UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
Processing non-culminating accomplishments across languages

Permalink
https://escholarship.org/uc/item/1mb5x4cc

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors
Bott, Oliver
Solstad, Torgrim
Michaelis, Jens

Publication Date
2024

Copyright Information
This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Processing Non-Culminating Accomplishments across Languages

Oliver Bott (oliver.bott@uni-bielefeld.de)
Department of Linguistics, Bielefeld University & CRC 1646, Project B01

Torgrim Solstad (torgrim.solstad@uni-bielefeld.de)
Department of Linguistics, Bielefeld University & CRC 1646, Project B01

Jens Michaelis (jens.michaelis@uni-bielefeld.de)
Department of Linguistics, Bielefeld University & CRC 1646, Project B01

Abstract

We investigated the processing and interpretation of aspectual coercion in the case of non-culminating accomplishments in English and German, two languages which differ in interesting ways in their aspectual properties. Two offline experiments employing an inference rating task provide evidence that non-culminating accomplishments in both languages actually involve a shift in interpretation. Four self-paced reading experiments furthermore show that this type of coercion isn't costly – neither in German, a language lacking grammatical aspect, nor in English with an aspectual opposition between progressive and perfective simple past forms. This lack of online effect in processing coercion was obtained in a first pair of experiments using adverbial modification (sentence-internally) within the verb phrase and in a second pair of experiments in which aspectual coercion was triggered in a subsequent discourse unit. Finding no processing difficulty is in line with previous results for this type of aspectual coercion in German, but is inconsistent with findings reported for English (Bott & Hamm, 2014). A final stops-making-sense experiment replicates the lack of effect for English and furthermore shows that the processing of non-culminating accomplishments does not incur a processing effect even in a task calling for immediate full interpretation. The present study thus adds to existing evidence that aspectual coercion does not necessarily incur processing costs (Pickering, McElree, Frisson, Chen, & Traxler, 2006).

Keywords: aspectual coercion; lexical aspect; pragmatic inferencing; non-culminating accomplishments; telicity

Introduction

Languages differ with respect to the grammatical means they have to express semantic characteristics of events such as their completed- or ongoingness, a distinction generally encoded by grammatical aspect (Comrie, 1977, i.a.). At the same time, event interpretation shows great flexibility with systematic transitions between event types (Moens & Steedman, 1988; Dölling, 2014, a.o.), such as the four lexical aspectual classes discussed by Vendler (1957). Consider (1).

(1) The boy ate the pizza for five minutes.

The verb phrase *eat the pizza* is typically taken to express a telic event of accomplishment type (Krifka, 1998). However, modification by a *for*-adverbial shifts or coerces the accomplishment into an atelic activity interpretation (Bott, 2010; Dölling, 2014): From the sentence without the adverbial we derive the culmination inference that the whole pizza has been eaten, but adverbial modification leads to the cancellation of this inference (Martin, 2019). Moreover, *for*-modification is typically used as a test for predicates of the activity type rather than accomplishments and it has been questioned whether sentences such as (1) are grammatical at all (Vendler, 1957; Dowty, 1979). It has, however, been recognized that even perfective accomplishments exhibit varying degrees of culmination requirements with clear differences between perfective accomplishments across a range of languages (Koenig & Muansuwan, 2000; Arunchalam & Kothari, 2011; Bott & Hamm, 2014; Martin, 2019; Nadathur & Filip, 2021; Minor, Mitrofanova, Guajardo, Vos, & Ramchand, 2023; Kasher & Hacohen, 2023). Crucially, Martin (2019) proposes for English that non-culminating interpretations of accomplishments in fact require the presence of a *for*-adverbial, otherwise the culmination takes the form of an entailment and cannot be suspended, cf. (2). Accomplishments can be further subdivided into different sub-classes (Rappaport Hovav, 2008) and incremental theme predicates such as *eat the pizza* – especially when appearing in combination with a particle such as *up* – have been claimed to strongly implicate culmination (Arunchalam & Kothari, 2011).

(2) #The boy ate (up) the pizza. He didn't finish it, though.

Generally, it has thus been assumed that (perfective) accomplishments give rise to a default inference of event completion, which can be canceled if the sentence context or broader pragmatic context requires it. It is natural to assume that such cancellation should come at a cost. Accordingly, previous research on the imperfective paradox, that is, suspension of a culmination inference in imperfective accomplishments, has provided evidence for enhanced processing load in the form of a sustained anterior negativity in ERPs (Baggio, van Lambalgen, & Hagoort, 2008). The empirical picture is not entirely clear, though. Bott (2010) investigated aspectual coercion in non-culminating accomplishments and found no indication of difficulty in German equivalents of (1). Another type of aspectual coercion, iterative interpretations of punctual eventive predicates (*sneeze for an hour*) was shown by Pickering et al. (2006) not to incur any coercion costs in natural reading (see Brennan and Pylkkänen (2008); Paczynski, Jackendoff, and Kuperberg (2014) for evidence to the contrary). Coercion difficulty also seems to be task-dependent. While it was found for various kinds of coercion in stops-making-sense tasks (Todorova, Straub, Badecker, & Frank, 2000; Bott, 2010, Exp. 1), it was not found in self-paced reading or eyetracking during reading without a secondary task (Pickering et al., 2006).

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.* ©2024 The Author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

To resolve the issue of differences in coercion costs, Bott and Hamm (2014) suggested that aspectual processing may be subject to cross-linguistic differences in grammatical aspect and also fine-grained differences in lexical aspect. Bott and Hamm investigated the interpretation and processing of non-culminating accomplishments in English and German. In their cross-linguistic self-paced reading study, they compared the processing of sentences with culminating and non-culminating accomplishments of the following types (the critical region for self-paced reading underlined).

(3) a. The architect built the monument <u>for two years</u> after the city had provided the money for it.
b. ... built the monument <u>in two years</u> ...
c. ... was building the monument <u>for two years</u> ...

(4) a. ... errichtete das Monument <u>zwei Jahre lang</u> ...
b. ... errichtete das Monument <u>in zwei Jahren</u> ...

The study contrasted subtractive coercion from an accomplishment into an activity (van Lambalgen & Hamm, 2005) in (3a) and (4a) with an aspectual control condition in (3b) and (4b) employing *in*-adverbials, one of the classical tests for accomplishments and telicity (Vendler, 1957; Dowty, 1979). In English, perfective accomplishments in the simple past were also compared with accomplishments in the progressive, which give rise to the imperfective paradox and thus easily allow for suspension of the culmination (Dowty, 1979, i.a.). Bott and Hamm reported enhanced processing difficulty in the English coercion condition (3a) relative to the control condition (3b) and the progressive control condition (3c), while the two conditions didn't differ in the German experiments. They proposed the Crosslinguistic Aspectual Variation (CAV) hypothesis, stating that English simple past accomplishments are pragmatically strengthened to a telic interpretation resulting in coercion costs, whereas German accomplishments are underspecified with respect to telicity and are thus smoothly adjusted to the sentence or discourse context.

The present study aims at replicating and extending the work just reviewed. We targeted the Bott and Hamm (2014) study for a number of reasons. Firstly, their cross-linguistic comparison involved different accomplishment verbs and sentence materials between the German and English experiments, making a direct comparison impossible. Secondly, even for English, coercion difficulty was only observed in constructions (3a)–(3c), but not in another set of conditions testing the stimuli in a slightly different word order. Thirdly, the study was underpowered (Brysbaert & Stevens, 2018) with only 30 participants contributing only 8 observations in each condition. And last but not least, no offline interpretation results were provided for the English part of the study, calling into question whether the hypothesized coercion operation was necessary at all.

The Present Study

The present study adopted the materials of the English experiment in Bott and Hamm (2014). The 48 items were translated to German yielding parallel sets of experiments (see the accompanying OSF archive). The experiments can be summarized as follows (E: English, G: German; Offline Inf.: Offline inference rating task, SPR: self-paced reading):

Exp. 1: Offline Inf. (G)	Exp. 2: Offline Inf. (E)
Exp. 3: SPR sentence (G)	Exp. 4: SPR sentence (E)
Exp. 5: SPR discourse (G)	Exp. 6: SPR discourse (E)
Exp. 7: Stops-making-sense task (E)	

The first two experiments employed an inference rating task in order to establish that coercion sentences such as (3a) and (4a) do not implicate the culmination, while baseline comparisons without adverbials in fact do so and are perceived to be as telic as the control conditions with *in*-adverbials in (3b) and (4b). These and the following experiments were exactly parallel in the two languages. Exps. 3 and 4 were self-paced reading experiments with no additional task testing processing difficulty during the real-time interpretation of these constructions. Exps. 5 and 6 modified the experimental items by separating the accomplishment and the culmination canceling expression across a sentence boundary. Since in none of these self-paced reading experiments a coercion effect was found, Exp. 7 employed a stops-making-sense task in a parallel experiment to Exp. 4 in order to see whether non-culminating accomplishments give rise to processing difficulty in a task calling for deep semantic interpretation and immediate full specification of the incoming sentence.

The study was first preregistered at *aspredicted.org* for Exps. 1–4 (link). After the analysis of these experiments, we decided to follow them up with Exps. 5–7 in two separate preregistrations, for Exp. 5/6 (link) and Exp. 7 (link).

Offline Interpretation – Exps. 1/2

The first two experiments assessed the interpretation of the English and German accomplishment items. These encompassed incremental theme verbs commonly viewed as telic with quantized objects like *build the monument* or *write the letter* (Krifka, 1998; Rappaport Hovav, 2008). Many of them encoded endpoints lexically by means of a particle/prefix (*eat <u>up</u>*, *pour <u>down</u>*, *<u>overfly</u>*), which trigger an indefeasible culmination entailment according to Martin (2019). The interpretation of imperfective and perfective accomplishments with respect to telicity is an open question. Thus, Baggio and van Lambalgen (2007); Baggio et al. (2008) claimed that past progressive accomplishments trigger a culmination inference per default, while Minor et al. (2023) proposed that they refer to ongoing, incomplete events. A similar picture emerges for the English perfective. While Bott and Hamm (2014) assumed it to involve event completion, Minor et al. (2023) proposed that simple past English accomplishments are ambiguous between complete and incomplete event interpretations.

Methods

Procedure The experiments employed an inference rating task in which participants were presented with the target sentences (e.g., *The architect built the monument*) along with a

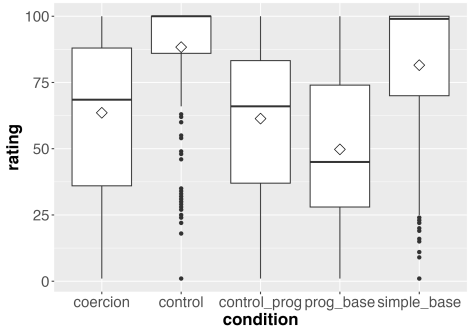

Figure 1: Inference ratings (means (◇) + boxplots) for English accomplishments in Exp. 1; *base*: baseline, *prog*: progressive (imperfective), *simple*: simple past (perfective).

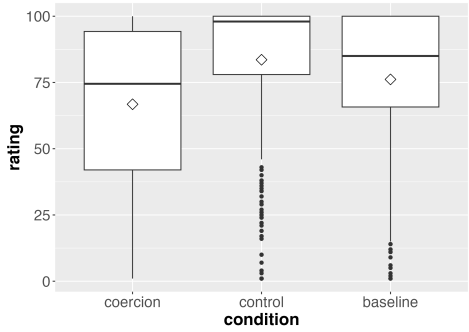

Figure 2: Inference ratings (means (◇) + boxplots) for German accomplishments in Exp. 2.

potential inference (e.g., *From this I conclude that the monument was completed*), for which they had to provide a rating with a slider with values ranging from 1 *no, definitely not* to 100 *yes, definitely*. For English, the three sentence conditions (3a), (3b) and (3c) were tested together with two baseline conditions without adverbials: a baseline condition in the simple past and another baseline condition in the past progressive. For the experimental items we always queried the culmination inference by asking whether the sentence implies that the incremental theme argument was complete. For German, the two conditions (4a) and (4b) were tested together with a baseline condition without an adverbial. The experiment was implemented with the Penn Controller for Ibex (Zehr & Schwarz, 2018) and conducted online. After instructions and an exercise with explicit feedback, the experiment followed in a single block without any feedback and no time pressure. Median experiment duration was approx. 15 minutes.

Materials The English items in Exp. 1 were constructed in five conditions, three with adverbials and two baseline conditions without adverbials. The German items in Exp. 2 were constructed in three conditions, two conditions with adverbials and a baseline condition without adverbial. Items were distributed to experimental lists in a Latin Square design.

32 fillers were added to each list: 12 fillers tested valid vs. invalid entailments, 5 fillers tested the presuppositions of presupposition triggers and 15 fillers tested scalar implicatures ranging from strong to invalid inferences (Ronai & Xiang, 2022). Thus, the experiment importantly included invalid inferences, that is, test cases contradicting an actual inference.

Participants 31 native English speakers (8 fem., 20 male, 2 non-binary, 1 unknown; mean age 23.5 y.o., range 19-34) and 30 native German speakers (17 fem., 12 male, 1 non-binary; mean age 22.5 y.o., range 19-31) were recruited from Prolific for a payment of £3.50.

Results and Discussion

Figures 1 and 2 show mean inference ratings and box plots for the experimental conditions of Exp. 1 and 2. The ratings of the fillers are omitted from the plots, but these, importantly, covered the whole range of the scale (cf. analyses on OSF).

The ratings of the **English** baseline conditions show that unmodified accomplishments in the simple past receive a complete event interpretation, contra Minor et al. (2023). The progressive baseline condition, on the other hand, is completely open to whether the event culminates or doesn't culminate, with ratings well in the middle of the scale, contra Baggio and van Lambalgen (2007); Baggio et al. (2008). These differences in results is most likely due to the different methods used across studies, that is, a choice between preparatory process and result state interpretations in Minor et al. and a binary probe task in Baggio et al.. The latter employed Dutch present perfect constructions with the perfect being known to display a high degree of semantic ambiguity in itself (Alexiadou, Rathert, & von Stechow, 2003).

Turning to the modified conditions, the coercion condition considerably weakened this inference, whereas the control condition with *in*-adverbials were clearly interpreted as telic. A linear mixed-effects regression (LMER) analysis with random intercepts and by-condition slopes of participants and items was conducted comparing the COERCION condition with the CONTROL condition (using R's *lme4* package, Bates, Mächler, Bolker, and Walker (2015)). This difference was significant ($\chi^2(1) = 33.76, p < .001$), showing that the two adverbials led to clearly different interpretations. This effect did not vary between items (non-sign. CONDITION by-item slope $\chi^2(2) = 2.51, p = .29$). Furthermore, a comparison between the CONTROL condition and the SIMPLE BASELINE condition revealed no reliable difference between these conditions ($\chi^2(1) = -1.00, p = .33$). The PROGRESSIVE CONTROL condition, on the other hand, was indistinguishable from the COERCION condition ($\chi^2(1) = 0.99, p = .32$) showing that these two conditions in fact gave rise to similar interpretations as assumed in Bott and Hamm (2014).

In **German** (Exp. 2), there was also a clear difference between COERCION and CONTROL conditions ($\chi^2(1) = 30.83, p < .001$), although somewhat smaller, as reflected in a significant interaction between CONDITION and LANGUAGE in a linear mixed-effects regression analysis compar-

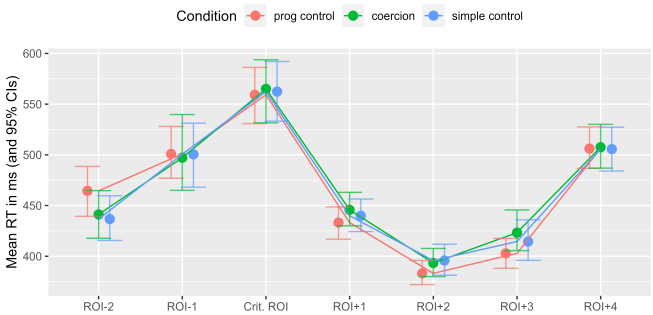

Figure 3: Mean reading times and bootstrapped confidence intervals in Exp. 3 testing coercion within English sentences.

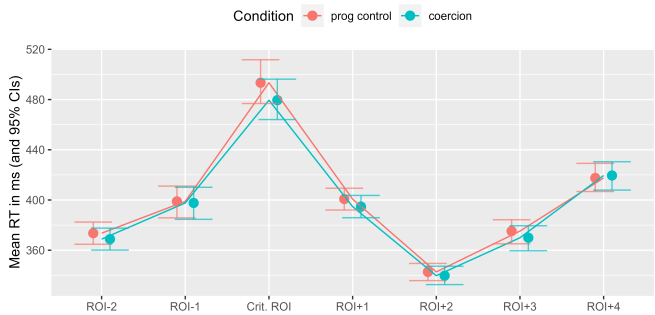

Figure 4: Mean reading times and bootstrapped confidence intervals in Exp. 4 testing coercion within German sentences.

ing COERCION vs. CONTROL across languages ($\chi^2(1) = 4.63, p < .05$). Like English, the German items did not vary significantly in terms of the size of the coercion effect, as indicated by an analysis of by-items slopes of CONDITION ($\chi^2(2) = 0.83, p = .66$). Diverging from English, the German BASELINE condition differed reliably from the CONTROL ($\chi^2(1) = 10.63, p < .01$) and the COERCION condition ($\chi^2(1) = 11.90, p < .01$). This suggests that German accomplishments in unmodified past sentences received telic interpretations overall, but were sometimes also interpreted as non-telic, an interpretation not readily available in English.

Self-Paced Reading – Exps. 3/4

The sentences pretested in the inference task were used in a self-paced reading experiment without a secondary task beyond simple comprehension questions. Exp. 3 tested the three English conditions in (3a), (3b) and (3c); Exp. 4 tested the two German conditions (4a) and (4b). With more than 1,000 observations per condition, the present experiments had considerably more statistical power than Bott and Hamm (2014)'s experiment on English with 240 observations per condition.

Methods

Materials The 48 items from the previous experiments were segmented into regions of interest (ROI) for moving window presentation with the following ROIs: ROI-1 subject phrase, ROI-2 verb, ROI-3 object phrase, ROI-4 critical adverbial region. In addition, two spillover regions, ROIs 5 and 6, were added as part of a subsequent *after*/*nachdem* sentence, specifying a temporally preceding event that did not interfere with the presence/absence of the culmination inference. ROI-5 was always the connective *after* and ROI-6 the following subject phrase (the remainder of the *after* sentence was segmented into two additional regions). The adverbial regions (ROI-4) were controlled for length with no sign. difference in the number of characters between *for*- and *in*-adverbials (English: $t(47) = .49, p = .63$; German: $t(47) = -.29, p = .78$). The items were distributed to lists according to a Latin Square design. To each list, the same set of 45 fillers was added, none of which involved aspectual coercion. Simple comprehension questions were formulated together with two answer alternatives for 16 experimental items and 15 fillers.

Procedure The experiment employed phrase-by-phrase self-paced reading with a moving window presentation. It was implemented in E-Prime 3.0 and conducted online via E-Prime Go 1.0. A typical experimental session had a median time on task of approximately 20 minutes.

Participants 60 native speakers were recruited from Prolific (link). After exclusion of participants with error rates above 15% on comprehension questions, 51 participants were included in the analysis of Exp. 3 (24 fem., 24 male, 3 other; mean age 26.6 y.o., range 18–40) and 50 participants were included in the analysis of Exp. 4 (21 fem., 29 male; mean age 24.7 y.o., range 18–34). Each participant was paid £5.

Data Analysis Reading times were first cleaned by eliminating outlier RT below 100ms and values 2.5 SDs above a participants mean for each ROI, leading to the exclusion of less than 4% of the data in each ROI/experiment.

The remaining RTs were subjected to LMER analyses with by-participants and by-items random intercepts and CONDITION slopes. The three levels of the fixed effect of CONDITION were treatment-coded into two contrasts, COERCION vs. CONTROL and COERCION vs. PROGRESSIVE CONTROL (the latter only for English), with COERCION as base category. Since the experiments involved multiple comparisons with analyses for three ROIs, α levels were Bonferroni-corrected to critical p-values of $.01\overline{6}$. Bonferroni corrected confidence intervals were fitted using the bootstrap function of R's bootstrap package (Efron & Tibshirani, 1993).

Results and Discussion

The mean reading times of Exp. 3 and 4 are shown in Figures 3 and 4, respectively. The only ROI where conditions differed was the verb ROI (ROI-2) of Exp. 3, which is expected due to length differences between past progressive and simple past forms. The LMER analyses corroborated this general lack of coercion effects. Neither the fixed effects of CONDITION nor the by-item random slopes of CONDITION contributed significantly to model fit (see analyses in OSF archive).

In the English experiment the fixed effect of condition including the contrasts COERCION vs. CONTROL and COERCION vs. PROGRESSIVE CONTROL was far from sig-

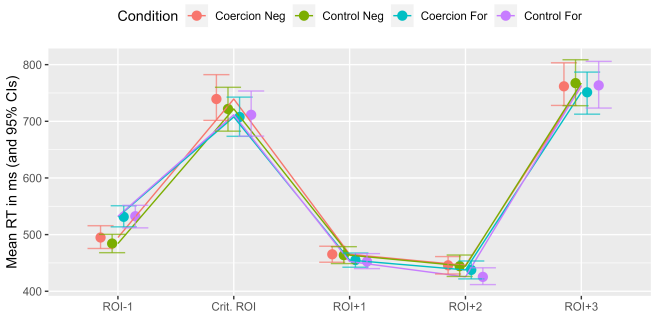

Figure 5: Mean reading times in Exp. 5 (English).

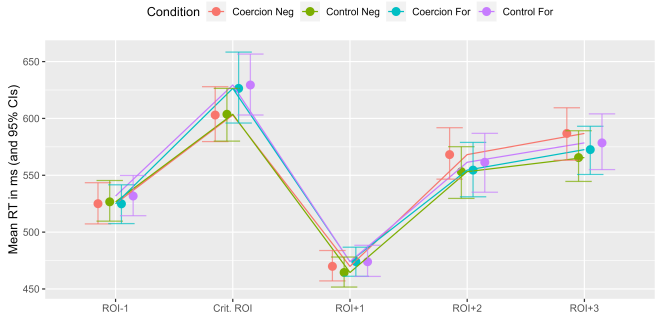

Figure 6: Mean reading times in Exp. 6 (German).

nificant for the critical adverbial ROI ($\chi^2(2) = 0.67, p = .72$) and for the two following spillover ROIs (ROI-5: $\chi^2(2) = 1.82, p = .40$; ROI-6: $\chi^2(2) = 3.05, p = .22$). Thus, the coercion condition was read as fast as the telic accomplishments in the simple past and progressive control conditions.

The qualitatively same pattern of effects was observed for German. The fixed effect of CONDITION did not significantly enhance model fit for the LMER analysis modeling the reading times of the adverbial ROI ($\chi^2(1) = 4.18, p = .04$) nor for the analyses of the two spillover ROIs (ROI-5: $\chi^2(1) = 1.15, p = .28$; ROI-6: $\chi^2(1) = 0.31, p = .58$).

In sum, the first set of self-paced reading experiments failed to replicate the processing differences for non-culminating accomplishments in English and German found by Bott and Hamm (2014). As the present study improved on that study in many respects, crucially also statistical power, we propose that subtractive coercion of accomplishments into activity readings is not taxing, neither in a language largely lacking grammatical aspect like German, nor in a language with grammaticalized aspectual distinctions like English.

Self-Paced Reading in Discourse – Exps. 5/6

To investigate the effects of canceling a culmination inference in subsequent discourse, we conducted another set of cross-linguistic reading time experiments. (5) illustrates the 2×2 design (vertical lines indicate segmentation). Whether culmination occurred was manipulated in the context sentence either using a simple past accomplishment (coercion condition) or the aspectual verb *begin* (control condition). The aspectual verb *begin* is a (lexical) aspectual operator particularly well-suited for our cross-linguistic investigation since it can be equally applied to English and German accomplishments. It selects for the initial part of the eventuality denoted by its complement (Egg, 2003), which for accomplishments is their preparatory process. Moreover, it could even trigger a *Non-Actuality Implicature* (Grant, Clifton, & Frazier, 2012) regarding the culmination of its complement eventuality.

(5) a. The architect | built | the monument | after | the city | had finally | provided | the money for it.
b. The architect | began building | the monument | after ...

The context sentence was immediately followed by the target sentence explicitly negating the culmination (NEGATION) or implicitly cancelling it using a *for*-type adverbial (FOR).

(6) a. He could | not complete it, however, | as | he had | a severe accident.
b. He did this | for more than two years, | until | ...

Methods

Materials, Procedure & Data Analysis The materials from the previous experiments were adopted to the design in (5) and (6). The target sentences were controlled for length making sure that there were no sign. differences between explicit negation and *for* targets. The context and the target sentence were presented as one text on a screen. The four resulting discourse conditions were distributed to four lists and 45 filler discourses were added to each list. The procedure was the same as in Exps. 3 and 4. A typical experimental session took about 25 minutes. The reading time data were analyzed in parallel fashion to the previous experiments.

Participants 90 native speakers of English and 90 native speakers of German were recruited from Prolific. Excluding participants who answered less than 80% of comprehension questions correctly, 82 participants were included into the analysis of the English experiment (43 fem., 34 male, 3 other; mean age 27.1 y.o., range 19–40) and 81 participants into the analysis of the German one (42 fem., 37 male, 2 other; mean age 27.5 y.o., range 19–40). Participants were paid £5.

Results and Discussion

As in the previous experiment, the inferential statistics did not provide evidence for coercion effects in English or German. We compared models including the maximal fixed effects structure with models including no fixed effects involving COERCION. These analyses revealed no reliable effects of coercion in English (crit. ROI: $\chi^2(3) = 1.69, p = .64$; 1st spillover ROI: $\chi^2(3) = 5.24, p = .16$; 2nd spillover ROI: $\chi^2(3) = 8.39, p = .04$) nor in German (crit. ROI: $\chi^2(3) = 4.11, p = .25$; 1st spillover ROI: $\chi^2(3) = 2.94, p = .40$; 2nd spillover ROI: $\chi^2(3) = 2.73, p = .44$). We also found there to be no significant by-items slopes due to COERCION, suggesting that the items behaved uniformly with respect to this

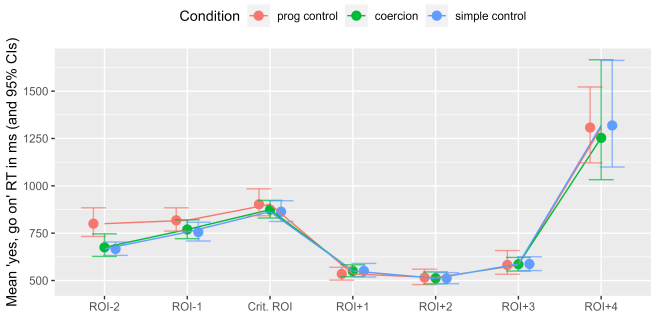

Figure 7: Mean RTs of *yes, go on* button presses in Exp. 7.

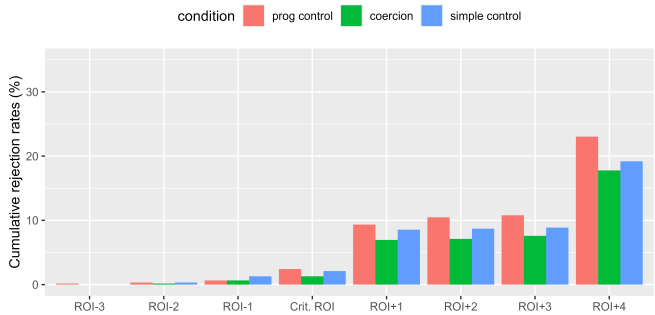

Figure 8: Mean cumulative rejection rates per ROI in Exp. 7.

general absence of effects. In sum, this strongly suggests that non-culminating accomplishments did not lead to measurable disruption during reading even if the culmination was canceled in subsequent discourse. This finding clearly stands in opposition to the conclusions of Baggio et al. (2008)'s study.

Stops-Making Sense Task – Exp. 7 (English)

The discussion in Pickering et al. (2006) vs. Todorova et al. (2000, i.a.) suggests that coercion effects may show up only in tasks involving a secondary task explicitly reflecting upon the meaning (see Bott (2010) for a similar point w.r.t. non-culminating accomplishments). More generally, this points to the role of underspecification in sentence comprehension (Sanford & Sturt, 2002; Ferreira & Patson, 2007) and in semantic underspecification accounts of coercion phenomena (Dölling, 2014; Egg, 2005, 2010). To test whether underspecification is an issue for English non-culminating accomplishments, our final experiment employed the stops-making sense task employed by Todorova et al. for iterative coercion.

Methods

Materials The experimental items were identical to those of Exp. 3. 77 new fillers were added to each experimental list: 39 non-sensical ones and 38 sensible ones. The nonsense fillers involved tense and aspectual mismatches (e.g., *he was knowing the answer*). Assuming that the coercion condition is in fact fully acceptable, about one third of the trials across the experiment were thus non-sensible and should be rejected.

Procedure and Analysis Participants read sentences self-paced in a moving window presentation. For each ROI they had to decide whether to read on or to reject the sentence as nonsensical, aborting a trial. For each ROI, we analyzed rejection rates using logistic mixed-regression modeling and the RTs of *yes, go on* button presses using linear mixed-regression models as in the previous experiments. A typical experimental session lasted about 35 minutes.

Participants 54 native English speakers were recruited via Prolific. 15 participants were excluded from the analysis due to poor performance on filler trials . Consequently, the analysis included 39 participants (23 fem., 16 male; mean age 27.8 y.o., range 18–40). They were paid £6.

Results and Discussion

Global rejection rates for nonsense fillers were 48.9% as compared to 13.7% for sensible fillers. The three experimental conditions broadly patterned with sensible fillers with rejection rates of 21.6% in coercion conditions, 17.8% in simple control and 18.9% in progressive control conditions, respectively (no sign. difference between conditions in GLMER: $\chi^2(2) = 3.58, p = .17$). The GLMER analyses of the critical ROI and the two spillover ROIs furthermore revealed no significant differences between the rejection rates mid-sentence (crit. ROI: $\chi^2(2) = 4.07, p = .13$; 1st spillover ROI: $\chi^2(2) = .97, p = .62$). The lack of effect was further corroborated by the RT analyses (crit. ROI: $\chi^2(2) = .91, p = .63$; 1st spillover ROI: $\chi^2(2) = .64, p = .73$). Thus, even when continuously having to consider the sensicality of the evolving sentence, coercion went unnoticed (but see Todorova et al. (2000) for iterative coercion effects in the same task).

General Discussion

We presented the results of seven experiments on the interpretation and processing of non-culminating accomplishments in English and German. The first two self-paced reading experiments provided evidence that modification by a *for*-adverbial does not lead to disruptions in reading pace, neither in German nor in English. This finding is inconsistent with Bott and Hamm's *Crosslinguistic Aspectual Variation* hypothesis. This general lack of coercion effect generalized to another pair of reading time experiments where the culmination inference was explicitly or implicitly defeated in subsequent discourse context. This finding challenges the claims in Baggio and van Lambalgen (2007); Baggio et al. (2008). The final experiment showed that even in a task requiring permanently checking the unfolding compositional interpretation, non-culminating accomplishments go unnoticed. In sum, the study shows that canceling or weakening the culmination inference does not pose a problem for the semantic/pragmatic parser, which flexibly adjusts the aspectual meaning to the larger discourse context. The results provide strong support for the flexibility of the interpretation process, as our inference task and the explicit and implicit cancelation of the culmination in our reading tasks constitute the most direct experimental tests of non-culminating accomplishments so far.

Acknowledgments

This research has been funded by the German Science Foundation (DFG) - project B01 *Coercion as a Creative Mechanism in Compositional Interpretation*, CRC 1646 *Linguistic Creativity in Communication*. We would like to thank Fritz Hamm, Fabienne Martin and Hana Filip for discussion and four anonymous reviewers for CogSci 2024 for their valuable feedback. Furthermore we would like to thank the audience at the Semantics Colloquium at the Heinrich Heine University, Düsseldorf and at the TELIC workshop, 2017 at the University of Stuttgart. Last but not least, we would like to thank Matthias Schrumpf for assisting with the experiments as well as the participants in O.B.'s Bachelor seminar *Experiments on Tense and Aspect* in the summer semester 2023 at Bielefeld University. The usual disclaimer applies.

References

Alexiadou, A., Rathert, M., & von Stechow, A. (Eds.). (2003). *Perfect explorations*. Berlin, New York: Mouton de Gruyter.

Arunchalam, S., & Kothari, A. (2011). An experimental study of Hindi and English perfective interpretation. In *JSAL* (Vol. 4, pp. 27–42). CSLI publications.

Baggio, G., & van Lambalgen, M. (2007). The processing consequences of the imperfective paradox. *Journal of Semantics*, *24*(4), 307–330.

Baggio, G., van Lambalgen, M., & Hagoort, P. (2008). Computing and recomputing discourse models: an ERP study. *Journal of Memory and Language*, *59*(1), 36–53.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

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

Bott, O., & Hamm, F. (2014). Cross-linguistic variation in the processing of aspect. In B. Hemforth, B. Schmiedtová, & C. Fabricius-Hansen (Eds.), *Psycholinguistic approaches to meaning and understanding across languages* (p. 83-109). New York: Springer (Studies in Theoretical Psycholinguistics 44).

Brennan, J., & Pylkkänen, L. (2008). Processing events: behavioral and neuromagnetic correlates of aspectual coercion. *Brain and Language*, *106*, 132–143. doi: 10.1016/j.bandl.2008.04.003

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of cognition*, *1*, 1–20. doi: 10.5334/joc.10

Comrie, B. (1977). In defense of spontaneous demotion: The impersonal passive. In P. Cole & J. M. Sadock (Eds.), *Syntax and semantics: Grammatical relations* (Vol. 8, p. 47-58). New York: Academic.

Dölling, J. (2014). Aspectual coercion and eventuality structure. In K. Robering (Ed.), *Events, Arguments, and Aspects: Topics in the Semantics of Verbs* (pp. 189–226). Amsterdam: John Benjamins. doi: 10.1075/slcs.152.05dol

Dowty, D. R. (1979). *Word Meaning and Montague Grammar*. Dordrecht: D. Reidel.

Efron, B., & Tibshirani, R. (1993). *Introduction to the bootstrap*. New York, London: Chapman and Hall.

Egg, M. (2003). Beginning Novels and Finishing Hamburgers: Remarks on the Semantics of *to begin*. *Journal of Semantics*, *20*, 163-191.

Egg, M. (2005). *Flexible semantics for reinterpretation phenomena*. Stanford: CSLI.

Egg, M. (2010). Semantic underspecification. *Language and Linguistics Compass*, *4*(3), 166–181.

Ferreira, F., & Patson, N. D. (2007). The 'good enough' approach to language comprehension. *Language and Linguistics Compass*, *1*(1-2), 71–83.

Grant, E., Clifton, C. J., & Frazier, L. (2012). The role of non-actuality implicatures in processing elided constituents. *Journal of Memory and Language*, *66*(1), 326–343.

Kasher, N., & Hacohen, A. (2023). Aspect processing across languages: A visual world eye-tracking study. *Glossa: a journal of general linguistics*, *8*(1), 1–14. doi: 10.16995/glossa.9680

Koenig, J.-P., & Muansuwan, N. (2000). How to end without ever finishing: Thai semi-perfectivity. *Journal of semantics*, *17*(2), 147–184.

Krifka, M. (1998). The Origins of Telicity. In S. Rothstein (Ed.), *Events and Grammar.* Dordrecht: Kluwer.

Martin, F. (2019). Non-culminating accomplishments. *Language and Linguistics Compass*, *13*, e1234: 1–20.

Minor, S., Mitrofanova, N., Guajardo, G., Vos, M., & Ramchand, G. (2023). Aspect processing across languages: A visual world eye-tracking study. *Frontiers in Language Sciences*, *1*. doi: 10.3389/flang.2022.1052205

Moens, M., & Steedman, M. (1988). Temporal ontology and temporal reference. *Computational Linguistics*, *14*(2), 15–28.

Nadathur, P., & Filip, H. (2021). Telicity, teleological modality, and (non-)culmination. In *Proceedings of the 39th West Coast Conference on Formal Linguistics* (p. preprint). University of Arizona.

Paczynski, M., Jackendoff, R., & Kuperberg, G. (2014, 04). When events change their nature: The neurocognitive mechanisms underlying aspectual coercion. *Journal of cognitive neuroscience*, *26*, 1905–1917. doi: 10.1162/jocn_a_00638

Pickering, M. J., McElree, B., Frisson, S., Chen, L., & Traxler, M. J. (2006). Underspecification and aspectual coercion. *Discourse Processes*, *42*(2), 131–155.

Rappaport Hovav, M. (2008). Lexicalized meaning and the internal temporal structure of events. In S. Rothstein (Ed.), *Theoretical and crosslinguistic approaches to the semantics of aspect* (pp. 13–42). Amsterdam: John Benjamins. doi: 10.1075/la.110.03hov

Ronai, E., & Xiang, M. (2022). Three factors in explaining scalar diversity. In D. Gutzmann & S. Repp (Eds.), *Proceedings of Sinn & Bedeutung 26* (pp. 716–733).

Sanford, A. J., & Sturt, P. (2002). Depth of processing in language comprehension: not noticing the evidence. *Trends in Cognitive Sciences*, *6*(9), 382–386.

Todorova, M., Straub, K., Badecker, W., & Frank, R. (2000). Aspectual coercion and the online computation of sentential aspect. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 22, pp. 1–6). Cognitive Science Society. Retrieved from https://escholarship.org/uc/item/14x15786

van Lambalgen, M., & Hamm, F. (2005). *The proper treatment of events*. Malden, Mass.: Blackwell.

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

Zehr, J., & Schwarz, F. (2018). *PennController for Internet Based Experiments (IBEX).* doi: 10.17605/OSF.IO/MD832