book*), an alternative view based on the Surprisal Theory of language processing would attribute the cost to the relative unpredictability of the complement noun in the coercion compared to the control condition, with no need to postulate coercion-specific mechanisms. In two experiments, monitoring eye-tracking and event-related potentials (ERPs), respectively, we sought to determine whether there is any evidence for coercion-specific processing cost above-and-beyond the difficulty predicted by surprisal, by contrasting coercing and control expressions with a further control condition in which the predictability of the complement noun was similar to that in the coercion condition (e.g., *bought the book*). While the eye-tracking study showed significant effects of surprisal and a marginal effect of coercion on late reading measures, the ERP study clearly supported the surprisal account. Overall, our findings suggest that the coercion cost largely reflects the surprisal of the complement noun with coercion specific operations possibly influencing later processing stages.

1. Introduction

Psycholinguistic research has traditionally exploited online processing measures, such as reading times and event-related potentials, as a window into the representations and mechanisms that underlie language comprehension. Recently, however, information theoretic accounts (Hale, 2001; Levy, 2008) suggest that the cognitive effort that is indexed by these paradigms may be generally seen as indexing a word's *surprisal*, subsuming any costs associated with construction-specific structure-building mechanisms, *per se*. Thus, while Surprisal Theory has been shown to account for a broad range of linguistic (Piantadosi, Tily, & Gibson, 2011, 2012) and processing phenomena (Levy, 2008), it weakens the relationship between observable measures and underlying mental representations – what Levy (2008) has dubbed the *causal bottleneck*. Specifically, if Surprisal Theory is correct, then "representational choices affect predictions about incremental processing difficulty *exclusively* through the conditional word probabilities they determine" (Levy, 2008 pp. 1132–1133).

Since there exists a considerable literature suggesting construction-specific mechanisms are necessary to explain so-called *coercion* phenomena, we take this as a test case to assess whether a construction-specific representational account of processing difficulty is needed above and beyond a surprisal-based explanation. If any such processing effort is subsumed by surprisal, this would imply that construction-specific mechanisms, while providing a legitimate representational hypothesis, would be unnecessary to account for the processing data.

Verbs like *begin*, *finish* and *enjoy* require their complements to express an event or activity. While in (1) the activity is explicitly conveyed by the verb phrase (VP) complement, in (2) it must be inferred from the context. Expressions like (2) are typically interpreted as involving the recovery of a covert event-meaning that instantiates some activity commonly related to the complement noun, such as *reading* or *writing*.

(1) The author began writing the book.
(2) The author began the book.

Much work in lexical semantic research has been concerned with determining which type of operation is responsible for the interpretation of such expressions. According to one influential

* Corresponding author at: Department of Language Science & Technology, Building C7.1, Saarland University, 66123 Saarbrücken, Germany.
E-mail address: delogu@coli.uni-saarland.de (F. Delogu).

http://dx.doi.org/10.1016/j.cognition.2016.12.017
0010-0277/

analysis (Pustejovsky, 1995), these verbs semantically select for event-denoting complements. When the default interpretation of the complement is of a different type, as in (2), a type-coercion operation converts its semantic type into one that satisfies the verb's restrictions. Complement coercion (Pustejovsky, 1995) is thus an enriched form of semantic composition by which an event-sense of the complement (e.g., *writing the book*) is recovered in order to resolve a type-mismatch (see also Jackendoff, 2002).

Consistent with this hypothesis, several studies employing behavioral (e.g., Frisson & McElree, 2008; McElree, Traxler, Pickering, Seely, & Jackendoff, 2001; McElree, Frisson, & Pickering, 2006; Pickering, McElree, & Traxler, 2005; Traxler, Pickering, & McElree, 2002; Traxler, McElree, Williams, & Pickering, 2005) as well as neuroscientific methodologies (Baggio, Choma, van Lambalgen, & Hagoort, 2010; Husband, Kelly, & Zhu, 2011; Kuperberg, Choi, Cohn, Paczynski, & Jackendoff, 2010; Pylkkänen & McElree, 2007) have found that expressions like (2) incur increased processing cost. For example, in an eye-tracking study, Traxler et al. (2002) tested coercion expressions, as in (2), against control sentences such as *The author wrote the book* and found longer total reading time on the verb and complement noun and longer regression-path time from the words immediately following the noun for the coerced expressions. This increase in processing cost was interpreted as indexing the mechanisms that incorporate an event-sense of the complement into the representation of the VP.

A common approach in the aforementioned studies is the use of a control condition that overtly instantiates the preferred interpretation that comprehenders assign to the event-sense implicit in the coercion condition, as established through cloze tasks (e.g., *The author began to the book*; see, for example, Traxler et al., 2002; Pylkkänen & McElree, 2007). One concern with this type of manipulation, however, is that the predictability of the target noun in the control condition (henceforth, the preferred condition) is generally higher than it is in the coercion condition. Traxler et al. (2002), for example, report a significant difference in cloze probabilities for the target nouns, with the coercion condition producing lower cloze proportions (.03) than the preferred condition (.19) (see also Kuperberg et al., 2010; Pylkkänen & McElree, 2007).[1] The generally low cloze proportions in both conditions were taken to indicate that the target nouns were sufficiently improbable to rule out an effect of predictability on reading times. Very low cloze proportions are also reported in those studies that tried to match cloze probabilities across conditions (Baggio et al., 2010; McElree et al., 2001). As pointed out by Smith and Levy (2013), however, the cloze task makes it difficult to measure predictabilities less than 5–10%, which may be problematic for the aforementioned studies given that the effect of word predictability on reading time is logarithmic across six orders of magnitude in estimated word probabilities. To quantify this relationship, Smith and Levy (2013) estimated trigram word probabilities from the British National Corpus and examined their influence on both eye-movements and self-paced reading time. Their findings clearly demonstrate that comprehenders are sensitive to differences in predictability even between highly unpredictable words. The question therefore arises as to what extent the processing cost attributed to coercion may in fact be accounted for by the predictability of the complement noun alone, in accordance with Surprisal Theory (Hale, 2001; Levy, 2008).

This concern was explicitly addressed by Pylkkänen and McElree (2007) using magnetoencephalography (MEG). In order to distinguish effects of coercion from effects of predictability, they compared low predictable coercing nouns with an entirely unpredictable control condition involving semantic anomaly (e.g., *the journalist astonished the article*). They expected sources sensitive to predictability to show a graded effect, with the violation condition eliciting the strongest effect, the preferred condition (*the journalist wrote the article*) the smallest, and the coercion condition (*the journalist began the article*) patterning in the middle. The results showed precisely this pattern between 300 and 400 ms following the onset of the noun in a left temporal source (M350), the magnetic correlate of the N400 ERP component (Halgren et al., 2002; Helenius, Salmelin, Service, & Connolly, 1998; Pylkkänen & Marantz, 2003). Between 350 and 500 ms, however, a MEG response specific to coercion was found in a frontal source, the anterior midline field (AMF) localizing to the ventromedial prefrontal cortex, which has been more recently discussed as an index of semantic composition (Bemis & Pylkkänen, 2011; Pylkkänen, Brennan, & Bemis, 2011).

Although this study provides evidence that coercion incurs its own processing cost, it does not rule out the possibility that such cost also reflects comprehenders' expectations about upcoming nouns. Furthermore, the use of an anomalous condition to control for predictability is not entirely unproblematic: Anomalous and coerced nouns are qualitatively different types of items, with the former being unpredictable by definition (and therefore producing 0 cloze probabilities) while the latter elicit very low, but still detectable, cloze values.

The picture is further complicated by the results of two ERP studies (Kuperberg et al., 2010; Baggio et al., 2010) employing the same design as the one used by Pylkkänen and McElree (2007). Both studies predicted increased activity in the N400 time-window for coerced and anomalous nouns relative to preferred ones, but a distinct scalp distribution for the two effects, reflecting the distinct neural sources detected with MEG. Crucially, the results revealed a typical N400 effect for the coerced nouns which did not differ in magnitude or topography from the N400 effect elicited by the anomalous nouns. This finding was taken to support the view that the coercion cost at least partially reflects the detection of a semantic mismatch between the properties of the verb and those of the object (Kuperberg et al., 2010), an interpretation that was first considered by Traxler et al. (2005), but then rejected on the basis of the MEG data (see also Frisson & McElree, 2008).

To summarize, while there is robust evidence that complement coercion expressions incur increased processing cost, it remains unclear to what extent observed behavioral and neurophysiological measures are directly indexing the incorporation of an event-sense into the semantic representation of the VP, or simply the relative unpredictability of the target noun.

As noted above, an explanation in terms of predictability would suggest that processing costs associated with complement coercion constructions may be most appropriately accounted for by Surprisal Theory (Hale, 2001; Levy, 2008). Under such an account, the cognitive effort required to process a given word is proportional to its surprisal – or information conveyed (Shannon, 1948) – which is defined as the negative log probability of a word given the preceding context, as shown in 1:

$$Surprisal(word_i) = -\log_2 P(word_i | word_{0\ldots i-1}) \quad (1)$$

A growing body of empirical evidence indicates that surprisal is a significant predictor of reading times, ERPs and fMRI activation (e.g., Frank, Otten, Galli, & Vigliocco, 2015; Hale, 2001; Henderson, Choi, Lowder, & Ferreira, 2016; Levy, 2008; Smith & Levy, 2013 and studies cited therein; Willems, Frank, Nijhof, Hagoort, & van den Bosch, 2015). As such, Surprisal Theory has

[1] Cloze norms are not reported in other eye-tracking studies, such as Frisson and McElree (2008) and Traxler et al. (2005). However, in all these studies, the preferred verbs reflected the preferred interpretation of the implicit event-sense in the coercion condition. This procedure, as previously discussed, may result in a predictability difference between preferred and coercion verbs.

the potential to offer an overarching explanation for processing cost across a broad range of constructions, and is thus more parsimonious than accounts which posit a specific coercion mechanism.

In the studies reported here, we seek to determine whether there is any evidence for coercion-specific processing cost – above and beyond the difficulty predicted by surprisal – by contrasting coercion (3) and preferred (4) conditions, with a congruent control condition (5) in which the predictability of the complement noun is lower than in the preferred condition but, crucially, similar to that in the coercion condition.

(3) John began the book.
(4) John read the book.
(5) John bought the book.

This design contrasts with previous studies in two respects, which are crucial to evaluating the predictions of surprisal and coercion accounts. Firstly, we did not include a zero cloze condition to assess predictability effects, as this cannot distinguish improbable from either implausible or impossible target nouns. Second, our stimuli are controlled on the basis of both cloze and corpus-derived conditional likelihood estimates. Crucially, predictability was matched on the basis of corpus-derived surprisal estimates rather than cloze norming, as surprisal takes into account the non-linear, logarithmic relationship between word predictability and processing effort observed by Smith and Levy (2013).

Surprisal theory predicts that the equally unpredictive coercion (3) and control (5) conditions should be similarly difficult to process compared to the preferred condition, where the target is more predictable (4). In contrast, according to all coercion accounts discussed above, coercion (3) should elicit difficulty above-and-beyond surprisal. That is, the coercion condition should be more difficult than both the neutral (5) and the preferred condition (4), since only the coercion condition requires the incorporation of an event-sense into the VP, an operation which has been claimed to engender additional processing effort.

To thoroughly evaluate these two sets of predictions – and enable comparison with previous studies – we conducted both an eye-tracking and an ERP experiment utilizing the experimental design sketched above. In the eye-tracking study (Experiment 1) we expect the influence of surprisal or coercion to be manifest as increased reading time. A particular advantage of eye-tracking is the potential to reveal whether effects of surprisal and coercion are found in early versus later measures: While surprisal effects have often been found in relatively early measures (Levy, 2008; Smith & Levy, 2013), coercion effects have been associated with later processing, as reliable effects of coercion have been mainly found in late reading measures such as total reading time on the verb and object region and regression-path time on the post-critical region (see Pickering et al., 2005; Traxler et al., 2005; but see Frisson & McElree, 2008 for an earlier effect). The goal of the ERP study (Experiment 2) was firstly to replicate previous findings demonstrating the sensitivity of the N400 to predictability, and then crucially to determine whether it independently indexes coercion, when predictability is held constant.

2. Experiment 1

The present study extends previous research by examining the processing of coercing expressions when the predictability of the complement noun is controlled. In previous eye-tracking studies (e.g., Traxler et al., 2002), the coercion cost for expressions like *began the book* emerged from the comparison with a control sentence in which the verb (e.g., *read*) instantiated the event-sense implicit in the coercion condition. As previously discussed, this manipulation resulted in a difference in the predictability of the complement noun, as *book* is much more predictable following *read* than *began*. In this experiment, we introduced a further control condition – the neutral condition – in which the verb (e.g., *bought*) was equally unconstraining as *began*. If the coercion cost is driven by the predictability of the complement noun, we expect the coercion condition to pattern with the neutral condition, with both conditions eliciting equally longer reading times than the preferred condition. If, on the other hand, the coercion cost also reflects type-shifting operations to build an event-sense of the complement, coercing expressions should elicit longer reading times than the two control conditions. We also manipulated the frequency of the complement noun to establish whether or not the predictions of the two accounts are robust across a range of noun frequencies. An obvious prediction for models that attribute the frequency effect to differential expectations for low- and high-frequency words (e.g., Levy, 2008), is that the predictability effect should be larger for low-frequency nouns than for high-frequency nouns. Although a large number of studies have observed additive rather than multiplicative effects of frequency and predictability (e.g., Ashby, Rayner, & Clifton, 2005; Kennedy, Pynte, Murray, & Paul, 2013; Kliegl, Grabner, Rolfs, & Engbert, 2004; Miellet, Sparrow, & Sereno, 2007; Rayner, Aschby, Pollatsek, & Reichle, 2004), the debate is still open (see Hand, Miellet, O'Donnell, & Sereno, 2010).

2.1. Method

2.1.1. Participants

Forty-eight participants from Saarland University took part in the experiment. They were all native speakers of German and had normal or corrected-to-normal vision. All participants in the experiment gave written informed consent and were paid for taking part in the study.

2.1.2. Materials

We created 48 items in 6 conditions, crossing verb type (coercion, neutral, and preferred) with noun frequency (higher, lower). An example of the materials is presented in Table 1 (see also Appendix A):

We first selected a list of 10 coercion verbs such as *beginnen* (begin) and *genießen* (enjoy) and 48 pairs of entity-denoting nouns such that each pair consisted of a higher and a lower frequency noun belonging to the same semantic category (e.g., *Song* and *Refrain*).[2] The average frequency of the complement nouns was estimated using the deWaC German corpus (Baroni, Bernardini, Ferraresi, & Zanchetta, 2009). Higher frequency nouns occurred on average 48.2 times per million words, lower frequency nouns occurred 3.6 times. The difference was significant, $t(47) = 3.25$, $p < .01$. Higher and lower frequency nouns did not differ significantly with regard to length (6.6 and 6.5 characters on average, respectively). Nouns and verbs were then combined to construct a set of 48 items. Unlike previous studies (e.g., Traxler et al., 2002; Pylkkänen & McElree, 2007), sentence subjects were created using proper names instead of more informative subjects like *the author* or *the student*.

The verbs in the preferred and neutral conditions were selected on the basis of the results of a completion study assessing the default event-sense associated with each coercing expression (see Traxler et al., 2002). Sixteen participants who did not participate in the eye-tracking study were presented with fragments such as "John begann das Buch zu _____." ("*John began the book to* _____.")

[2] In order to create the 48 pairs, the verbs had to be repeated. We also repeated neutral verbs to minimize differences between these two conditions.

Table 1
Example of the materials used in Experiment 1.

Factors	
Verb type	Noun frequency [high/low]
Coercion	Peter begann [das Buch/das Exposé] im Urlaub. *Peter began [the book/the report] on vacation.*
Neutral	Peter kaufte [das Buch/das Exposé] im Urlaub. *Peter bought [the book/the report] on vacation.*
Preferred	Peter las [das Buch/das Exposé] im Urlaub. *Peter read [the book/the report] on vacation.*

and were asked to complete the sentence with an infinitive verb. The verbs used in the preferred condition occurred on average 62% of the time (65% for higher frequency nouns, 59% for lower frequency nouns).[3] Neutral verbs never occurred as a completion to the sentence fragments. This procedure resulted in coercion verbs being longer than neutral and preferred verbs (see Table 2). A one-way ANOVA with condition as within-item factor revealed a significant effect of condition, $F(2,94) = 10.45$, $p < .001$. Coercion verbs were significantly longer than neutral verbs, $t(47) = 2.13$, $p < .05$, and preferred verbs, $t(47) = 4.12$, $p < .0001$. Neutral verbs were significantly longer than preferred verbs, $t(47) = 2.66$, $p < .05$. We also examined the frequency per million words of the verbs in the deWaC corpus (see Table 2). The ANOVA on frequency data yielded a significant effect of condition, $F(2,94) = 3.993$, $p < .05$, with neutral verbs being more frequent than coercion verbs, $t(47) = 2.067$, $p < .05$, and preferred verbs, $t(47) = 1.992$, $p < .06$. Coercion and preferred verbs did not differ from each other ($t < 1$). These potential covariates will be addressed in the statistical analyses of the eye-movement data.

In order to ensure that the resulting set of stimuli showed the desired properties in terms of noun predictability, we first carried out a cloze study. Thirty-four participants were presented with the experimental sentences up to and including the determiner following the verb (e.g., *"Peter begann das _____."*). The results of the norming study are shown in Table 2. Concerning lower frequency target nouns, the mean cloze probabilities, as expected, were very low (coercion: .001; neutral: 002; preferred: .02). The cloze probabilities of higher frequency nouns were .011 (coercion), .013 (neutral), and .18 (preferred). A one-way ANOVA performed on higher frequency items produced a significant effect of condition, $F(2,94) = 22.102$, $p < .001$. The preferred condition produced higher cloze probabilities than the coercion condition, $t(47) = 4.71$, $p < .001$, and the neutral condition, $t(47) = 4.74$, $p < .001$. The coercion and neutral conditions did not differ from each other ($t < 1$).

We also estimated the surprisal of the complement noun from a trigram language model. To obtain bigram and trigram probabilities we first used the Google books n-gram corpus for German (available at http://storage.googleapis.com/books/ngrams/books/datasetsv2.html). The corpus, however, includes only n-grams that occur at least 40 times. As a consequence, some of our trigrams, in particular those containing lower frequency nouns, were not represented in the corpus. Probabilities were then estimated for the higher frequency items. Surprisal estimates were computed as the - log_2 P(noun| verb, determiner). The average surprisal for high frequency nouns in the coercion, neutral, and preferred conditions was 12.8, 12.9, and 7.4, respectively. The ANOVA revealed a significant effect of condition, $F(2,90) = 24.35$, $p < .0001$.[4] Preferred nouns were lower in surprisal than coercion nouns, $t(45) = 5.67$, $p < .0001$, and neutral nouns, $t(45) = 7.05$, $p < .0001$. Nouns in the coercion and neutral conditions did not differ from each other ($t < 1$).

To overcome the sparse data problem, and compute surprisal for the whole set of items, we estimated probabilities using the result counts of the Google search engine, following guidelines outlined in Rayson, Charles, and Auty (2012). We searched for bigrams and trigrams using the quotation marks operator to get the exact phrase (e.g., "kaufte das Buch") and we narrowed the results by language (German) and domain (.de). The average surprisal for the nouns in the coercion, neutral, and preferred condition is shown in Table 2. The ANOVA produced a reliable effect of condition, $F(2,90) = 37.07$, $p < .0001$. Nouns in the preferred condition were lower in surprisal than coerced nouns, $t(45) = 8.06$, $p < .0001$, and neutral nouns, $t(45) = 7.89$, $p < .0001$. Coercion and neutral nouns did not differ from each other ($t < 1$).

Stimuli were divided into 6 lists of 48 items, so that one version of each item appeared in each list and no participants saw more than one version of any given item. Experimental items were displayed along with 128 filler sentences. All items were followed by simple yes-no comprehension questions, designed to ensure participants were paying attention.

2.1.3. Procedure

Participants' eye-movements were monitored by an SR Research EyeLink-II head-mounted eye-tracker running at 500 Hz sampling rate and with a spatial resolution of 0.01°. Viewing was binocular, but only the participant's dominant eye was tracked. Stimuli presentation and recording of latencies were controlled by Experiment Builder (SR Research, Ltd. Kanata, Ontario, Canada). Participants were instructed to read at their normal speed and answer simple comprehension questions. The experiment began with the adjustment of the cameras. Next, a brief calibration procedure was performed during which participants had to look at a fixation dot in nine different positions on the screen. This procedure was repeated throughout the experiment anytime measurement accuracy appeared insufficient. A reading trial always started with the presentation of a fixation dot vertically centered on the left of the screen. Participants had to fixate the dot in order for the sentence to be revealed. The first character of the sentence replaced the fixation dot. Once participants had read the text, a comprehension question appeared to which participants responded by pressing one of two keys on the response box.

2.1.4. Data analysis

Fixations less than 80 ms were incorporated into larger fixations within one character. Fixations less than 40 ms were discarded, as readers do not extract much information during such short fixations (Rayner & Pollatsek, 1989). Fixations over 1200 ms were also discarded, as they usually indicate tracker loss.

We report data from four eye-movement measures: *first-pass time*, the sum of all the fixations made in a region until an eye-movement leaves the region either to the left or the right; *Number of regressions out*, the number of regressive saccades crossing the left edge of the scoring region; *Regression-path time*, the sum of fixations from first entering into a region from the left to the time the region is first exited to the right (this measure includes fixations made to re-inspect earlier portions of the text); and *total time*, the sum of all fixations made within a region.

Statistical analyses were conducted in three regions: the region containing the verb (*begann/las/kaufte*), the critical region containing the complement noun (*das Buch*), and the post-critical region containing the last words in the sentence (*im Urlaub*), where spillover or wrap-up effects could be observed.

Data for the eye-movement measures were analyzed using linear mixed-effects models with participants and items as crossed random effects. The models were evaluated using the lme4 package (version 1.1–5) within the statistical software R 3.0.2 (Bates & Sarkar, 2007). For all eye-movement measures reported here, a

[3] When the most frequent verb appeared insufficiently constraining (e.g., *see*), the second more frequent verb was selected.

[4] Two items that had no determiner were excluded from the analysis.

Table 2
Stimulus properties from Experiment 1.

Sentence type	Verb		Noun	
	Length	Frequency	Cloze	Trigram surprisal
Coercion				
John began the book	7.5 (2.3)	12.1 (23.7)	.006 (.02)	11.9 (4.2)
Neutral				
John bought the book	6.7 (1.6)	47.9 (114.8)	.008 (.03)	11.5 (4.0)
Preferred				
John read the book	5.6 (2.6)	15.1 (16.1)	.11 (.21)	6.7 (3.1)

Stimulus properties are presented as means, with standard deviations in parentheses.

linear mixed model was constructed, incorporating the fixed effects of Verb Type (three levels: coerced, neutral and preferred), Noun Frequency (two levels: higher and lower) and their interaction. The models also included two control predictors: verb length (number of characters), log transformed verb frequency and their interaction. All the continuous variables were centered prior to analyses, while the categorical fixed effects were deviation-coded.

As suggested by Barr, Levy, Scheepers, and Tily (2013), in evaluating the models we started with the maximal structure of random effects supported by the design, which included crossed random intercepts for both participants and items as well as random slope parameters for the main effects of Verb Type and Noun Frequency and their interaction. However, because the maximal models did not converge, we simplified the random effects structure by building the maximal model with no correlation parameters (see Barr et al., 2013). Model comparisons examining the main effects of Verb Type and Noun Frequency and their interaction were tested by means of likelihood ratio tests. Main effects were tested by comparing the base model (which only included the length and log frequency of the verb as fixed factors and the random factors) to the same model but with the factor Verb Type or Noun Frequency added. The interaction was tested by comparing the full model to a model containing the two main effects of Verb Type and Noun Frequency. The results of these tests are reported in Table 3. Pairwise comparisons were carried out on the subset of data corresponding to the relevant pair of conditions. We report coefficients, standard errors and t values for the significant effects of Verb Type and Noun Frequency. A given coefficient was judged to be significant at $\alpha = .05$ if the absolute value of t exceeded 2 (Baayen, 2008).

2.2. Results

2.2.1. Eye-tracking data

Participants scored above 90% accuracy on the comprehension questions in all the experiments reported here. Table 4 presents participants' means for each measure in each region.

2.2.1.1. First-pass time

There were no significant effects in first-pass time on the verb region. In the object region, there was an effect of Noun Frequency, with longer first-pass time for lower frequency nouns than higher frequency nouns (Estimate = 54.277, SE = 12.387, t = 4.382). No other significant effects were found in this region or in the post-critical region.

2.2.1.2. Number of regressions out

There were no significant effects in the number of regressions out in any of the regions of analysis.

2.2.1.3. Regression-path time

The analysis of regression-path data showed no significant effects in the verb region. In the object region, there was an effect of Noun Frequency, with longer regression-path time for lower frequency nouns than higher frequency nouns (Estimate = 64.595, SE = 17.009, t = 3.798). In the post-critical region, the analyses revealed an effect of Verb Type. Consistent with the surprisal account, pairwise comparisons showed longer regression-path time for the coercion condition compared to the preferred condition (Estimate = 132.049, SE = 48.476, t = 2.724) and for the neutral condition compared to the preferred condition (Estimate = 105.197, SE = 41.178, t = 2.555), but no difference between the coercion and the neutral condition ($t < 1$).

2.2.1.4. Total time

The total time data in the verb region revealed an effect of Verb Type, which was consistent with the surprisal account. Pairwise comparisons revealed longer total reading time for coercion verbs compared to preferred verbs (Estimate = 90.373, SE = 15.516, t = 5.825), and longer total time for neutral verbs compared to preferred verbs (Estimate = 67.088, SE = 18.522, t = 3.622), while neutral and coercion verbs did not differ from each other ($t < 1$).

In the object region, there was an effect of Noun Frequency, with longer total reading time for lower frequency nouns than higher frequency nouns (Estimate = 98.9697, SE = 19.2300, t = 5.147). The analyses also revealed an effect of Verb Type, which was consistent with the coercion account. Pairwise comparisons showed longer total reading time for coerced nouns compared to neutral nouns (Estimate = 48.7600, SE = 21.1537, t = 2.305), and to preferred nouns (Estimate = 86.8026, SE = 25.5609, t = 3.396). The difference between neutral and preferred nouns did not reach significance ($t < 1.8$). No significant effects were found in the post-critical region.

2.3. Discussion

Our main findings are summarized in Fig. 1, which shows the effects of Verb Type in the verb, object and post-critical region.

Interestingly, the effects of Verb Type were observed in the same measures as those observed in previous eye-tracking studies on coercion, i.e., in total reading time on the verb and the complement noun and regression-path time on the post-noun region (e.g., McElree et al., 2006; Traxler et al., 2002; Traxler et al., 2005). However, the current study allowed us to tease apart coercion from surprisal effects. In particular, the regression-path time on the post-noun region and the total time on the verb region were consistent with the surprisal account, which predicts that the coercion and neutral conditions should be similarly difficult to process compared to the preferred condition. Total reading time on the object region, however, showed a cost specific to coercion, as total time was longer for coerced nouns compared to neutral nouns. Since the early reading-time measures on the object region did not show

Table 3
Results of likelihood ratio tests to evaluate the main effects of Verb Type (VT), Noun Frequency (NF) and their interaction.

	Verb region		Noun region		Spillover	
	χ^2 (df)	p	χ^2 (df)	p	χ^2 (df)	p
First-pass time						
VT	2.13 (2)	.34	2.78 (2)	.25	1.52 (2)	.47
NF	.45 (1)	.50	16.57 (1)	<.0001	1.04 (1)	.31
VT× NF	1.84 (2)	.40	.56 (2)	.75	3.35 (2)	.19
Number of regressions out						
VT	1.81 (2)	.40	3.01 (2)	.22	5.61 (2)	.07
NF	1.79 (1)	.18	.04 (1)	.83	.42 (1)	.51
VT × NF	.35 (2)	.84	2.03 (2)	.36	1.57 (2)	.46
Regression-path time						
VT	1.31 (2)	.52	5.18 (2)	.08	11.44 (2)	<.001
NF	.08 (1)	.78	12.27 (1)	<.0001	.69 (1)	.41
VT × NF	.74 (2)	.69	2.57 (2)	.28	.77 (2)	.68
Total time						
VT	27.45 (2)	<.0001	9.68 (2)	<.001	3.21 (2)	.20
NF	1.14 (1)	0.28	20.72 (1)	<.0001	1.66 (1)	.20
VT× NF	1.73 (2)	0.42	.51 (2)	.77	3.22 (2)	.20

a Verb Type effect, it is likely that the cost in total time originated during re-inspections of the noun possibly coming from the post-noun region.

As a stronger test of whether the total-time effect held above-and-beyond any effect due to surprisal, we considered trigram surprisal estimates as a continuous predictor and compared a model including Verb Type, the control predictors and (centered) trigram surprisal with another model without the Verb Type factor but otherwise identical to the previous one. The inclusion of trigram surprisal rendered the Verb Type effect marginally significant ($\chi^2(2) = 5.2505$, $p = .072$), somewhat diminishing the extent to which the evidence supports an effect of coercion beyond surprisal.

3. Experiment 2

The results of Experiment 1 provide support for the surprisal account of the coercion cost, while also offering some evidence that coercion operations additionally influence later processing. The results, however, are unclear as to the precise time-course of surprisal and coercion effects. Surprisal effects emerged on the post-critical region, albeit in regression-path time, a reading measure reflecting initial difficulty and the time to over-come it (Clifton, Staub, & Rayner, 2007). Coercion effects emerged in a later measure, but on the noun region. In Experiment 2, we used ERPs to further assess the nature of the effect on the noun region. ERP studies use rapid serial visual presentation (RSVP) to help prevent eye-movement artifacts in the EEG. Unlike normal reading, with this procedure words cannot be skipped or re-inspected, but each word must be read for a fixed amount of time. Therefore, ERPs reflect the incremental processing of each word in a sentence. Furthermore, they capture what is reflected in early reading measures, not contaminated by regressive eye-movements (see, e.g., Dambacher & Kliegl, 2007).

In the present ERP experiment, we tested a subset of items from Experiment 1, focusing on the verb-type manipulation only, as no interaction with frequency was observed in Experiment 1. Based on previous findings, we expected effects of our manipulation to be manifest in the N400 component. The N400 has been shown to be sensitive to a word's predictability, with larger N400 amplitudes for lower than higher predictable words (e.g., DeLong, Urbach, & Kutas, 2005; Frank, Otten, Galli, & Vigliocco, 2013; Kutas & Hillyard, 1984). The N400 was also found to be modulated by coercion expressions, with larger N400 amplitudes for coerced than preferred nouns (Baggio et al., 2010; Kuperberg et al., 2010). Baggio et al. (2010) and Kuperberg et al. (2010) also found that anomalous nouns, which are unpredictable by definition, elicited an N400 effect similar in amplitude and topography to that elicited by coerced nouns. As argued in the introduction, however, these previous studies may not have adequately controlled for the predictability of the complement noun. Also, the inclusion of an anomalous condition to control for predictability may be problematic, as it involves comparing qualitatively different types of items (improbable vs. implausible nouns). Crucially, the items in our study were designed to control for noun predictability by matching the predictability of coerced nouns with that of non-coerced but plausible ones (the neutral condition), thus avoiding a comparison with zero cloze expressions.

The surprisal account predicts larger N400 amplitudes for coerced and neutral nouns compared to preferred ones and, crucially, no differences in the N400 amplitude or topography between the coerced and the neutral nouns. The coercion account, on the other hand, predicts larger N400 amplitudes for coerced than neutral and preferred nouns, reflecting the cost for type-shifting operations.

3.1. Method

3.1.1. Participants

Twenty-four right-handed native German speakers from Saarland University who did not take part in Experiment 1 participated in this experiment. All had a normal or corrected-to-normal vision and no history of neurological or psychiatric disorders. All participants gave their written informed consent and were paid for taking part in this study.

3.1.2. Materials

We selected a subset of items from Experiment 1 (see Appendix A). Since in Experiment 1 the noun frequency manipulation did not produce any interaction, in Experiment 2 we only considered the verb-type manipulation. The final set of stimuli, therefore, included 48 items in 3 conditions: coercion, neutral, preferred.

The verbs used in the preferred condition occurred on average 62% of the time in the completion study (see Experiment 1). The properties of the stimuli are shown in Table 5.

The average length of coercion, neutral and preferred verbs was 7.3, 6.7, and 4.5 characters respectively. A one-way ANOVA with condition (coercion vs. neutral vs. preferred) as within-item factor

Table 4
Experiment 1: Mean reading times (first-pass time, regression-path time, total time) and number of regressions out by region.

Measure	Verb region (*begann*) M(SD)	Noun region (*das Buch*) M(SD)	Spillover (*im Urlaub*) M(SD)
First-pass time			
Coercion-HF	274 (96)	332 (88)	514 (171)
Coercion-LF	267 (69)	394 (126)	493 (161)
Neutral-HF	261 (66)	328 (76)	475 (173)
Neutral-LF	271 (73)	380 (115)	495 (163)
Preferred-HF	242 (54)	330 (86)	486 (155)
Preferred-LF	250 (55)	380 (130)	521 (195)
Regressions out			
Coercion-HF	.15 (.17)	.27 (.28)	.97 (.53)
Coercion-LF	.13 (.16)	.24 (.22)	.98 (.49)
Neutral-HF	.14 (.15)	.24 (.25)	.98 (.51)
Neutral-LF	.13 (.15)	.21 (.25)	.93 (.48)
Preferred-HF	.13 (.14)	.23 (.18)	.84 (.42)
Preferred-LF	.10 (.13)	.26 (.23)	.91 (.46)
Regression-path time			
Coercion-HF	336 (123)	487 (185)	1111 (562)
Coercion-LF	324 (87)	541 (180)	1111 (598)
Neutral-HF	324 (107)	456 (171)	1109 (500)
Neutral-LF	324 (103)	501 (166)	1105 (516)
Preferred-HF	293 (80)	443 (125)	952 (437)
Preferred-LF	302 (99)	540 (197)	1017 (510)
Total time			
Coercion-HF	510 (202)	531 (194)	601 (247)
Coercion-LF	477 (162)	640 (263)	584 (257)
Neutral-HF	474 (188)	508 (179)	573 (226)
Neutral-LF	458 (186)	595 (212)	592 (249)
Preferred-HF	361 (138)	487 (189)	537 (238)
Preferred-LF	365 (162)	589 (220)	594 (284)

HF = High Frequency, LF = Low Frequency; Standard deviations in parentheses.

produced a significant effect of condition, $F(2,94) = 35.70$, $p < .0001$. Preferred verbs were shorter than coercion verbs, $t(47) = 8.1$, $p < .001$, and neutral verbs, $t(47) = 6.15$, $p < .0001$. The difference between coercion and neutral verbs was marginally significant, $t(47) = 1.78$, $p < .1$.

The ANOVA on frequency data revealed a marginal effect of condition, $F(2,94) = 2.731$, $p < .08$. Coercion verbs were more frequent than preferred verbs, $t(47) = 2.39$, $p < .05$. Neutral verbs were marginally more frequent than preferred verbs $t(47) = 1.93$, $p < .06$. Coercion and neutral verbs did not differ from each other ($t < 1.3$).

The cloze probabilities of the complement nouns patterned with those of the eye-tracking stimuli. For higher frequency nouns, cloze probabilities for both the coerced and neutral targets were .01, and for the preferred targets .17. The ANOVA with condition as within-item factor produced a significant effect of condition, $F(2,94) = 10.537$, $p < .01$. The preferred condition produced higher cloze proportions than the coercion condition, $t(47) = 2.96$, $p < .01$, and the neutral condition, $t(47) = 3.06$, $p < .01$. The coercion and the neutral conditions did not differ from each other, $t < 1$.

Finally, we recomputed the average surprisal for the complement nouns as estimated from Google.

The ANOVA revealed a significant effect of condition, $F(2,94) = 34.65$, $p < .0001$. Preferred nouns were lower in surprisal than coercion nouns, $t(47) = 7.49$, $p < .0001$, and neutral nouns, $t(47) = 7.99$, $p < .0001$. Coerced and neutral nouns did not differ from each other ($t < 1$).

Each participant saw all three versions of each stimulus in a pseudorandom order so that the effect of repetition at the sentence was equal in all conditions (see Pylkkänen & McElree, 2007). Moreover, to make repetition less obvious, the three versions of each stimulus appeared with a different proper name. Experimental items were combined with 168 unrelated filler sentences of various length.

3.1.3. Procedure

Participants sat in a comfortable chair in a dimly lit room. At the beginning of each trial a display screen appeared in which participants were prompted to press a button when they were ready to begin the trial. Next, a fixation point was displayed for 500 ms at the center of the screen, followed by the first word of the sentence. Following Baggio et al. (2010), each word appeared for 300 ms with an ISI of 300 ms before the onset of the next word. After 1000 ms of blank screen following the last word of each sentence, either a simple comprehension question requiring a yes-no answer appeared, or the next trial began. Comprehension questions appeared on 25% of trials to ensure participants were paying attention. At the beginning of the experiment, participants performed a practice session of six trials.

3.1.4. Electrophysiological recording

EOG and EEG were recorded from 26 active electrodes (actiCAP, Brain Products) embedded in an elastic cap according to the 10–20 system. Data were recorded using FCz as reference and AFz as ground. No filters were applied during recording. The vertical EOG was measured from two electrodes, one below and one above the left eye, and the horizontal EOG from one electrode placed at the outer canthus of each eye. Electrode impedance was kept below 5 kOhm for all scalp electrode sites, and below 10 kOhm for the EOG electrodes. The EEG signal was amplified by a BrainAmps DC amplifier (Brain Products) and sampled at 500 Hz.

3.1.5. Data analysis

The offline processing of the EEG data was performed using Brain Vision Analyzer 2 (Brain Products). The EEG signal was filtered (30 Hz low-pass filter) and then re-referenced off-line to the average of the left and right mastoid electrodes. Segments time-locked to the target nouns were extracted with an interval

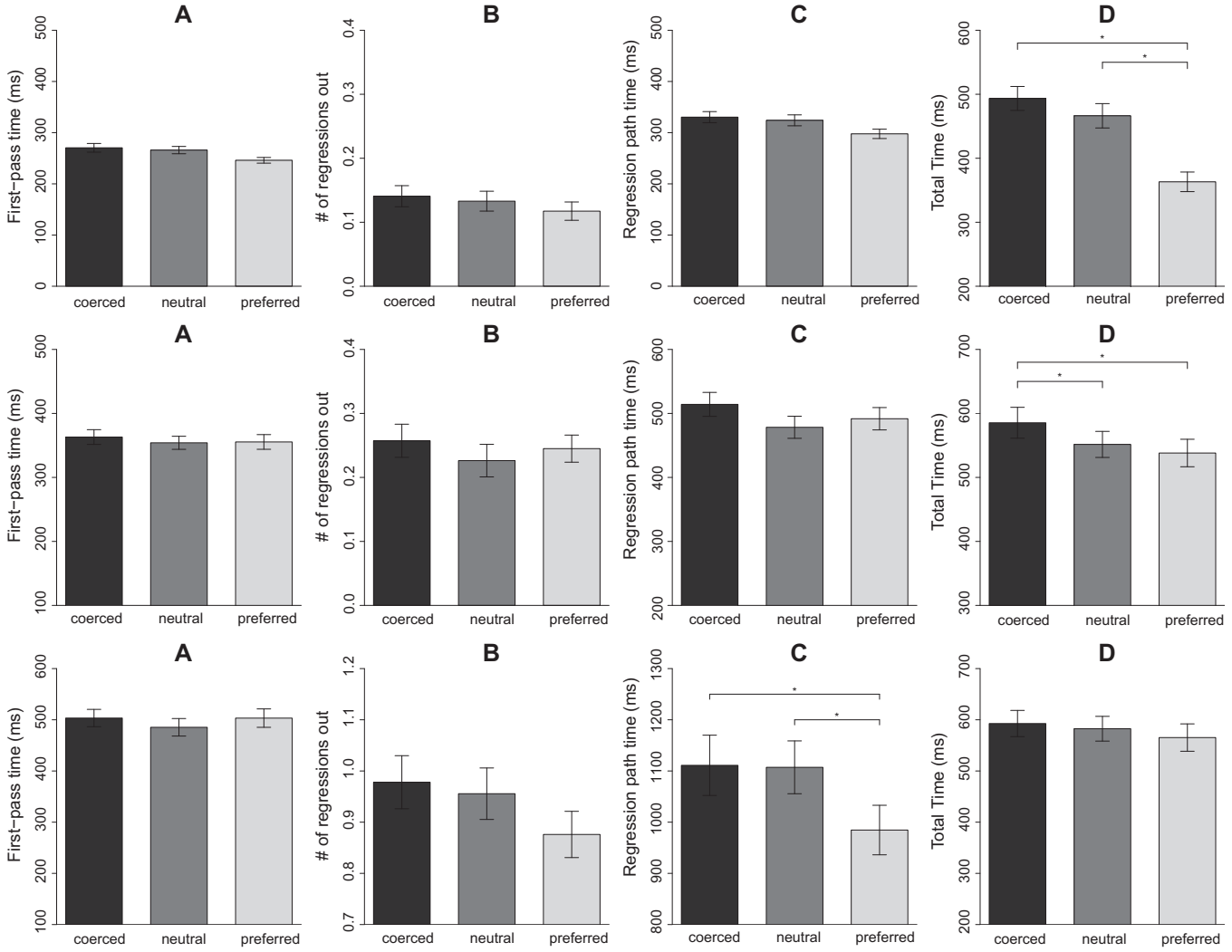

Fig. 1. Verb-type effects in the verb region (top panel), object region (middle panel) and post-critical region (bottom panel) – A: first-pass time; B: number of regressions out; C: regression-path time; D: total time. Confidence intervals indicate SEM.

Table 5
Stimulus properties from Experiment 2.

Sentence type	Verb		Noun	
	Length	Frequency	Cloze	Trigram surprisal
Coercion				
John began the book	7.3 (2.1)	22.0 (30.3)	.007 (.02)	12.2 (6.1)
Neutral				
John bought the book	6.7 (1.9)	50.6 (148.1)	.006 (.03)	12.7 (6.5)
Preferred				
John read the book	4.5 (1.8)	9.6 (9.5)	.10 (.20)	6.6 (4.9)

Stimulus properties are presented as means, with standard deviations in parentheses.

of 200 ms preceding and 1000 ms following the onset of the stimulus. Averaged ERPs were formed from segments free of ocular and muscular artifacts (detected using a semi-automatic procedure). Baseline correction used the 200 ms interval preceding the onset of the stimulus. Following Kuperberg et al. (2010), ANOVAs were carried out at a midline column, containing three electrode sites (Fz, Cz, Pz), and two lateral columns: the medial column, containing electrodes FC1, C3, CP1, FC2, C4, CP2 and the lateral column, containing electrodes F3, FC5 CP5, P3, F4, FC6, CP6, P4. Within-subject factors were Verb-Type (VT: coercion, neutral, preferred), Anterior-Posterior (AP) Distribution (with number of levels depending on the electrode sites along the AP axis) and Hemisphere (two levels) for the lateral analyses. In all ANOVAs, the Greenhouse-Geisser correction was used in cases with more than one degree of freedom in the numerator (Greenhouse & Geisser, 1959).

3.2. Results and discussion

Grand-average ERPs time-locked to the complement noun in all conditions at a subset of electrodes are presented in Fig. 2.

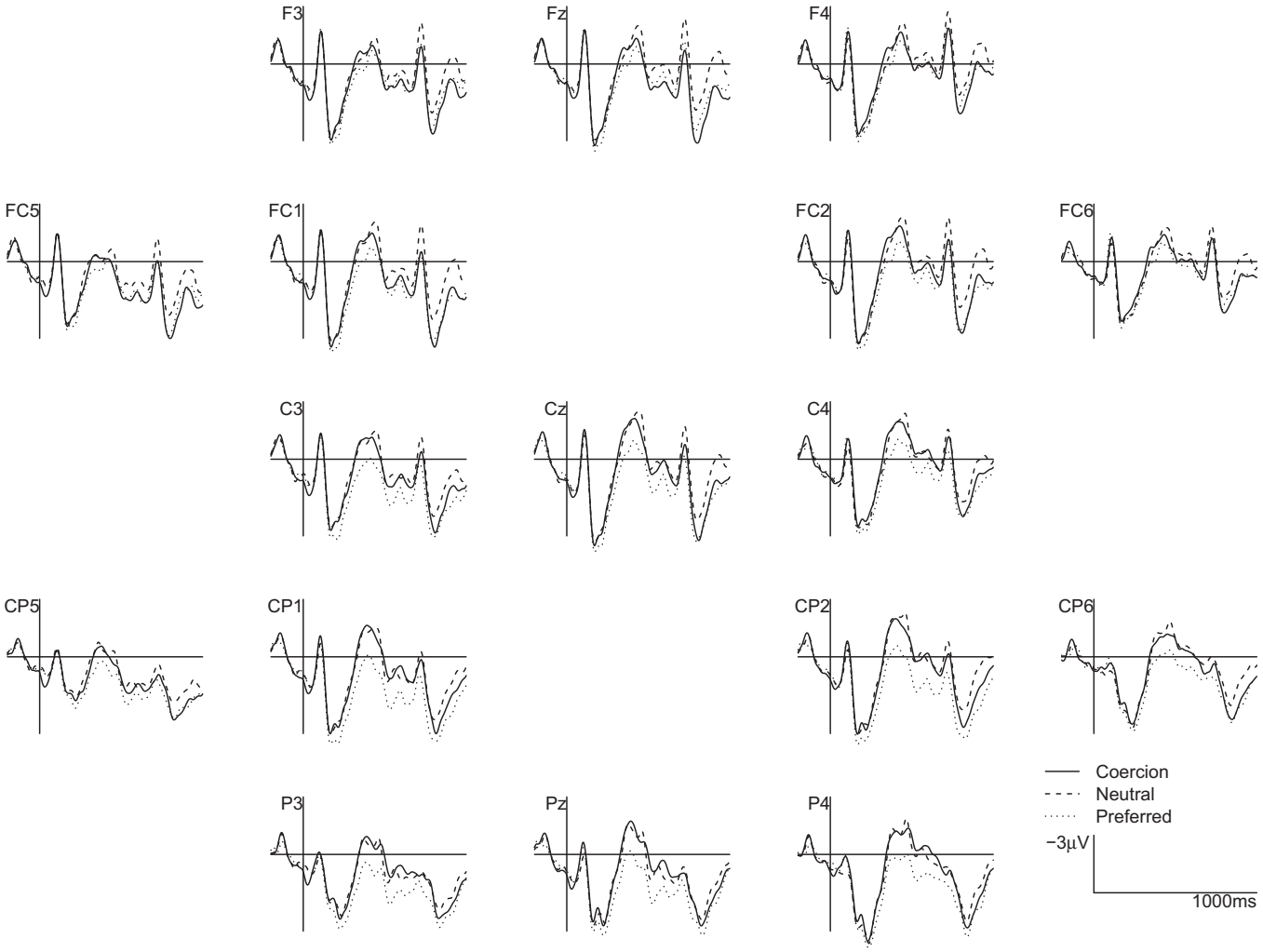

Fig. 2. Grand-average waveforms to complement nouns in all three conditions from frontal, central, and parietal electrode sites. Negative voltages are plotted upward.

Visual inspection of the ERP waveforms shows a difference between conditions in the N400 time-window, with a larger negativity for coerced and neutral nouns compared to preferred nouns. Both effects start around 200 ms, peak at 400 ms and appear to have a centro-parietal distribution, which is typical of the N400. The ANOVAs in the 300–500 ms time-window confirmed this conclusion: there was a significant main effect of condition and an interaction between condition and AP distribution at midline electrodes (condition: $F(2,46) = 3.56$, $p < .05$, $\eta_p^2 = .13$; interaction: $F(4,92) = 3.12$, $p < .05$, $\eta_p^2 = .13$), medial electrodes (condition: $F(2,46) = 4.75$, $p < .05$, $\eta_p^2 = .17$; interaction: $F(4,92) = 3.76$, $p < .01$, $\eta_p^2 = .14$), and, marginally, at lateral electrodes (condition: $F(2,46) = 3.03$, $p < .06$, $\eta_p^2 = .12$; interaction: $F(6,138) = 3.14$, $p < .01$, $\eta_p^2 = .12$). Pairwise comparisons (see Table 6) showed that the coerced nouns elicited a larger N400 than the nouns in the preferred condition. Similarly, the nouns in the neutral condition evoked a larger N400 compared to the nouns in the preferred condition. The amplitudes of the N400 to the coerced and neutral nouns did not differ significantly from each other.

We also carried out ANOVAs between 700 and 1000 ms, where Baggio et al. (2010) found an effect specific to coercion, with coerced nouns eliciting a larger negativity than preferred and anomalous nouns. The analyses revealed no evidence for such an effect. If anything, there was a marginally significant main effect of condition at medial electrodes, $F(2,46) = 2.77$, $p < .08$, $\eta_p^2 = .11$. Pairwise comparisons at these sites showed a significant difference only between the neutral and the preferred condition, $F(1,23) = 4.74$, $p < .05$, $,\eta_p^2 = .17$, (all other comparisons, $p > .1$).

Overall, these results support the surprisal account of the coercion cost, which predicts (i) higher processing cost for the nouns with higher surprisal (i.e., the coerced and the neutral nouns) compared to the preferred nouns and (ii) no differences between the coerced and the neutral nouns. Consistent with a vast ERP literature showing N400 predictability effects (e.g., Kutas & Federmeier, 2000; Kutas & Federmeier, 2011; see also Federmeier & Kutas, 1999, 1984), our results indicate that the N400 effect elicited by coerced nouns mainly reflects their contextual likelihood rather than the detection of a semantic mismatch between the verb and the object (Kuperberg et al., 2010) or some stage of coercion (Baggio et al., 2010). Both accounts would have predicted coerced nouns to elicit a larger N400 than neutral nouns, an effect that we did not observe in our study.

4. General discussion

While previous behavioral and neurophysiological studies on coercion consistently showed that coercing expressions incur increased processing cost, it remained unclear to what extent the

Table 6
ANOVAs on ERPs to complement nouns across the N400 time window (300–500 ms).

Effect		*F (df)*	*p*	η_p^2
Coercion versus Preferred				
Midline	VT	4.70 (1,23)	.04	.17
	VT × AP	5.20 (2,46)	.01	.18
Medial	VT	6.26 (1,23)	.02	.21
	VT × AP	6.21 (2,46)	.01	.21
	VT × H	.16 (1,23)	.69	.01
Lateral	VT	3.65 (1,23)	.07	.14
	VT × AP	4.38 (3,69)	.01	.16
	VT × H	.40 (1,23)	.52	.02
Neutral versus Preferred				
Midline	VT	5.14 (1,23)	.03	.18
	VT × AP	2.82 (2,46)	.07	.11
Medial	VT	7.01 (1,23)	.01	.23
	VT × AP	5.52 (2,46)	.01	.19
	VT × H	.30 (1,23)	.59	.01
Lateral	VT	4.56 (1,23)	.04	.17
	VT × AP	5.01 (3,69)	.01	.18
	VT × H	.15 (1,23)	.70	.01
Coercion versus Neutral				
Midline	VT	.01 (1,23)	.92	.00
	VT × AP	.82 (2,46)	.44	.03
Medial	VT	.01 (1,23)	.92	.00
	VT × AP	.03 (2,46)	.96	.00
	VT × H	.09 (1,23)	.77	.00
Lateral	VT	.04 (1,23)	.83	.00
	VT × AP	.33 (3,69)	.81	.00
	VT × H	.03 (1,23)	.87	.00

Note: VT = main effect of Verb Type; VT × AP = Verb Type × Anterior–Posterior distribution; VT × H = Verb Type × Hemisphere interaction.

observed effects directly index the incorporation of an event-sense of the complement into the semantic representation of the VP, or simply the relative unpredictability (surprisal) of the target noun.

Experiment 1 used eye-tracking to compare the processing of coercing expressions like *began the book* to that of two non-coercing conditions differing in the predictability of the complement noun: The noun in the preferred condition was significantly lower in surprisal than the noun in the coercion and neutral conditions. The pattern of regression-path time on the post-critical region and of total time on the verb region revealed similarly increased processing costs for the two conditions with higher surprisal, consistent with a surprisal account of coercion. The total time on the noun region, however, showed an effect specific to coercion, as reflected by longer reading time for the coercion condition compared to the two non-coercing expressions. When trigram surprisal estimates were considered as a predictor of reading times, however, the coercion effect was only marginally significant. In Experiment 2, we examined ERPs for similar materials and found evidence exclusively for the surprisal account: The N400 effect elicited by coercing nouns was similar in amplitude and topography to that elicited by neutral nouns. We start by discussing the ERP results, as they more obviously diverge from previous ERP findings on complement coercion that cannot be straightforwardly explained in terms of predictability.

It is well established in the ERP literature on language comprehension that the N400 component is sensitive to the predictability of a word in its context, as estimated not only by cloze probability (e.g., DeLong et al., 2005; Kutas & Hillyard, 1984), but also by surprisal (Frank et al., 2013). The N400 modulation to the noun in our study is fully explainable in terms of predictability, with no evidence that it might index some stage of coercion. The current account, however, seems unable to explain the results of Kuperberg et al. (2010) and Baggio et al. (2010), where coerced nouns were compared to preferred and anomalous nouns violating the selectional restrictions of the verb (as in *The author astonished the book*). Even though both of those studies attempted to control the predictability of the complement noun across conditions, they found that coerced and anomalous nouns elicited larger N400s than preferred nouns, and that the N400 effect elicited by coerced nouns was similar in amplitude and topography to that elicited by anomalous nouns. Thus, these previous results appear to conflict with the current study in two ways: (a) they observed a smaller N400 amplitude elicited by the preferred condition even though predictability was claimed to be controlled across conditions and (b) the anomalous condition did not elicit a larger N400 amplitude than the coercion condition, which would have been expected given the cloze probability difference between the anomalous and the coercion conditions (0 cloze for anomalous nouns vs. low – but still greater than 0 – cloze for the coercing nouns).

Concerning (a), Baggio et al. (2010) used items associated with very low cloze values (.061 for coerced nouns and .086 for preferred nouns). As pointed out by Smith and Levy (2013), small absolute differences in expectation for low-predictability words may produce relatively large effects on processing difficulty. The cloze task makes it difficult to estimate true differences in predictability for low-predictability words, leaving open the possibility that the ERP effect observed by Baggio and colleagues actually reflects very small differences in predictability.

Concerning (b), we consider several reasons why the anomalous condition did not elicit the largest N400. Firstly, in Kuperberg et al.'s experiment, the violation condition elicited a P600 effect relative to both the preferred and the coercion conditions. The absence of an N400 effect in the comparison of anomalous and coercing nouns could potentially be attributed to an overlap between components: The positivity elicited by the anomalous condition might have 'pulled down' the N400 amplitude in this

condition, thereby masking the N400 effect. This interpretation, however, cannot straightforwardly account for Baggio et al.'s results, in which, if anything, a larger positivity to the anomalous condition was evident in a very late time-window starting around 750–800 ms after the onset of the noun and only on parietal sites.[5]

There is, however, another possible explanation for the absence of an N400 effect, which relates to the nature of the N400 component. Although the functional interpretation of the N400 is a matter of ongoing debate (for an overview see Lau, Phillips, & Poeppel, 2008; Kutas & Federmeier, 2011), a growing body of evidence supports the view that the N400 amplitude indexes processes associated with lexical activation and retrieval (e.g., Federmeier & Kutas, 1999; Lau, Almeida, Hines, & Poeppel, 2009; Kutas & Federmeier, 2000; see also Brouwer, Fitz, & Hoeks, 2012 for an overview). On this view, reduced N400 amplitudes reflect facilitated retrieval of lexical information when context can pre-activate features of the critical word.

Not only can this perspective account for the sensitivity of the N400 to a word's offline predictability, but also for those cases in which an N400 effect that *would* be expected on the basis of cloze probabilities is not evoked. Several studies have failed to observe an N400 effect in response to anomalies in role reversed sentences – such as *The javelin has the athletes thrown* (Hoeks, Stowe, & Doedens, 2004), or *For breakfast the eggs would eat* (e.g., Kuperberg, Kreher, Sitnikova, Caplan, & Holcomb, 2007) – relative to their plausible (and more predictable) counterparts (see also Kim & Osterhout, 2005; Kolk, Chwilla, van Herten, & Oor, 2003; Kuperberg, Sitnikova, Caplan, & Holcomb, 2003; van Herten, Chwilla, & Kolk, 2006). As pointed out by Brouwer et al. (2012) and Chow and Phillips (2013), the absence of N400 effects in such cases may be due to a facilitatory (or priming) effect arising from the lexical association between the critical word and the preceding words in the sentence (e.g., *javelin-athlete-throw* or *breakfast-eggs-eat*). Similarly, the lexical association between the arguments of the verb in the example reported in Baggio et al. (2010) and Kuperberg et al. (2010) (i.e., *author-book*) might partially explain why the coercion and anomalous conditions elicited similar N400 amplitudes. While we are aware that this interpretation is only speculative, we can at least rule it out as an alternative explanation of our own ERP findings, since we used semantically empty proper names as the subjects of our sentences.

Our results therefore indicate that the N400 effect elicited by coercing nouns most likely reflects their contextual likelihood rather than some stage of coercion. Notably, this finding is consistent with previous MEG data on complement coercion (Pylkkänen & McElree, 2007) in which the magnetic counterpart of the N400 component – the so-called M350 – was only sensitive to the noun's predictability, whereas an MEG response specific to coercion emerged in a frontal source (the AMF) and in a later time-window. Further evidence that the coercion operation may occur at late processing stages is given by the results of Traxler et al. (2002), who tested a crucial prediction of the coercion account, namely that coercing verbs combined with NP complements denoting an event should not incur additional processing cost, since no type-shifting operation is required. In Experiment 2, they contrasted sentences like *The boy began the fight/the puzzle* with control sentences such as *The boy saw the fight/ the puzzle* and found greatest difficulty when entity NPs followed eventive verbs, but only in late reading measures (i.e., second-pass and total reading time) on the NP region. It is important to note that the control verbs (e.g., *saw*) were similarly unconstraining to the neutral verbs used in our study. Therefore, when predictability is controlled, coercion effects emerge on measures of later processing, a result that would corroborate our marginal effect of coercion in the total time on the complement noun.

Somewhat surprisingly, however, noun predictability did not affect reading times on the noun itself, but only those on the region following it. This result seems to be not fully consistent with previous eye-movement studies in which predictability effects have been detected not only on the spillover region, but also on the target region and in early reading measures (Ehrlich & Rayner, 1981; Frisson, Rayner, & Pickering, 2005; Rayner, Binder, Ashby, & Pollatsek, 2001; Rayner & Well, 1996; Smith & Levy, 2013; but see Calvo & Meseguer, 2002). One possible explanation for this discrepancy is that, unlike those studies, the context in our stimuli consisted of only three words, with the predictive component being only the verb. Although expectations can be generated in such a limited context (as demonstrated by cloze measures), their effects might arise later in time, shortly after the critical words. Nonetheless, clear effects of noun predictability were found for the object noun in the ERP responses.

The present findings provide further support for the generality and importance of Surprisal as a linking theory relating language comprehension processes with on-line measures of processing effort. While we do find some evidence for coercion specific effects in later reading time measures, these were only marginally significant when item-specific trigram surprisal was included as a predictor. This suggests that any costs associated specifically with comprehending coercion expressions are largely subsumed by the more general costs associated with their higher surprisal. The absence of a processing index of coercion should not, however, be taken as evidence against representational accounts, which are presumably necessary to explain how these constructions receive their interpretation.

In identifying the information conveyed by the words of the sentence (Hale, 2001; Shannon, 1948) as central to determining processing effort, Surprisal Theory is consistent with probabilistic models of comprehension. Essentially, the theory hypothesises that processing effort is determined by the expectedness of an incoming word, given all possible analyses of the prior input, rather than being determined by the cost of constructing those analyses. In the context of the present findings, this suggests that the observed increase in processing effort for coercion constructions (at least the sort investigated here), is not a consequence of recovering enriched representations *per se*, but rather their lower predictiveness prior to the coerced noun, and thus higher surprisal.

The results highlight the consequences of the *causal bottleneck* that Surprisal Theory entails (Levy, 2008), namely that it may be difficult to draw conclusions about highly specific processing mechanisms from our current inventory of empirical measures, if those measures are indeed dominated by contextual probability. Nonetheless, the pervasiveness of surprisal-driven effects, does provide increasing support for rational theories of language comprehension in which gradient notions of predictability and probability play a central role.

Acknowledgments

This research was partially supported by the SFB 1102 "Information density and linguistic encoding" awarded by the Deutsche Forschungsgemeinschaft (DFG). We thank Vera Demberg for useful discussion and advice concerning the surprisal estimates for our stimuli. We thank Nathaniel J. Smith and two anonymous reviewers for their helpful comments and suggestions.

[5] This effect was actually driven by the fact that the N400 elicited by coercing nouns was more sustained than that elicited by anomalous nouns. However, as the authors argue in the discussion, the sustained negativity can also be seen as a sequence of N400s elicited by the noun and the subsequent words. In any case, in the time-window typical of the N400 (300–500 ms), there was no significant difference between the effects evoked by coercing and anomalous nouns.

Appendix A. Items used in Experiment 1 and 2 and their translation in English

The first verb represents the coercion condition, the second verb represents the neutral condition and the third verb the preferred condition. The first noun is of higher frequency, the second of lower frequency. The first 24 items were used in the EEG study (Experiment 2), with minor changes.

(1) Peter begann/kaufte/las das Buch/Exposé im Urlaub.
Peter began/bought/read the book/report on vacation.

(2) Diego genoß/kaufte/aß das Dessert/Nugat mit großer Freude.
Diego enjoyed/bought/ate the dessert/nougat with great pleasure.

(3) Felix beherrschte/benutze/spielte die Violine/Oboe seit seiner Kindheit.
Felix mastered/used/played the violin/oboe since his childhood.

(4) Sina probierte/bekam/aß die Suppe/Brühe als Vorspeise.
Sina tried/got/ate the soup/broth as appetizer.

(5) Lukas begann/bemerkte/sang den Song/Refrain erst spät.
Lukas began/noticed/sang the song/refrain slowly.

(6) Lisa bevorzugte/lagerte/trank den Champagner/Cider sehr kalt.
Lisa preferred/stored/drank the champagne/cider very cold.

(7) Sonja schaffte/verkaufte/las den Roman/Krimi vor dem Morgengrauen.
Sonia managed/sold/read the novel/crime story before the dawn.

(8) Daniel genoß/hatte/spielte die Gitarre/Harfe seines Vater.
Daniel enjoyed/had/played the guitar/harp of his father.

(9) Silke bevorzugte/kaufte/trug die Kette/Spange mit dem roten Stein.
Silke preferred/bought/wore the necklace/clip with the red stone.

(10) Mark begann/verkaufte/malte das Gemälde/Fresko Freitag Abend.
Mark began/sold/painted the painting/fresco Friday evening.

(11) Nora ertrug/verstaute/hörte das Horn/Saxofon abends um zehn.
Nora endured/stored/heard the horn/saxophone in the evening at ten.

(12) Sina bevorzugte/bekam/goß die Pflanze/Orchidee von ihrem Freund.
Sina preferred/got/watered the plant/orchid of her friend.

(13) Vera genoß/bekam/roch das Parfum/Duftwasser von ihrer Freundin.
Vera enjoyed/got/smelled the perfume/toilet water of her friend.

(14) Andreas probierte/bekam/trank den Cocktail/Longdrink mit einem Strohhalm.
Andreas tried/got/drank the cocktail/long drink with a straw.

(15) Katharina begann/sah/backte den Kuchen/Stollen am Tag des Geburtstages.
Katherina began/saw/baked the cake/stollen on the day of the birthday.

(16) Lisa plante/machte/kochte die Nudeln/Spaghetti für den Hauptgang.
Lisa planned/made/cooked the noodles/spaghetti for the main course.

(17) Alexander bevorzugte/lagerte/spielte das Klavier/Spinett im Salon.
Alexander preferred/stored/played the piano/spinet in the living room.

(18) Hannah begann/beobachtete/mischte die Substanz/Tinktur Montag Abend.
Hannah began/observed/mixed the substance/tincture on Monday afternoon.

(19) Manuel liebte/kaufte/aß die Chips/Cracker von Aldi.
Manuel loved/bought/ate the chips/crackers from Aldi.

(20) Silke begann/bekam/aß die Pizza/Calzone mit großem Appetit.
Silke began/received/ate the pizza/calzone with great appetite.

(21) Michael bevorzugte/verstaute/hörte die Musik/Balladen aus den 80er Jahren.
Michael preferred/stored/heard the music/ballads from the 80s.

(22) Antonia begann/lagerte/aß die Knödel/Klöße rechtzeitig.
Antonia began/stored/ate the dumpling/dumpling in time.

(23) Miriam genoß/bekam/las die E-Mail/Fabel ihrer Kinder.
Miriam enjoyed/received/read the email/tale of her children.

(24) Florian bevorzugte/besorgte/schluckte die Tabletten/Pillen nach dem Essen.
Florian preferred/found/swallowed the tablets/pills after lunch.

(25) Melanie hasste/erklärte/stellte die Frage/Falle immer wieder.
Melanie hated/clarified/asked the question/trap over and over.

(26) Lena probierte/bekam/trank das Gemisch/Gebräu bevor sie losging.
Lena tried/got/drank the mixture/brew before she left.

(27) Wolfgang plante/bemerkte/veröffentlichte die Anzeige/Annonce schon einen Tag vorher.
Wolfgang planned/noticed/published the advertisement/advertisement already one day in advance.

(28) Joachim beherrschte/verstaute/steuerte das Boot/Kanu wie immer.
Joachim mastered/stored/steered the boat/canoe as always.

(*continued on next page*)

Appendix A (*continued*)

(29) Tom genoß/machte/besuchte die Vorlesung/Feier mit seinen Freunden.
Tom enjoyed/made/attended the lecture/cerimony with his friends.

(30) Philipp ertrug/bekam/hörte die Kritik/Schelte jeden Tag.
Philipp endured/received/heard the criticism/scolding every day.

(31) Fabian beherrschte/benutzte/fuhr das Motorrad/den Roller wie sein eigenes.
Fabian mastered/used/drove the motorcycle/the scooter like his own one.

(32) Hannah hasste/besorgte/packte den Koffer/Ranzen schon jetzt.
Hannah hated/found/packed the bag/wallet by now.

(33) Tim genoß/kennt/erzählte die Geschichte/Saga über den Hobbit.
Tim enjoyed/knew/told the story/saga about the Hobbit.

(34) Manuel beherrschte/benutzte/bediente den Computer/Rechner so oder so.
Manuel mastered/used/operated the computer/calculator one way or the other.

(35) Lukas genoß/kaufte/fuhr den Wagen/Schlitten am Wochenende.
Lukas enjoyed/bought/drove the car/sled on the weekend.

(36) Florian liebte/bekam/schaute den Film/Trailer über den berühmten Rennfahrer.
Florian loved/got/watched the movie/trailer on the famous racer

(37) Katharina bevorzugte/kaufte/streichelte das Kaninchen/Frettchen mit dem weichen Fell.
Katharina preferred/bought/petted the bunny/ferret with the soft fur.

(38) Diego beherrschte/machte/sprach den Dialekt/Slang wie alle anderen.
Diego mastered/made/spoke the dialect/slang as all the others.

(39) Lars hasste/benutzte/hielt die Reden/Vorträge immer wieder.
Lars hated/used/held the speeches/talks over and over.

(40) Tanja genoß/nahm/trug das Kleid/Trikot in Grün.
Tanja endured/took/wore the dress/jersey in green.

(41) Jonas bevorzugte/verkaufte/spielte die Geige/Fiedel mit der verzierten Schnecke.
Jonas preferred/sold/played the violin/fiddle with the decorated scroll.

(42) Chris bevorzugte/erklärte/spielte Fußball/Wasserball bei gutem Wetter.
Chris preferred/explained/played football/water polo with good weather.

(43) Malte plante/machte/besuchte die Veranstaltung/Gala Freitag Abend.
Malte planned/made/attended the meeting/gala Friday evening.

(44) Hannah beherrschte/bemerkte/fuhr den Traktor/Stapler von Anfang an.
Hannah mastered/noticed/drove the tractor/lift truck from the beginning.

(45) Lisa bevorzugte/machte/studierte Psychologie/Chirologie an der Universität des Saarlandes.
Lisa preferred/made/studied psychology/chirology at the university of Saarland.

(46) Dan genoß/kaufte/roch die Rose/Minze mit großer Freude.
Dan enjoyed/bought/smelled the rose/mint with great pleasure.

(47) Tanja ertrug/bemerkte/hörte den Krach/Rabatz in der Frühe.
Tanja endured/noticed/heard the noise/racket in the early morning.

(48) Eva genoß/konnte/löste das Rätsel/Mysterium auf Anhieb.
Eva enjoyed/knew/solved the puzzle/mystery at first go.

References

Ashby, J., Rayner, K., & Clifton, C. (2005). Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. *Quarterly Journal of Experimental Psychology, 58A*, 1065–1086.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.

Baggio, G., Choma, T., van Lambalgen, M., & Hagoort, P. (2010). Coercion and compositionality. *Journal of Cognitive Neuroscience, 22*(9), 2131–2140.

Baroni, M., Bernardini, S., Ferraresi, A., & Zanchetta, E. (2009). The wacky wide web: A collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation, 43*(3), 209–226.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*, 255–278.

Bates, D. M., & Sarkar, D. (2007). lme4: Linear mixed-effects models using s4 classes. R package version 0.9975-13.

Bemis, D. K., & Pylkkänen, L. (2011). Simple composition: A magnetoencephalography investigation into the comprehension of minimal linguistic phrases. *Journal of Neuroscience, 31*(8), 2801–2814.

Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about semantic illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research, 1446*, 127–143.

Calvo, M. G., & Meseguer, E. (2002). Eye movements and processing stages in reading: Relative contribution of visual, lexical, and contextual factors. *Spanish Journal of Psychology, 5*, 66–77.

Chow, W. Y., & Phillips, C. (2013). No semantic illusions in the semantic P600 phenomenon: ERP evidence from mandarin chinese. *Brain Research, 1506*, 76–93.

Clifton, C. J., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. In R. V. Gompel, M. Fisher, W. Murray, & R. L. Hill (Eds.) (pp. 341–372). Oxford: Elsevier Ltd..

Dambacher, M., & Kliegl, R. (2007). Synchronizing timelines: Relations between fixation durations and n400 amplitudes during sentence reading. *Brain Research, 1155*, 147–162.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*(89), 1117–1121.

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

Federmeier, K., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language, 41*, 469–495.

Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2013). Word surprisal predicts N400 amplitude during reading. In *Proceedings of the 51st annual meeting of the association for computational linguistics* (pp. 878–883). ACL.

Frank, S. L., Otten, L. J., Galli, G., & Vigliocco, G. (2015). The ERP response to the amount of information conveyed by words in sentences. *Brain and Language, 140*, 1–11.

Frisson, S., & McElree, B. (2008). Complement coercion is not modulated by competition: Evidence from eye-movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(1), 1–11.
Frisson, S., Rayner, K., & Pickering, M. J. (2005). Effects of contextual predictability and transitional probability on eye movements during reading. *Experimental Psychology: Learning, Memory, and Cognition, 31*, 862–877.
Greenhouse, S., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika, 24*, 95–112.
Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the second meeting of the north american chapter of the association for computational linguistics on language technologies* (pp. 1–8). Stroudsburg, PA, USA: Association for Computational Linguistics.
Halgren, E., Dhond, R. P., Christensen, N., Petten, C. V., Marinkovic, K., Lewine, J. D., & Dale, A. (2002). N400-like magnetoencephalography responses modulated by semantic context, word frequency, and lexical class in sentences. *Neuroimage, 17*, 1101–1116.
Hand, C. J., Miellet, S., O'Donnell, P. J., & Sereno, S. C. (2010). The frequency-predictability interaction in reading: It depends on where you're coming from. *Journal of Experimental Psychology - Human Perception and Performance, 36*, 1294–1313.
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.
Henderson, J. M., Choi, W., Lowder, W. M., & Ferreira, F. (2016). Language structure in the brain: A fixation-related FMRI study of syntactic surprisal in reading. *NeuroImage, 132*, 293–300.
Hoeks, J. C. J., Stowe, L. A., & Doedens, G. (2004). Seeing words in context: The interaction of lexical and sentence level information during reading. *Cognitive Brain Research, 19*(1), 59–73.
Husband, E. M., Kelly, L. A., & Zhu, D. C. (2011). Using complement coercion to understand the neural basis of semantic composition: Evidence from an FMRI study. *Journal of Cognitive Neuroscience, 23*(11), 3254–3266.
Jackendoff, R. (2002). *Foundations of language*. New York: Oxford University Press.
Kennedy, A., Pynte, J., Murray, W. S., & Paul, S. A. (2013). Frequency and predictability effects in the Dundee corpus: An eye movement analysis. *Quarterly Journal of Experimental Psychology, 66*(3), 601–618.
Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language, 52*, 205–225.
Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology, 16*, 262–284.
Kolk, H. H. J., Chwilla, D. J., van Herten, M., & Oor, P. J. (2003). Stucture and limited capacity in verbal working memory: A study with event-related potentials. *Brain and Language, 85*, 1–36.
Kuperberg, G. R., Choi, A., Cohn, N., Paczynski, M., & Jackendoff, R. (2010). Electrophysiological correlates of complement coercion. *Journal of Cognitive Neuroscience, 22*(12), 2685–2701.
Kuperberg, G. R., Kreher, D. A., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2007). The role of animacy and thematic relationships in processing active english sentences: Evidence from event-related potentials. *Brain and Language, 100*, 223–238.
Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research, 17*(1), 117–129.
Kutas, M., & Federmeier, K. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*(12), 463–470.
Kutas, M., & Federmeier, K. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62*, 621–647.
Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*, 161–163.
Lau, E., Almeida, D., Hines, P. C., & Poeppel, D. (2009). A lexical basis for n400 context effects: Evidence from MEG. *Brain and Language, 111*, 161–172.
Lau, E., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the n400. *Nature Reviews Neuvoscience, 9*, 920–933.
Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*(3), 1126–1177.
McElree, B., Frisson, S., & Pickering, M. J. (2006). Deferred interpretations: Why starting dickens is taxing but reading dickens isn't. *Cognitive Science, 30*, 115–124.
McElree, B., Traxler, M., Pickering, M. J., Seely, R., & Jackendoff, R. (2001). Reading time evidence for enriched composition. *Cognition, 78*, B17–B25.
Miellet, S., Sparrow, L., & Sereno, S. C. (2007). Word frequency and predictability effects in reading French: An evaluation of the E-Z reader model. *Psychonomic Bulletin and Review, 14*(4), 726–769.
Piantadosi, S., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences, 108*(9), 3526.
Piantadosi, S., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition, 122*, 280–291.
Pickering, M. J., McElree, B., & Traxler, M. (2005). The difficulty of coercion: A response to de almeida. *Brain and Language, 93*, 1–9.
Pustejovsky, J. (1995). *The generative lexicon*. Cambridge, Massachusetts: The MIT Press.
Pylkkänen, L., Brennan, J., & Bemis, D. K. (2011). Grounding the cognitive neuroscience of semantics in linguistic theory. *Language and Cognitive Processes, 26*(9), 1317–1337.
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. (2007). An meg study of silent meaning. *Journal of Cognitive Neuroscience, 19*, 1905–1921.
Rayner, K., Aschby, J., Pollatsek, A., & Reichle, E. D. (2004). The effects of frequency and predictability on eye fixations in reading: Implications for the EZ reader model. *Journal of Experimental Psychology - Human Perception and Performance, 30*, 12.
Rayner, K., Binder, K. S., Ashby, J., & Pollatsek, A. (2001). Eye movement control in reading: Word predictability has little influence on initial landing positions in words. *Vision Research, 41*(7), 943–954.
Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Hillsdale, NJ: Erlbaum.
Rayner, K., & Well, A. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review, 3*, 504–509.
Rayson, P., Charles, O., & Auty, I. (2012). Can google count? Estimating search engine results consistency. In A. Kilgarriff & S. Sharoff (Eds.), *Proceedings of the seventh web as corpus workshop (wac7)* (pp. 23–30).
Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal, 27*(3), 379–423.
Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition, 128*, 302–319.
Traxler, M. J., McElree, B., Williams, R., & Pickering, M. (2005). Context effects in coercion: Evidence from eye movements. *Journal of Memory and Language, 53*, 1–25.
Traxler, M. J., Pickering, M. J., & McElree, B. (2002). Coercion in sentence processing: Evidence from eye-movements and self-paced reading. *Journal of Memory and Language, 47*, 530–547.
van Herten, M., Chwilla, D. J., & Kolk, H. H. J. (2006). When heuristics clash with parsing routines: ERP evidence for conflict monitoring in sentence perception. *Journal of Cognitive Neuroscience, 18*(7), 1181–1197.
Willems, R. M., Frank, S. L., Nijhof, A. D., Hagoort, P., & van den Bosch, A. (2015). Prediction during natural language comprehension. *Cerebral Cortex*, 1–11.