Title

Context reduces coercion costs — Evidence from eyetracking during reading

Permalink

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 39(0)

Author

Bott, Oliver

Publication Date

2017

Peer reviewed

Context reduces coercion costs – Evidence from eyetracking during reading

Oliver Bott (oliver.bott@uni-tuebingen.de)

Project *Composition in Context* (CiC), DFG Priority Program *XPrag.de*

SFB 833, Nauklerstraße 35, 72074 Tübingen, Germany

Abstract

This paper presents an eyetracking during reading experiment that investigated the role of supportive context on processing aspectual coercion. Coercion sentences in need of aspectual enrichment were embedded in discourse contexts providing the necessary information for successful interpretation. The findings of the reported experiment show that context information can be used immediately without disrupting reading of coercion sentences. The lack of coercion costs in supportive discourse contexts provides experimental evidence for the proposed *Composition in Context Hypothesis* and against theories that view semantic composition as largely encapsulated from context. Furthermore, the present experiment investigated the role of inter-individual differences in verbal working memory capacity on the immediate use of contextual information in computing coerced interpretations.

Keywords: Aspectual coercion; Discourse context; Semantic processing; Eyetracking during reading; Working memory capacity

Introduction

A central assumption in semantics is that interpretation is governed by the principle of compositionality (see, e.g., Pelletier, 1994). The prinholds that the meaning of a complex expression is entirely determined by the meaning of its parts and their syntactic combination. However, linguistic expressions are at the same time highly context dependent, and the language interpretation system is therefore not only dependent on the parts of complex expressions in a bottom-up fashion, but also has to be open to top-down influences of the context of utterance. The present study investigates the interplay between sentential and contextual information during the on-line composition of the event interpretation. In particular, I tested whether contextual information is immediately used to resolve compositional conflicts during online interpretation.

Cases of coercion have prominently figured in studies on the time course of compositional interpretation, see e.g. Piñango and Deo (2016) for an overview of studies on complement coercion, and Bott (2010) and Paczynski, Jackendoff, and Kuperberg (2014) for aspectual coercion. (1) displays a compositional conflict calling for aspectual coercion.

(1)   # Yesterday, Peter jogged in only thirty minutes.

When uttered out of the blue, sentence (1) is hardly interpretable. This is because an *in-adverbial* requires a telic event predicate of the accomplishment type (Vendler, 1957), but *Peter jog-* expresses an atelic activity. Under a coercion analysis *Peter jog-* has therefore to be shifted into an accomplishment, i.e. the event representation of the activity has to be enriched by adding a culminating event. The required operation can be summarized as follows:

(2)   [in thirty minutes[**Peter jog-**]] $\rightsquigarrow_{coerce}$ [in thirty minutes[ADD CULMINATION[**Peter jog-**]]]

The inserted coercion operator **ADD CULMINATION** is a function that takes as input an activity and outputs an accomplishment (see, e.g., Dölling, 2014, for semantic representations of various coercion operators). The coercion operation solves the compositional problem. After the inclusion of the appropriate type shifting operator the resulting representation can be interpreted fully compositionally. Interestingly, the semantic problem in (1) seems to completely disappear once the sentence is embedded in an appropriate discourse context. Consider (1) in the context of (3).

(3)   Half a year ago, Peter started to jog **four kilometers**$_{culmination}$ every day. When he began, he was quite slow but now he is really fast.

Based on the pragmatic literature (Recanati, 2010) two theoretical alternatives can be contrasted on how compositional interpretation might make use of contextual information. The *Composition in Context (CiC)* hypothesis predicts immediate availability of contextual information (e.g., Nieuwland & van Berkum, 2006). Accordingly, the bounded path *four kilometers* from the preceding context should be immediately available when composing the adverbial with the rest of the target sentence in (1). Alternatively, however, compositional interpretation may operate in strictly locally in a bottom-up fashion (Cappelen & Lepore, 2005). According to this view, which may be characterized as *Encapsulated Composition (EC)* hypothesis, contextual information is only considered when the sentence information is not sufficient: Either to resolve compositional conflicts or to interpret context dependent expressions that are context dependent. In fact, in coercion theories it is standardly assumed that coercion operations are locally triggered by temporary semantic mismatch, i.e. many coercion analyses employ the EC hypothesis (cf. de Swart, 1998, p. 8).

According to the EC hypothesis the initial interpretation of (1) in the context of (3) should result in an aspectual mismatch that is only resolved in a second processing step. This should lead to measurable disruption during online interpretation. The *CiC* hypothesis, by contrast, predicts no processing costs of coercion sentences relative to non-coercing controls because the contextually given culmination can go right into the composed meaning.

Context effects on coercion have only been investigated in a small number of online studies so far. Traxler, McElree, Williams, and Pickering (2005) report a series of self-paced reading and eyetracking during reading experiments in which

they presented target sentences involving complement coercion (*the student began the book*: [**begin**[**the book**]] $\leadsto_{coerce}$ [**begin**[**to do something involving**[**the book**]]]) and manipulated the preceding discourse context. Here is a sample item from Traxler et al. (2005, Exp. 3/4); c1/2 correspond to coercion vs. control contexts, t1/2 are coercion vs. control targets, critical regions underlined.

(4) c1) The student started <u>a book</u> in his dorm room.
     c2) The student read <u>a book</u> in his dorm room.
     t1) Before he started <u>it</u>, he checked his e-mail.
     t2) Before he read <u>it</u>, he checked his e-mail.

Contextual facilitation was found in cases in which the context sentence either itself required complement coercion (c1), or explicitly introduced the relevant event (c2). In their Exp. 3 and 4, Traxler et al. (2005) observed coercion costs for c1 relative to c2. Crucially, however, the anaphor *it* in the coercion target t1 was read as fast as in the control condition t2 showing that the anaphor could be immediately linked to the relevant event in the discourse representation. This is prima facie evidence in favour of the *CiC* hypothesis. However, there is another explanation consistent with the EC hypothesis too. It is also possible that *it* was not interpreted as anaphoric to *the book* but rather as an event anaphor, which would make coercion unnecessary in the first place. The following example adopted from Asher (1993, p. 233) illustrates the difference between individual and event anaphora.

(5)   If Timmy$_i$ hits$_j$ John$_k$, it$_{*i/j/*k}$ will cause a fight.

As for aspectual coercion, Bott (2010, Exp. 3) reports a self-paced reading experiment investigating whether German contexts such as (3) facilitate aspectual enrichment in German versions of (1). The findings provide preliminary evidence that supportive context can eliminate coercion costs observed for this type of sentences when presented out of context. After telic contexts such as (3) the reading times of coercion sentences did not differ from control targets allowing for plain compositional interpretation. However, this interpretation of the results is complicated by the fact that the comparison crucially involved a direct comparison between different adverbials (German *in x time* vs. *for x time*).

Processing along the lines of the EC or the CiC hypothesis does not have to be an either or choice but could be subject to inter-individual differences. In particular, verbal working memory capacity may be an important constraining factor for being able to employ a highly context dependent processing strategy as put forward in the CiC hypothesis. Note that the CiC haypothesis presupposes full accessibility of all the relevant contextual information. Processing along the lines of the CiC hypothesis can therefore be expected to require more working memory capacity than the strictly local interpretation of the EC hypothesis. Existing research on sentential context effects has provided evidence that low-span readers make even less use of the immediate sentential context than high-span readers (van Petten, Weckerly, McIsaac, & Kutas, 1997).

Thus, smooth aspectual enrichment in line with the CiC hypothesis may be especially expected for high-span readers, whereas low-span readers should be more likely to exhibit coercion costs. Inter-individual differences in working memory capacity have not been taken into account in coercion studies so far.

The present experiment studied contextual facilitation effects in aspectual coercion with four major modifications relative to prior research. First of all, the present study employed eyetracking during reading – an online method that provides us with a richer picture about the time course of interpretation than self-paced reading. Secondly, a larger set of experimental items was tested, and these materials were set up in such a way that the critical region was kept identical across conditions. Thirdly, the materials were more carefully pretested concerning their offline interpretation than the ones used in Bott (2010, Exp. 3). Finally, contextual facilitation effects were related to participants' verbal working memory capacity as measured by the reading-span task (Daneman & Carpenter, 1980).

The experimental design of the present study included a coercion and a control condition as well as a mismatch condition. Two kinds of contexts were constructed. Both, telic contexts such as (3), and atelic contexts such as (6) introduced a repetitive event (contexts translated from German). The only difference is that the *telic context* (3) establishes a series of telic, bounded events (e.g., *jog four kilometers*) while the *atelic context* (6) introduces a series of atelic, unbounded activities instead. Both types of contexts put emphasis on the actual duration of the respective events at reference time *now*.

(6)   Half a year ago, Peter started to jog every day. When he began, he had to stop after a short time but now he can run for quite a long time.

Target sentences were of two types manipulating the adverbial: telic sentences (7-a) including German *in*-adverbials, and atelic sentences (7-b) with German *for*-adverbials.

(7) a. Als es ihm vorhin gelang, in nur dreißig
         When he just   managed in only thirty
         Minuten zu <u>joggen</u>,...
         minutes to <u>jog</u>...
         'When he managed to jog in only 30 min. ...'
     b. Als es ihm vorhin gelang, ganze dreißig
         When he just   managed for   thirty
         Minuten zu <u>joggen</u>,...
         minutes to <u>jog</u>...
         'When he managed to jog for 30 min. ...'
     c. ...war er sehr stolz auf sich.
         ...he was very proud of himself.

Discourse conditions were as follows. The coercion condition was constructed by combining telic contexts (3) with telic targets (7-a). The control condition combined atelic contexts (6) with atelic targets (7-b). For the aspectual mismatch condition atelic contexts (6) were paired with telic targets (7-a).
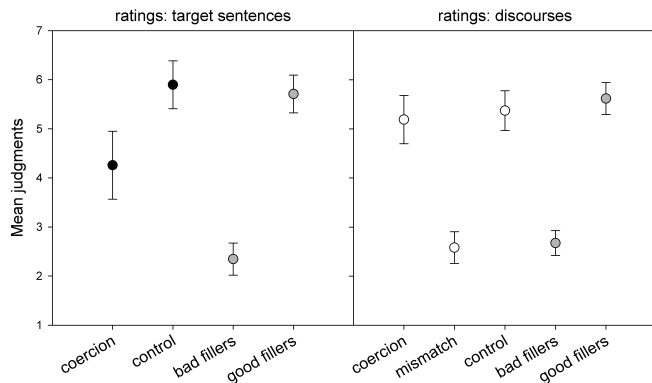
Figure 1: Mean judgments in Pretest 1 (left-hand side) and Pretest 2 (right-hand side). Error bars represent 95% confidence intervals (by-participants analysis).

Pretests

Two pretests were conducted. The first pretest was a sentence acceptability judgment experiment testing the telic (7-a) and the atelic (7-b) target sentences out of context. The second pretest was a discourse acceptability rating experiment that queried the felicity of the discourses in the coercion, the control, and the mismatch condition, respectively. The predictions are straightforward. Due to their need of coercion, telic target sentences should be less acceptable than atelic target sentences when encountered out of context. Supportive context should increase the acceptability of the coercion condition, though. After a telic context, a telic target should become as acceptable as the control condition. The mismatch condition should be judged as infelicitous.

Pretest 1

Method   20 native German speakers (mean age: 26.9 years; 17 female) participated in the pretest for a payment of €5. Participants rated the acceptability of the telic and atelic target sentences on a scale from 1–7 from completely unacceptable to fully acceptable.

Target sentences were taken from the set of 24 items created for the eyetracking study. All were constructed following the scheme exemplified in (7). Pronouns were replaced by the proper names from the contexts (3)/(6). Two lists were constructed using a Latin square design and 100 filler sentences were added to each list. 60 of the fillers were infelicitous ('*bad fillers*') while the others were fully acceptable ('*good fillers*'). Ten participants were randomly assigned to each list.

Paricipants were tested individually in a quiet computer pool. Sentence materials were presented in randomized order in a single block which was preceded by a short practice of five trials. An experimental session took less than 30 minutes.

Results and discussion   The mean judgments are shown in Figure 1. As predicted, acceptability of telic target sentences was judged significantly worse than atelic target sentences

$(t_1(19) = -5.4, p < .01; t_2(23) = -9.52, p < .01)$. The latter were judged even slightly better than the well-formed fillers (*atelic targets*: 5.90; *well-formed fillers*: 5.70) suggesting that the target sentences in the control condition are in fact fully acceptable. The telic target sentences received mean ratings of 4.24 and were thus well above the nonsensical fillers with a mean rating of 2.34. Even though the telic target sentences were perceived as not fully acceptable when presented out of the blue, participants seemed to be aware of the fact that these sentences are in fact well-formed if embedded in an appropriate discourse context.

Pretest 2

Method   30 new participants (mean age: 26.2 years; 16 female), all native speakers of German, took part in the pretest for a payment of €5. Participants rated the acceptability of the discourses in the coercion, the control, and the mismatch condition on a scale from 1–7. In addition to the 24 experimental items 66 filler discourses were included (33 acceptable and 33 incoherent discourses). 20 of the incoherent filler discourses were globally incoherent, e.g. *Lisa is very bad in maths. [. . .] . So, she wasn't surprised when she got an A.*, and 13 were locally incoherent, e.g. *. . . the jockey sat in his horse . . . .* The items plus the fillers were distributed to three lists in a Latin square design. Ten participants were randomly assigned to each list. The procedure was the same as in the previous pretest

Results and discussion   The mean judgments are also shown in Figure 1. As predicted, telic targets preceded by a telic context made the coercion condition fully acceptable. Paired t-tests revealed that the coercion condition did not differ reliably from the control condition $(t_2(29) = -1.65, p = .11; t_1(23) = -.76, p = .46)$. Both, coercion (mean rating: 5.10) and control (mean rating: 5.27) received ratings in the range of the good fillers (mean rating: 5.82). As expected, the mismatch condition was judged similar to the incoherent fillers. Repeated measures ANOVAs revealed that mismatch was judged significantly worse than coercion and control $(F_1(2,58) = 81.77, p < .01; F_2(2,46) = 122,56, p < .01)$.[1]

Taken together, the results of the pretests show that the materials tested in the eyetracking study fully meet the assumptions stated in the introduction. The coercion targets are not fully interpretable on their own but require contextual support. Embedded in a telic context, however, the telic targets become fully acceptable. After atelic contexts, however, telic target sentences result in aspectually incoherent discourses.

Eyetracking Experiment

The EC and the CiC hypotheses make fundamentally different predictions regarding the online processing of the three discourse conditions.

---

[1]In all analyses in the present paper including the three-level factor DISCOURSE CONDITION the degrees of freedom were corrected by applying the Greenhouse-Geisser correction. In the text the uncorrected degrees of freedom are reported.

According to the EC hypothesis, the coercion operation is triggered by a temporary semantic mismatch during the intial interpretation of atelic activity verbs modified by *in*-adverbials. Therefore, during first-pass reading the coercion condition should pattern with the mismatch condition and lead to processing difficulty relative to the control condition.

By contrast, the CiC hypothesis predicts smooth interpretation of the coercion condition. Coercion should therefore pattern with the control condition, and it should only be the mismatch condition that causes processing difficulty.

In order to assess effects of inter-individual differences in working memory capacity, between group analyses were conducted for high- versus medium- vs. low-span readers.

Methods

Participants 48 new participants (mean age: 24.1 y., range 20 – 32 y.; 40 female) all native speakers of German with normal or corrected-to-normal vision took part in the experiment for a payment of €8. Based on their performance in Daneman and Carpenter's (1980) reading span task they were divided into three groups. The first group consisted of 10 participants with a reading span of 3.0, the median value in the sample. This group was included as MEDIUM SPAN readers in ANOVA analyses including the between factor READING SPAN. The group of LOW SPAN readers consisted of 19 participants with a mean reading span of 2.4 (range 2.0 – 2.5). The group of HIGH SPAN readers consisted of 19 participants with a mean reading span of 3.7 (range 3.5 – 4.5).

Materials The sentence materials were identical to those used in the pretest and the items were always presented on three lines. The first line contained the first context sentence, the second line the second context sentence, and the third line the target sentence. The target sentences were split up in ten regions of interest (ROIs):

(8)   Als es ihm| heute| gelang,|$_{adv1}$ in nur|$_{adv2}$ dreißig|$_{adv3}$ Minuten|$_{verb}$ zu joggen,|$_{coda1}$ war er|$_{coda2}$ sehr stolz|$_{coda3}$ auf sich.|

The critical region was the verb ROI. Note that any aspectual mismatch only becomes evident at this ROI. For instance, *als es ihm heute gelang, in nur dreißig Minuten vier Kilometer weit zu joggen, . . .* (*when he managed today to jog four kilometers in only thirty minutes . . .*) would be fully acceptable even after an atelic context.

Apparatus and procedure Eye movements of the dominant eye were recorded with an SR Research Ltd. Eyelink 1000 eyetracker. The trial began with the presentation of a screen which served as calibration check and for drift correction with a yellow dot in the position where the centre of the first word would appear. If no fixation on the dot was registered within five seconds, recalibration was enforced. Otherwise, texts were presented. After reading, participants had to move their eyes to a yellow dot in the right bottom corner of the screen which triggered the presentation of the judgment screen. Judgments had to be provided by pressing the left or the right button of a gamepad.

The experiment started with five discourses for practice, followed by the individually randomized presentation of the experimental trials in three blocks. A typical experimental session lasted less than 45 minutes. Immediately after the eyetracking experiment each participant was subjected to an experimenter-administered version of the reading span task (Friedman & Miyake, 2004). Reading span was scored as follows. The highest stage for which at least two out of a total of three sequences could be correctly recalled determined a participant's basic reading span. If she was able to correctly recall one sequence from an even higher stage, a value of 0.5 was added to this basic value.

Eyetracking analysis Prior to all analyses the eyetracking data were preprocessed. Two trials with major track loss were exluded, and all fixations immediately preceding or following a blink were eliminated. All fixations shorter than 80 ms and further than 0.5 degrees from the last or next fixation as well as fixations longer than 800 ms were eliminated. Preprocessing affected 2.7% of all fixations.

The analyzed measures of first-pass reading included *first fixation durations*, *first-pass times*, and *first-pass regression ratios*, i.e. the proportions of regressions made during first-pass reading. Measures related to rereading included *second-pass time* and the proportions of *regressions in*.

Results

The coercion condition was accepted 83.3% of all trials, control was accepted 86.9%, but mismatch was rejected 66.5%. Thus, discourses in the coercion and in the control condition were generally accepted while the aspectual mismatch condition was generally rejected. The analysis of judgment RTs, corrected for outliers by eliminating all RTs more than 2.5 standard deviations above a participant's mean RT, revealed no significant differences between the three discourse conditions ($F_{1/2} < 1$).

Table 1 presents the descriptive statistics for the eyetracking measures related to first-pass parsing. The findings for the measures related to rereading are shown in Table 2.

First-pass reading Immediately when readers encountered the verb region, *first fixation durations* were longer in the mismatch condition than in the control condition ($t_1(47) = 2.29, p < .05$; $t_2(23) = 2.12, p < .05$). By contrast, verbs in the coercion condition were read equally fast as verbs in the control condition ($|t_{1/2}| < 1$). Before or after the critical verb ROI there were no significant differences in first fixation durations for any of the ROIs in the target sentences.

The analyses of *first-pass times* further corroborated this finding. At the verb ROI a clear mismatch effect was found ($t_1(47) = 3.85, p < .01$; $t_2(23) = 3.72, p < .05$), but coercion did not differ from control ($|t_{1/2}| < 1$). The mismatch effect was again limited to the verb ROI, and discourse conditions did not differ reliably from each other at any target ROI.

Table 1: Mean *first fixation durations (FFD)*, mean *first-pass times (FPT)*, and mean *first-pass regression ratios (FPRR)* of the target sentences in the eyetracking experiment. Note: ROI *a3* corresponds to the final adverbial region, *verb* to the critical verb region, and *c1–c3* to the three sentence final ROIs.

		a3	verb	c1	c2	c3
FFD	*contr.*	205	251	238	252	256
(in ms)	*coerc.*	210	247	247	247	264
	mism.	213	266	240	243	257
FPT	*contr.*	217	311	290	333	403
(in ms)	*coerc.*	219	317	317	312	424
	mism.	219	345	304	307	417
FPRR	*contr.*	14.6	18.5	6.4	25.4	45.8
(in %)	*coerc.*	15.7	16.9	9.3	26.3	48.3
	mism.	16.2	27.1	12.2	21.9	51.2

Mismatch not only slowed down reading speed during first-pass reading, it also gave rise to more regressions from the verb ROI. The analysis of *first-pass regression ratios* showed that readers launched more regressions from mismatching verbs than from verbs in the control condition ($t_1(47) = 2.23, p < .05; t_2(23) = 3.08, p < .01$). Again, coercion did not differ from control ($|t_{1/2}| < 1$).

Taken together, the findings from the three measures reflecting the first-pass reading of the critical verb ROI show that the initial interpretation of the coercion targets was as smooth as that of the control targets. Without a preceding telic context in the mismatch condition, however, first-pass reading was severely disrupted.

Rereading The analyses of the *proportions of regressions in* revealed that readers regressed back to the three adverbial ROIs in the mismatch condition. For all three ROIs reliable mismatch effects were observed (first ROI: $t_1(47) = 2.52, p < .05; t_2(23) = 2.82, p < .05$; second ROI: $t_1(47) = 2.70, p < .05; t_2(23) = 2.84, p < .01$; third ROI: $t_1(47) = 2.02, p < .05; t_2(23) = 2.16, p < .05$). Also, on the verb ROI a mismatch effect was found that was marginally significant by subjects and significant by items ($t_1(47) = 1.87, p = .07; t_2(23) = 2.18, p < .05$). The coercion analyses showed that the adverbial ROIs did not receive more regressions in the coercion condition than in the control condition (all $|t_{1/2}| < 1$). However, it turned out that readers regressed more often back into the verb ROI than they did in the control condition. This was reflected by a (by-items marginally) significant coercion effect ($t_1(47) = 2.46, p < .05; t_2(23) = 1.97, p = .06$).

A similar pattern of effects was observed in the *second-pass times* too. The mismatch condition led to longer second-pass times than the control condition persisting from the adverbial ROIs (first ROI: $t_1(47) = 2.83, p < .01; t_2(23) = 2.23, p < .05$; second ROI: $t_1(47) = 2.40, p < .05; t_2(23) = 2.64, p < .05$; third ROI: $t_1(47) = 1.70, p = .10; t_2(23) = 1.80, p = .08$) to the verb ROI ($t_1(47) = 4.71, p < .01$;

Table 2: Mean *second-pass times (SPT)*, and mean proportions of *regressions in (RI)* of the relevant ROIs of the target sentences.

		a1	a2	a3	verb	c1	c2
SPT	*contr.*	81	106	56	81	88	115
(in ms)	*coerc.*	85	104	59	143	115	132
	mism.	129	160	89	167	87	108
RI	*contr.*	35.7	22.8	19.7	4.6	–	–
(in %)	*coerc.*	37.9	26.3	22.0	9.2	–	–
	mism.	47.5	33.5	30.7	8.1	–	–

$t_2(23) = 3.90, p < .01$). Also, a coercion effect was present, and, consistent with what was observed for the *regressions in*, this effect was limited to the verb ROI ($t_1(47) = 3.25, p < .01; t_2(23) = 2.46, p < .05$).

Taken together, the analyses of late eyetracking measures – besides substantial mismatch effects – show that participants were more likely to reread the verbs in the coercion condition than in the control condition.

Analyses contingent on reading span In order to investigate whether early and late effects were modulated by inter-individual differences in working memory capacity ANOVAs with the within-factor DISCOURSE CONDITION and the between factor READING SPAN (three levels: HIGH SPAN vs. MEDIUM SPAN vs. LOW SPAN) were computed analyzing the first and second-pass times of the verb ROI. Table 3 presents the mean reading times of the verb ROI split up by groups.

The analysis of *first-pass times* only revealed a significant main effect of DISCOURSE CONDITION ($F_1(2, 90) = 4.61, p < .05$), i.e. the above reported mismatch effect. Neither the main effect of READING SPAN nor its interaction with DISCOURSE CONDITION reached significance (both $F_1 < .05$). Thus, the three reading span groups did not differ with respect to their first-pass reading times of the verb.

The analysis of *second-pass times* also revealed no differences between the three groups. The main effect of DISCOURSE CONDITION was reliable ($F_1(2, 90) = 11.88, p < .01$), but neither the main effect of READING SPAN nor the interaction reached significance (both $F_1 < 1.3$).

To summarize, the three reading span groups had strikingly similar patterns of results. The three groups showed equally sized early effects of aspectual mismatch and late coercion effects that only started during rereading the sentence.

Discussion

The present study investigated whether coercion sentences embedded in supportive discourse context lead to measurable processing costs during their initial interpretation. According to the EC hypothesis, the compositional system operates strictly bottom-up, and the coercion targets should therefore lead to temporary aspectual mismatch during the initial inter-

Table 3: Mean first- (*FPT*) and second-pass times (*SPT*) of the verb ROI split up by reading span groups.

		FPT (ms)	*SPT* (ms)
low span	*contr.*	325	73
	coerc.	324	106
	mism.	353	126
medium span	*contr.*	313	72
	coerc.	316	158
	mism.	344	230
high span	*contr.*	296	95
	coerc.	309	171
	mism.	338	174

pretation and subsequent context-driven repair. During first-pass reading target sentences in the coercion and the mismatch condition should exhibit qualitatively similar processing effects. The CiC hypothesis, by contrast, predicts smooth interpretation of the coercion targets if embedded in a supportive context because the culminating event from the context should be immediately available.

The findings of the present eyetracking experiment unambiguously provide evidence against the EC hypothesis. While aspectual mismatch led to substantial processing difficulty during first-pass reading, none of the three analyzed early eyetracking measures indicated any difficulty in the coercion condition. The processing of aspectual coercion involved a qualitatively different time course than aspectual mismatch.

Do the findings support the CiC hypothesis, then? Above, it was stated that supportive context should completely eliminate all coercion costs. The coercion effects observed during rereading of the coercion targets may therefore be taken as prima facie evidence against the CiC hypothesis. However, the exposition of this hypothesis in the introduction was grossly oversimplistic. On closer consideration, the hypothesis is fully consistent with overall higher processing demands in the coercion than in the control condition as long as difficulty only emerges *after* a smooth first composition step. Two potential sources of difficulty come to mind. Under the CiC hypothesis comprehenders still have to be able to exactly recall the culmination introduced two sentences before. Note that in order to make proper sense of the coercion target it is important to know the exact length of the path. Obviously, to jog a distance of three miles in half an hour is plausible but thirty miles is not. Therefore, successful coercion crucially depends on a precise representation of the context in all its particulars. Another potential source of difficulty may be attributed to the metalinguistic evaluation demanded from the participants: since the target sentences were exactly the same in the coercion and the mismatch condition, the metalinguistic evaluation may have been more difficult in the coercion than in the control condition.

Both explanations, the (un-)availability of the culminating event in working memory as well as difficulty during metalin-guistic evaluation, would be consistent with the CiC hypothesis. The above reported analyses taking into account participants' working memory capacity suggest that the second explanation is more likely than the first.

References

Asher, N. (1993). *Reference to abstract objects in discourse*. Dordrecht: Springer.

Bott, O. (2010). *The processing of events*. Amsterdam: John Benjamins.

Cappelen, H., & Lepore, E. (2005). *Insensitive semantics*. Oxford: Blackwell.

Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450–466.

de Swart, H. (1998). Aspect shift and coercion. *Natural Language & Linguistic Theory*, *16*(2), 347–385.

Dölling, J. (2014). Aspectual coercion and eventuality structure. In K. Robering (Ed.), *Events, arguments and aspects – topics in the semantics of verbs* (p. 189-226). Amsterdam: John Benjamin.

Friedman, N. P., & Miyake, A. (2004). The reading span test and its predictive power for reading comprehension ability. *Journal of Memory and Language*, *51*, 1–25.

Nieuwland, M. S., & van Berkum, J. J. (2006). When peanuts fall in love: N400 evidence for the power of discourse. *Journal of Cognitive Neuroscience*, *18*(7), 1098–1111.

Paczynski, M., Jackendoff, R., & Kuperberg, G. (2014). When events change their nature. *Journal of Cognitive Neuroscience*, *26*(9), 1905–1917.

Pelletier, F. J. (1994). The principle of semantic compositionality. *Topoi*, *13*(1), 11–24.

Piñango, M. M., & Deo, A. (2016). Reanalyzing the complement coercion effect through a generalized lexical semantics for aspectual verbs. *Journal of Semantics*, *33*(2), 359–408.

Recanati, F. (2010). *Truth-conditional pragmatics*. Oxford: Oxford University Press.

Traxler, M., McElree, B., Williams, R., & Pickering, M. (2005). Context effects in coercion. *Journal of Memory and Language*, *53*, 1–25.

van Petten, C., Weckerly, J., McIsaac, H., & Kutas, M. (1997). Working memory capacity dissociates lexical and sentential context effects. *Psychological Science*, *8*(3), 238–242.

Vendler, Z. (1957). Verbs and times. *The Philosophical Review*, *66*(2), 143–160.