Proceedings of SALT 35: 144–163, 2025

On aspectual coercion in *before*-clauses: Evidence from processing*

Johanna Alstott
Massachusetts Institute of Technology

Abstract Sentences with accomplishments in *before*-clauses are *prima facie* ambiguous between a strong reading and a weak reading, and two types of theories of these sentences have been proposed. Under-specification theories posit that the *before*-clauses in question are not actually ambiguous: their only LF corresponds to the weak reading, with the strong reading falling out as a subcase. Rett's (2020) ambiguity theory, by contrast, claims that the strong reading surfaces by default, with the weak reading only arising if we coerce the embedded accomplishment into an achievement. Under standard psycholinguistic assumptions, coercion incurs a processing cost, and hence, I argue, Rett's (2020) theory makes processing predictions that under-specification theories do not. I test these predictions via two self-paced reading experiments and find moderate evidence that they are borne out.

Keywords: coercion, temporal connectives, lexical aspect, language processing

1 Introduction

As originally observed by Heinamäki (1974), sentences with accomplishments in *before*-clauses, such as (1a), are *prima facie* ambiguous (i.e. they can be judged true in two types of scenarios), unlike parallel sentences with *after*.

(1) a. Ben was nervous before Eli climbed the mountain.
b. Ben was nervous after Eli climbed the mountain.

To see this difference between (1a) and (1b), suppose that Eli climbed the mountain in three hours (6pm-9pm) and consider three scenarios for when Ben was nervous (Scenario A, 4-5pm; Scenario B, 7-8pm; Scenario C, 10-11pm). (1a) is clearly true if Ben was nervous before Eli started his climb (Scenario A), but it can also be judged true–with a bit more effort–if he was nervous before Eli finished (Scen. B). Conversely, (1b) is only true if Ben was nervous after Eli finished his climb (Scen. C); if he was merely nervous after Eli started (Scen. B), (1b) is false.

* For their valuable insights and feedback, I thank Athulya Aravind, Martin Hackl, Shota Momma, Kai von Fintel, Sabine Iatridou, and audiences at MIT LingLunch, CreteLing 2024, and SALT 35.

There are two types of theories of why (1a) has both a "before-start" reading (Scenario A) and a "before-finish" reading (Scenario B): under-specification theories and Rett's (2020) ambiguity theory. Under-specification theories (Heinamäki 1974; Beaver & Condoravdi 2003; Condoravdi 2010;[1] Krifka 2010), noting that there is an entailment relation between the two putative "readings" of (1a), posit that (1a) is not really ambiguous. On this view, (1a)'s sole LF has a before-finish meaning ("Ben was nervous before Eli finished climbing the mountain"). (1a) is true in a before-start context, on this view, because this context is a subcase of a before-finish context: if Ben was nervous before Eli started, then he was nervous before Eli finished.

Against under-specification theories, Rett (2020) proposes that in the absence of silent material, (1a) has a before-start meaning ("Ben was nervous before Eli started climbing the mountain"). To account for why (1a) can be judged true in a before-finish scenario like Scenario B, Rett (2020) assumes that a silent **completive coercion** operator (which I call COMPLET) can attach to an accomplishment like *Eli climbed the mountain* and make it get interpreted as *Eli finished climbing the mountain*. In other words, Rett (2020) claims that *before Eli climbed the mountain* (without silent material) means "before Eli started climbing the mountain," whereas *before* COMPLET *Eli climbed the mountain* means "before Eli finished climbing the mountain." To account for why (1b) is not ambiguous in the same way as (1a), Rett (2020) claims that inserting COMPLET in an *after*-clause would be vacuous.

Rett (2020) argues that positing COMPLET is not *ad hoc* but rather independently needed to capture cases where an *at*-adverbial modifies an accomplishment, such as (2). In particular, Rett (2020) notes that (2) has a reading paraphraseable as "Eli finished climbing the mountain at 7pm sharp," and she accounts for this reading by assuming that the *at*-adverbial forces COMPLET to apply to *Eli climbed the mountain*.

(2) Eli climbed the mountain at 7pm sharp.

The debate between under-specification theories and Rett's (2020) theory is unsettled, and the present study contributes to the debate by identifying and testing two online processing predictions that Rett's (2020) theory makes and under-specification theories do not. In a nutshell, Rett's (2020) theory unique processing predictions because coercion has a distinct psycholinguistic profile (Todorova, Straub, Badecker & Frank 2000; Husband, Beretta & Stockall 2006; Brennan & Pylkkänen 2008, 2010; Lukassek, Prysłopska, Hörnig & Maienborn 2017). In particular, under standard assumptions, parsers do not insert coercion operators by default and thus show processing slowdowns in a coercion sentence once they know that coercion is needed

1 Condoravdi (2010) posits that **atelic** *before/after* clauses are ambiguous between an LF with a silent *earliest* operator and one with a silent *max*. Nonetheless, her theory of telic cases like (1a) is an under-specification theory: for her, (1a) has just one LF, the one with *earliest*. LFs with *max* are, she claims, ill-formed when the temporal clause is telic (or non-cumulative more broadly, see p. 902).

to repair a mismatch with an adverbial or with the overall context. Using this perspective on coercion, past studies have tested coercion vs. non-coercion theories of various data, arguing for coercion if the claimed coercion cases have an online cost (Piñango, Zurif & Jackendoff 1999; Todorova et al. 2000; McElree, Traxler, Pickering, Seely & Jackendoff 2001; Traxler, Pickering & McElree 2002; Brennan & Pylkkänen 2008, 2010; Lukassek et al. 2017, a.o.). However, no prior work examines Rett's (2020) notion of completive coercion under the lens of online processing.

Once we take past work on the processing of coercion into account, we see that if Rett (2020) is correct, (1a)'s before-finish reading and (2) should incur the same sort of processing cost as other coercion phenomena. Absent additional assumptions, under-specification theories do not predict that the processing of (1a) should be affected by context (so long as the sentence is true) and are agnostic about (2).

I test these predictions via two self-paced reading experiments. In Experiment 1, I test whether cases like (2) have the processing profile predicted by the completive coercion account, and I find that they do. In Experiment 2, I test whether the before-finish reading of sentences like (1a) has the processing profile predicted by Rett's (2020) theory; I find that before-finish readings do incur a processing cost, though the online effect found in Experiment 2 differs from the effect found in Experiment 1 in subtle ways. Taken together, my results are broadly compatible with Rett's (2020) theory, though the subtle processing differences between (2) and the before-finish reading of (1) require Rett (2020) to make some additional assumptions.

The rest of this paper is structured as follows. Section 2 provides further background on theories of *before/after* (section 2.1) and aspectual coercion (section 2.2). Section 3 lays out the processing predictions made by Rett's (2020) theory, while sections 4-5 report on Experiments 1-2. Section 6 sums up the results and concludes.

2 Background

2.1 Accomplishments in *before*-clauses: Under-specification or ambiguity?

This subsection expounds an example of an under-specification theory and contrasts it with Rett 2020. I illustrate the theories using (1), ignoring tense and grammatical aspect so as to stay faithful to the theories I am summarizing (which do not consider these features). Throughout, I make three formal assumptions: (a) propositions are sets of time-intervals, type $\langle i,t \rangle$; (b) a stative like ⟦Ben was nervous⟧ contains the maximal interval throughout which Ben was nervous and its subintervals; (c) accomplishments like ⟦Eli climbed the mountain⟧ are singletons. For example, if Ben was nervous 4-5pm, then ⟦Ben was nervous⟧ contains [4pm,5pm] and its subintervals. And if Eli climbed the mountain 6pm-9pm, ⟦Eli climbed the mountain⟧={[6pm,9pm]}. These claims about *Aktionsart* (Dowty 1979) are standard.

2.1.1 An under-specification theory of (1)

As an exemplar of an under-specification theory, I present a version of Anscombe (1964) and Krifka's (2010) quantificational approach to *before/after*, minimally modified so as to be compatible with the standard semantics for accomplishments mentioned in the previous paragraph.[2] On this view, (1a) receives the LF in (3a), the derivation for which involves the lexical entry for *before* given in (3b). Note that I use "$t'' \leq_\forall t'$" as shorthand for "every subinterval of t'' precedes or overlaps t'."

(3) a. [[Ben was nervous][before [Eli climbed the mountain]]]
b. $[\![\text{before}]\!] = \lambda q_{\langle i,t \rangle}.\ \lambda p_{\langle i,t \rangle}.\ \exists t'\ [p(t') = 1 \wedge \neg\exists t''\ [q(t'') = 1 \wedge t'' \leq_\forall t']]$

This entry for *before* yields the following truth-conditions for (1a).

(4) $[\![(1a)]\!] = 1$ iff $\exists t'\ [t' \in [\![\text{Ben was nervous}]\!] \wedge$
$\neg\exists t''\ [t'' \in [\![\text{Eli climbed the mountain}]\!] \wedge t'' \leq_\forall t']]$

To see that these truth-conditions are correct, let us suppose again that $[\![\text{Eli climbed the mountain}]\!] = \{[6\text{pm},9\text{pm}]\}$. In this scenario, (4) can be restated as: "there is a time-point t' such that (a) Ben was nervous at t'; (b) not all subintervals of [6pm,9pm] precede or overlap t'." If not all subintervals of [6pm,9pm] precede or overlap t', then at least part of the runtime of Eli climbing the mountain must come after t'. In other words, for the conditions in (4) to be satisfied, Ben must have been anxious sometime before Eli finished climbing the mountain, as desired.[3]

Turning to *after*, Anscombe (1964) posits the lexical entry in (5a), which, when plugged into (1b), yields the truth-conditions in (5b).

(5) a. $[\![\text{after}]\!] = \lambda q_{\langle i,t \rangle}.\ \lambda p_{\langle i,t \rangle}.\ \exists t'\ [p(t') = 1 \wedge \exists t''\ [q(t'') = 1 \wedge t'' < t']]$
b. $[\![(1b)]\!] = 1$ iff $\exists t'\ [t' \in [\![\text{Ben was nervous}]\!] \wedge$
$\exists t''\ [t'' \in [\![\text{Eli climbed the mountain}]\!] \wedge t'' < t']]$

2 To elaborate: Anscombe's (1964) theory (built upon in Krifka 2010) is an influential view that uses quantificational entries for *before/after* and seeks to avoid positing that *before/after*-sentences are ambiguous. Anscombe (1964) herself only considered atelic examples, but Beaver & Condoravdi (2003) note that her lexical entries work for examples like (1) too if we make a non-standard assumption about accomplishments: namely, that $[\![\text{Eli climbed the mountain}]\!]$ contains the telos of the construction rather than the runtime. While it is possible to reconcile Anscombe's (1964) theory with (1) in this way, it is equally simple (and still the spirit of the theory) to reconcile the two by slightly tweaking Anscombe's entries, and hence I choose that option here. See fn. 3 for further discussion.

3 Following up on fn. 2: the original Anscombe-Krifka entry for *before* uses $t'' \leq t'$ ("t'' precedes or overlaps t'") instead of $t'' \leq_\forall t'$. If we replaced $\leq_\forall$ with $\leq$, (4) would require that **no** subinterval of [6pm,9pm] precede or overlap t', i.e. require that t' precede [6pm,9pm]. These truth-conditions are only verified in a "before-start" scenario and are hence too strong. To remedy this state of affairs, we can either use $\leq_\forall$ instead of $\leq$ (as done here), or we can keep $\leq$ but assume that $[\![\text{Eli climbed the mountain}]\!] = \{9\text{pm}\}$ rather than $\{[6\text{pm},9\text{pm}]\}$ (as Beaver & Condoravdi 2003 suggest).

In our scenario where ⟦Eli climbed the mountain⟧ = {[6pm,9pm]}, (5b) can be restated as "there is a Ben-be-nervous time t' s.t. [6pm,9pm] precedes t′." That is, (5b) requires that Ben be nervous sometime after Eli finished his climb, as desired.[4]

2.1.2 Rett 2020

Having presented an example of an under-specification theory, I now contrast it with Rett 2020. Rett (2020)-style entries for *before/after* are given in (6). These entries are slightly simplified relative to her true entries but are equivalent for our purposes.[5]

(6) a. $[\![\text{before}]\!] = \lambda \mathrm{q}_{\langle i,t\rangle}.\ \lambda \mathrm{p}_{\langle i,t\rangle}.\ \exists \mathrm{t}' [\mathrm{p}(\mathrm{t}') = 1 \wedge \forall \mathrm{t}'' [\mathrm{q}(\mathrm{t}'') = 1 \rightarrow \mathrm{t}' < \mathrm{t}'']]$
b. $[\![\text{after}]\!] = \lambda \mathrm{q}_{\langle i,t\rangle}.\ \lambda \mathrm{p}_{\langle i,t\rangle}.\ \exists \mathrm{t}' [\mathrm{p}(\mathrm{t}') = 1 \wedge \forall \mathrm{t}'' [\mathrm{q}(\mathrm{t}'') = 1 \rightarrow \mathrm{t}' > \mathrm{t}'']]$

(7) shows the predicted truth-conditions for (1a-b). Note that since *Eli climbed the mountain* is a singleton, there is only one time that satisfies the restrictor of the universal quantifier in (7): namely, the runtime of Eli climbing the mountain.[6]

(7) ⟦[[Ben was nervous] [$\boxed{\textbf{before}}$/after [Eli climbed the mountain]]]⟧ = 1 iff
$\exists \mathrm{t}' [\mathrm{t}' \in [\![\text{Ben was nervous}]\!] \wedge \forall \mathrm{t}'' [\mathrm{t}'' \in [\![\text{Eli climbed the mt.}]\!] \rightarrow \mathrm{t}' \boxed{<}/\!> \mathrm{t}'']]$

(7) gives the correct truth-conditions for (1b) but not for (1a). To see this, let us suppose again that Eli climbed the mountain in three hours (6-9pm) and consider three scenarios for when Ben was nervous (Scenario A, 4pm-5pm; Scenario B, 7-8pm; Scenario C, 10-11pm). (1b) is correctly predicted to be true in Scenario C only, as that is the only scenario where there is a Ben-be-nervous time that follows the runtime of Eli climbing the mountain. The analogous prediction for (1a) is that it should only be true in Scenario A (the scenario where there is a Ben-be-nervous time that precedes Eli's runtime), but it is intuitively true in Scenario B too.

To capture the before-finish reading, Rett posits that whenever we judge (1a) true in a scenario like Scenario B, we are coercing *Eli climbed the mountain* into an achievement (*Eli finished climbing the mountain*). Rett (2020) formalizes this idea using a **completive coercion** operator, which she claims is only defined for telics.[7]

4 Note that this theory is "non-uniform" in that the entries for *before/after* differ not just in the presence vs. absence of negation but also in their use of $\leq_\forall$ vs. $<$. This non-uniformity is not easily avoided: if one used $\leq_\forall$ in the entry for *after*, we would wrongly predict that (1b) is true in a scenario where Ben was nervous at 9pm but wasn't afterwards. Beaver & Condoravdi (2003) consider this non-uniformity a problem for the quantificational theory; see Krifka (2010) for an attempt to remedy the problem.

5 In particular, Rett's entries invoke *maximality* rather than universal quantification over times, but the two notations are truth-conditionally equivalent.

6 The universal quantification in (6), while more or less vacuous in (7), plays a crucial role in Rett's account of cases where the embedded clause denotes a non-singleton; see section 6 for discussion.

7 $\sqsubseteq$ denotes the subinterval relation. Also, Rett (2020) cites Dölling 2014 for the notion of completive coercion, but it seems to me that Dölling's notion of completive coercion is different from Rett's.

(8) $\textsc{complet}(p) = \{\iota t.\ \exists t'\ [p(t') = 1 \wedge t \sqsubseteq t' \wedge \forall t''\ [t'' \sqsubseteq t' \rightarrow [t > t'' \vee t \sqsubseteq t'']]]\}$

In words, COMPLET takes a singleton *p*, finds the latest point contained in *p*'s unique member, and returns a set containing that point. For example, if ⟦Eli climbed the mountain⟧ = {[6pm,9pm]}, then COMPLET(⟦Eli climbed the mountain⟧) = {9pm}.[8]

(9a) shows the predicted truth-conditions for (1a) when COMPLET is inserted: while the coercion-less LF for (1a) was only true in Scenario A, (9a) is also true in Scenario B. COMPLET can in principle attach to a telic *after*-clause too, but it would be vacuous: (9b), just like the coercion-less LF for (1b), is only true in Scenario C.

(9) a. ⟦[[Ben was nervous] [before [COMPLET [Eli climbed the mountain]]]]⟧ = 1 iff $\exists t'\ [t' \in$ ⟦Ben was nervous⟧ and $\forall t''\ [t'' \in \{9pm\} \rightarrow t' < t'']]$

b. ⟦[[Ben was nervous] [after [COMPLET [Eli climbed the mountain]]]]⟧ = 1 iff $\exists t'\ [t' \in$ ⟦Ben was nervous⟧ and $\forall t''\ [t'' \in \{9pm\} \rightarrow t' > t'']]$

As mentioned in section 1, Rett (2020) argues that COMPLET is independently needed to deal with examples like (2). She also observes (on the basis of a 17-language survey) that all languages seem to allow before-start readings of *before*, but not all permit before-finish readings; this fact arguably favors a view (such as Rett's (2020) own) in which before-start readings are always "basic" and before-finish readings are derived via a mechanism that not all languages have.

Taken together, these arguments provide reasons for thinking that Rett's (2020) theory may have conceptual advantages. However, there is a dearth of English-internal empirical arguments that favor Rett's (2020) theory of *before*/*after* over an under-specification theory (or vice versa), and hence the debate between the two theories is not settled. This paper contributes to the debate by identifying and testing two processing predictions that Rett's (2020) theory makes, predictions that become clear once we take a closer look at the existing literature on coercion phenomena.

2.2 Coercion: Theoretical and experimental perspectives

There are a variety of phenomena for which coercion-based theories have been proposed. In addition to the cases discussed so far (to which I return in section 3), two particularly well-studied cases involve sentences like (10a-b). In (10a), an entity-denoting phrase (*the book*) takes on an eventive meaning ('reading the book') in the presence of a verb like *finish*; in (10b), a punctual verb (*jump*) takes on an iterative meaning ('jump repeatedly') in the presence of a *for*-adverbial.

8 Since COMPLET(p) needs *p* to be a singleton, we run into an issue with non-singleton accomplishments (*climbed a mountain*). To remedy this issue, we could redefine COMPLET(p) so as to find the interval in *p* with the earliest telos and return the singleton containing that telos. Then, *before* COMPLET *Eli climbed a mountain* will mean "before the telos of Eli's first climb," which seems correct.

(10) a. I finished the book that you told me to read.
b. The zebra jumped for an hour.

Proponents of coercion-based theories generally assume that sentences with coercion contain a covert operator that is inserted to repair an otherwise ill-formed sentence (Moens & Steedman 1988; Pustejovsky 1991, and subsequent work). In some cases, the proposed ill-formedness that requires repair is a compositional semantic mismatch. For example, proponents of a coercion theory of (10a) assume that there is a type mismatch between *finish* and entity-denoting expressions; to resolve the mismatch, we insert a **complement coercion** operator, which coerces *the book* into having an eventive meaning (Jackendoff 1997; McElree et al. 2001; Traxler et al. 2002). In other cases, the proposed ill-formedness that requires repair can either be thought of as semantic or pragmatic. For example, some proponents of a coercion theory of (10b) posit that there is a compositional mismatch between *for*-adverbials and punctual verbs like *jump* (de Swart 1998). Others suggest that the semantic derivation for (10b) is mismatch-free but that the resulting meaning ("the zebra jumped once, and it lasted for an hour") clashes with world knowledge (Dölling 2014). Both variants of the coercion theory posit that (10b) is repaired by insertion of an **iterative coercion** operator, whose meaning is "repeatedly."

There are yet other coercion cases where the proposed ill-formedness that requires repair can only be thought of as pragmatic (Dölling 2014). As an example, consider *Mike swam for twenty years*, whose most natural reading is a habitual one ("Mike was a swimmer for twenty years") rather than a literal one ("Mike swam non-stop for twenty years"). It would be implausible to claim that the habitual reading arises because of a compositional mismatch between *swam* and *for twenty years*; after all, the literal meaning is still accessible in the right (fantastical) context. Rather, Dölling (2014) argues, the habitual reading arises because the literal meaning conflicts with world knowledge, necessitating insertion of a habitual operator.

Coercion theories make strong predictions about real-time processing (Todorova et al. 2000; Brennan & Pylkkänen 2008; Lukassek et al. 2017, etc.), which hold regardless of whether the proposed ill-formedness that requires repair is semantic or pragmatic (Brennan & Pylkkänen 2008). In particular, if a sentence is underlyingly ill-formed and requires emergency insertion of a covert operator, we expect that parsers encountering such sentences will show processing slowdowns upon realizing that operator-insertion is necessary; these slowdowns reflect the additional cognitive operations involved in inserting the operator.

Because coercion theories make distinctive processing predictions, psycholinguistic studies have been an important testbed for coercion theories of various phenomena. The majority of such studies look at cases like (10a-b), finding that these sentences are indeed hard to process in the region where the coercion account predicts a slowdown (McElree et al. 2001 and subsequent work for (10a); Piñango et al. 1999 and subsequent work for (10b)). For example, all studies on (10a) known to me have found that (10a) is hard to process at or one word after *book*; this is expected under the coercion account, as parsers have seen *finish* + an entity-denoting complement at that point and thus know that complement coercion is required.

3 The processing predictions of Rett 2020

In this section, I put Rett 2020 in conversation with the broader coercion literature by (a) sharpening Rett's invocations of coercion using the language of "ill-formedness and repair" that characterizes the coercion literature; (b) contrasting the processing predictions of Rett's (2020) theory with those of under-specification theories.

One can sharpen Rett's (2020) theory of sentences like (11) either by appealing to a semantic mismatch or by appealing to a pragmatic violation. If one opts for the semantic route, one would say that there is a compositional mismatch between the durative VP *climbed the mountain* and the punctual *at*-adverbial. If one opts for the pragmatic route, one would say that the LF for (11) is mismatch-free but that its meaning ("at 7pm sharp, Eli both started and finished climbing the mountain") clashes with world knowledge that mountains cannot be climbed in an instant. Regardless of how one conceptualizes the ill-formedness of (11), insertion of COMPLET repairs the sentence because COMPLET(⟦Eli climbed the mountain⟧) is punctual.

(11) At 7pm sharp, Eli climbed the mountain.

Both the semantic and pragmatic variants of the theory predict that in a context where Eli climbed a single mountain and reached the summit at 7pm, (11) should be hard to process at or just after *climbed*; at that point, the parser has seen the beginning of an accomplishment but know that the climb ended at 7pm, so they should know that completive coercion is required.[9] **Experiment 1** tests this prediction.

We cannot sharpen Rett's (2020) theory of (12) by appealing to a compositional issue, as the derivation for (12) does not crash without COMPLET, according to Rett; rather, the coercion-less LF merely expresses a different reading than the LF with COMPLET. However, it is possible to sharpen her theory of (12) by assuming that insertion of COMPLET is only licensed when the coercion-less reading conflicts with contextual knowledge. In other words, COMPLET is only ever inserted to repair a mismatch between the "default," before-start truth-conditions of (12) and the context.

(12) Ben was nervous before Eli climbed the mountain.

9 Since *climbed* can have atelic continuations (*climbed mountains*), one might think that parsers will not know whether to insert COMPLET until reaching *the*; after all, Rett posits that COMPLET selects for telics only. But Eli only climbed one mountain in the context, so the atelic continuation is implausible.

Under this theory, we might expect that in a context that is only compatible with a before-finish reading (and in which Eli only climbed one mountain), (12) should be hard to process at or just after *climbed*. At that point, parsers have seen the beginning of an accomplishment in a *before*-clause but know that only a before-finish reading is compatible with the context; they thus spot the mismatch between (12)'s default truth-conditions and the context and know to insert COMPLET. Crucially, reading times at or just after *climbed* in (12) should be longer in a before-finish context than in a before-start context, since the default truth-conditions of (12) do not conflict with a before-start context. **Experiment 2** tests this prediction.

Under-specification theories do not share these predictions about (12). Absent further assumptions, such theories expect that the processing of (12) should be unaffected by context (so long as the sentence is true) and are agnostic about (11).

I conclude this section with a caveat. Past work has found online costs for coercions that repair a semantic mismatch (e.g. complement coercion) and for coercions that repair an issue that can either be thought of as semantic or pragmatic (iterative coercion). However, no processing studies I know of look at claimed coercion cases whose ill-formedness must be thought of as pragmatic (e.g. (12)'s before-finish reading or the habitual reading of *Mike swam for twenty years*, see sec. 2.2). Thus, it remains unknown whether semantic and pragmatic coercion have the exact same processing profile. Past work has assumed that they should have the same signature in behavioral paradigms like self-paced reading (Brennan & Pylkkänen 2008), however, so the online predictions of Rett's theory are best stated as above.[10]

4 Experiment 1

4.1 Methodology

4.1.1 Participants and materials

110 self-reported native English speakers from Prolific participated in Experiment 1. The study took ∼25 minutes, and participants were paid £4.20 apiece.

Participants in Experiment 1 saw 26 trials, each of which had three phases. First, participants read a context (e.g. (13)) sentence-by-sentence. Then, participants read a target sentence word-by-word using a self-paced reading interface, and we measured how long it took them to read each word. Finally, participants rated the naturalness of the target sentence given the context on a 5-point Likert scale. We included a

10 As Brennan & Pylkkänen (2008) note, paradigms like magnetoencephalography (MEG) are better-suited than behavioral paradigms to address the question of whether a given instance of coercion is motivated by a compositional mismatch or a world-knowledge violation. This is because in MEG, world knowledge violations trigger a so-called "N400" effect that compositional mismatches do not. Self-paced reading has no comparable way to tease apart compositional vs. pragmatic violations.

naturalness-rating component because it allowed us to construct attention checks (see sec. 4.2) and because ratings can provide further insight into participants' behavior.

On critical trials, participants saw one of four types of target sentences. For example, participants would read one of the sentences in Table 1 after reading (13).[11] The first two types of target sentences were completive coercion sentences with *at*-modifiers (e.g. (a) in Table 1) and controls with *in*-modifiers (b). Sentences like (b) are suitable controls because while (b) is structurally akin to (a), no theory predicts that (b) should be costly. The other two target sentence types were cases of complement coercion (c) and controls with non-coercive verbs (d). These two trial types served as a sanity check: replicating the finding that (c) is costlier than (d)–see sec. 2.2–helps show that my paradigm is sensitive enough to capture coercion effects.

(13) Context: Hector is a lifelong hiker and his daughter Emma wants to learn more about the outdoors, so they went on a hiking trip. On the first night, Hector bragged about how quickly he could set up a tent, and Emma bet him that he couldn't do it in 10 minutes. He started building the tent at 8:55pm, and the tent was fully set up **at 9pm**./*five minutes after he started*.

	Completive	Complement
Coercion	(a) **At 9pm sharp Hector built the humble tent.**	(c) Hector finished the tent in front of Emma.
Control	(b) *In five minutes Hector built the humble tent.*	(d) Hector built the tent in front of Emma.

Table 1 Sample stimuli for Experiment 1, which was designed in PCIbex (Schwarz & Zehr 2021); click here for a sample trial with (13) + (a).

While most previous studies on coercion did not use contexts, I felt that contexts were needed to ensure that (a) was interpreted with completive coercion (*at 9pm, Hector finished*) rather than in another way. Note that the italicized phrase in (13) was absent from (a)'s context, while the bolded phrase was absent from (b)'s context. While I could have used, e.g., *at 9pm* in both contexts, this might have made the first few words of (a) easier to process than (b) (because *at 9pm* is repeated), and I wanted to minimize the chance of pre-verbal differences between (a) and (b). Trials with targets like (c)-(d) used contexts containing both the bolded and italicized phrases (or analogous phrases in other items), e.g. *at 9pm, five minutes after he started*.

To make the task more visually engaging, I had each trial unfold next to a picture. For example, trials using the context in (13) unfolded next to a stylized picture of a

11 As detailed below, both experiments had item sets for a variety of predicates (not just *build the hut*), including *climb the mountain*. I show our build-the-hut stimuli here just for a bit of variation.

tent in the woods. The pictures, which were generated by the Adobe Firefly software, were the same across the four trial types, so their use should not affect the results.

Besides the trial types in Table 1, there were two filler types: ungrammatical fillers, whose target sentences violated subcategorization (e.g. *During the meeting Adam meowed the confused student*), and grammatical fillers, some of whose target sentences were true given the context and some of whose target sentences were not.

4.1.2 Procedure

Participants first saw instructions and a practice trial (available here). Then, they were sorted into one of four lists, each of which had 16 critical trials and 10 fillers (presented in random order). The 10 fillers were the same across lists and consisted of 4 ungrammatical fillers and 6 grammatical fillers (4 true, 2 false). For the critical trials, I created 16 item sets like the item set in (13) + Table 1 and used a Latin-square design where each participant saw each item set once. Hence, participants saw 4 exemplars of each of the 4 critical trial types. Each item set used a different accomplishment, e.g. *built the tent*, *baked the cake*, *climbed the mountain*, etc.

4.2 Results

I analyzed the data in R (R Core Team 2025), version 4.5. For analyses on the self-paced reading results, I excluded data from participants who gave a rating of 3 or higher on at least three out of the six trials with clearly unnatural target sentences (ungrammatical fillers and false fillers). By this criterion, 31 participants were excluded.[12] I also excluded reading times (RTs) that were more than three standard deviations from the mean for each word type ("word type" meaning subject, verb, postverbal determiner, etc.). In both experiments, I opted to use log-transformed RTs rather than raw RTs, as the former were more normally distributed.

Figure 1 sums up logRTs for the analyzed subset of the data. Its left panel shows logRTs for complement coercion trials like (c) in Table 1 and controls like (d) around the region of interest for complement coercion. Recall that following past studies, we expect longer RTs for trials like (c) vis-à-vis controls like (d) at the direct object noun or noun+1. The right panel shows logRTs for trials like (a) in Table 1 and controls like (b) around the region of interest for completive coercion. Recall that Rett (2020) expects longer RTs for trials like (a) vis-à-vis controls at verb or verb+1.

12 This exclusion rate is quite high, especially since the exclusion criterion I used is not very strict. Even though my exclusion criterion yielded a high exclusion rate, I decided not to tweak it *post hoc* since it was chosen before data collection. One possible reason for the high exclusion rate is that Likert ratings are just not an ideal way to gauge attention; after all, even attentive participants might not all use a Likert scale in the same way. Hence, I gauged attention differently in Experiment 2; see 5.1.1.

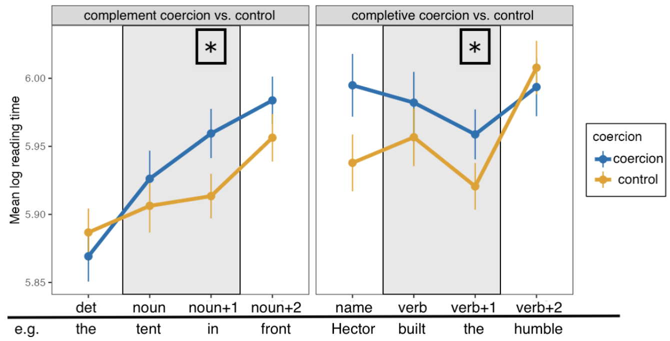

Figure 1 Mean logRTs in Experiment 1, with the regions of interest in each panel highlighted in grey. Error bars are SEM.

To test these hypotheses, I fit a pair of mixed-effects linear regression models using the *lmerTest* package in R (Kuznetsova, Brockhoff & Christensen 2017). Both models used data from all four critical trial types, with the first model being fit on critical word logRTs (i.e. the verb in trials like (a) and (b) in Table 1; the direct object noun in trials like (c) and (d)) and the second model being fit on critical+1 logRTs (verb+1 in trials like (a) and (b); noun+1 in trials like (c) and (d)). The models predicted logRT from trial type, with the maximal random effects structure that converged: by-item random intercepts as well as by-participant random intercepts and slopes for trial type. The models also included word length, log lexical frequency,[13] and previous word logRT as covariates.[14] Using the *faux* package (DeBruine 2025), I coded the "trialtype" variable so as to extract three contrasts: (I) completive coercion vs. matched controls; (II) complement coercion vs. matched controls; (III) completive coercion vs. complement coercion. I am only interested in (I) and (II), but I had to include a third contrast because the "trial type" variable had four levels. I only interpreted the p-values corresponding to (I) and (II).

The model on critical word logRTs found no significant differences between trial types. The model on critical+1 logRTs, by contrast, found that completive coercion sentences had significantly longer logRTs than matched controls at verb+1 (β = 0.03, SE = 0.0145, p = 0.02), and complement coercion sentences had significantly longer logRTs than matched controls at noun+1 (β = 0.03, SE = 0.014, p = 0.01).

Turning to naturalness ratings, Table 2 shows mean ratings for the four critical trial types (for reference, ungrammatical fillers' mean rating was 1.7). I analyzed the ratings data via two mixed-effects ordinal logistic regressions fit with the *ordinal* package (Christensen 2023). The models predicted rating from trial type, with by-

13 Frequencies were retrieved from the English Lexicon Project (Balota, Yap, Hutchison, Cortese, Kessler, Loftis, Neely, Nelson, Simpson & Treiman 2007), following Brennan & Pylkkänen 2010.

14 The R formula is: logRT $\sim$ trialtype + length + prevlogRT + freq + (1+trialtype | subjID) + (1 | item).

	completive	complement
coercion	3.56 (± 0.066)	3.58 (± 0.066)
control	4.15 (± 0.052)	3.89 (± 0.058)

Table 2 Mean naturalness ratings in Experiment 1, with SEM in parentheses.

item random intercepts and by-participant random intercepts and slopes for trial type. The first model used data from completive coercion trials, matched controls, and ungrammatical fillers, while the second used data from complement coercion trials, matched controls, and ungrammatical fillers. In each model, the 3-level "trial type" variable was Helmert-coded to extract two contrasts: ungrammatical vs. grammatical (where "grammatical" pools coercion+control) and coercion vs. control.

My first model found that ungrammatical fillers were less natural than completive coercion trials+controls pooled (β = -1.39, SE = 0.15, $p < 0.001$), and completive coercion trials were less natural than controls (β = -0.53, SE = 0.08, $p < 0.001$). The other model found that ungrammatical fillers were less natural than complement coercion trials+controls (β = -1.28, SE = 0.19, $p < 0.001$), and complement coercion trials were less natural than controls (β = -0.24, SE = 0.07, $p < 0.001$).

4.3 Discussion

Experiment 1 bore out the prediction of Rett's (2020) theory: sentences like (a) in Table 1 had slower logRTs than controls like (b) just after the main verb. Experiment 1 also replicated past findings about complement coercion: sentences like (c) in Table 1 had slower logRTs than controls like (d) just after the direct object noun. This sanity check shows that one can detect known coercion effects in my paradigm. Finally, sentences like (a) and (c) were less natural than controls, much like certain other cases of coercion (Todorova et al. 2000; Brennan & Pylkkänen 2010).

5 Experiment 2

Experiment 1 found an online cost for cases like (2) in self-paced reading. I now turn to whether before-finish readings of *before*-clauses incur a similar cost.

5.1 Methodology

5.1.1 Participants and materials

120 self-reported native English speakers from Prolific participated in Experiment 2 and were paid £4.50 apiece for their participation in the ~30 minute study. Partici-

pants saw 24 trials with the same basic structure as Experiment 1 (context, target sentence, then rating). However, Experiment 2 had three critical trial types rather than four (Table 3). On no-coercion trials like (a) in Table 3, participants read sentences with accomplishments in *before*-clauses after seeing before-start contexts like Context A, in which Emma's irritation emerged before Hector started building the tent. Completive trials like (b) used the same target sentences as no-coercion trials but used before-finish contexts like Context B, in which Emma's irritation emerged before Hector finished building the tent but not before he started. Finally, complement trials used complement-coercion sentences with *finish* and contexts like Context B. As before, trials like (c) were included as a sanity check; following past work, we expect (c) to have longer RTs than (a) at or just after *tent*.

(14) a. **Context A**: Hector is a lifelong hiker and his daughter Emma wants to learn more about the outdoors, so they went on a hiking trip in the woods. It got very windy, so they decided to build a tent and wait in it until the wind stopped. Hector pulled out the camping equipment at 4pm and took longer than usual to set up the tent because of the wind; however, the tent was fully up by 4:30. Emma, who had become irritable when the wind started (2pm), regained her composure once she saw the full tent at 4:30.

b. **Context B**: Same first three sentences as (14a), then: Emma, who had become very irritable when Hector made a big mistake and had to start over at 4:15pm, regained her composure once she saw the full tent at 4:30pm.

Trial type	Context	Target Sentence
(a) no-coercion	A	Emma was irritable before Hector built the tent in the woods.
(b) completive	B	Emma was irritable before Hector built the tent in the woods.
(c) complement	B	Emma was irritable before Hector finished the tent in the woods.

Table 3 Sample stimuli for Experiment 2.

As participants read a context, they saw timelines to help them keep track of events that would be mentioned in the target sentence. For example, on the trials in Table 3, a timeline showing the duration of Hector building the tent appeared once participants started to read the third sentence of the context. Once they moved to the last sentence of the context, the timeline morphed into one that showed the duration of Hector building the tent and the duration of Emma's irritation. To see what the timelines looked like, click here to run through (a)-(b) on PCIbex. I used timelines because I felt that it would be hard to keep track of all the information otherwise.

As in Experiment 1, there were both ungrammatical and grammatical fillers. Ungrammatical fillers used target sentences with selectional violations, as before. Some grammatical fillers used target sentences that made true claims about temporal order, e.g. *Jane was ready after Sue loaded the car...* in a context where Sue loaded the car 5-5:30am and Jane was ready 6:30am. Other grammatical fillers involved false claims about temporal order, e.g. *Alex was relieved before Jim wrote the thesis...* in a context where Jim wrote the thesis 7am-7pm and Alex was relieved 8pm.

One further difference between the two experiments bears mention. In Experiment 1, participants rated the target sentence's naturalness on every trial. In Experiment 2, by contrast, participants inputted a naturalness rating on critical trials and ungrammatical fillers but had to make a binary "True or False" judgment instead on grammatical fillers. I made this choice because I hypothesized that "True or False" judgments might be a better attention check than 5-point Likert scale ratings.

5.1.2 Procedure

After a practice trial (available here), participants were sorted into one of four lists, each of which had 12 critical trials (4 x 3 types) and 12 fillers (4 true, 6 false, 2 ungrammatical). The ungrammatical fillers, grammatical+true fillers, and two of the grammatical+false fillers were the same across lists. The other four grammatical+false fillers drew from the same contexts as critical trials but used false *after*-sentences (e.g. (14a) + *Emma was irritable after Hector built the tent in the woods*). For these fillers and critical trials, I created 16 item sets like the one in Table 3 and adopted Experiment 1's Latin-square design. Trial order was fully random.

5.2 Results

Very long and short RTs (defined as in section 4.2) were excluded, along with data from the 30 participants who answered "True" on three or more of the six grammatical+false fillers.[15] Figure 2 sums up logRTs on the analyzed subset of the data. Its left panel shows logRTs for complement trials like (c) in Table 3 vs. no-coercion trials like (a) around the region of interest for complement coercion, while its right panel shows logRTs for completive trials like (b) vs. no-coercion trials in and around the region of interest for completive coercion (verb, verb+1).

For complement and no-coercion trials, I fit two models predicting logRT from trial type (no-coercion vs. complement), with the same covariates and random

15 This exclusion rate of 30/120 (25%) is lower than Exp. 1's rate (31/110, 28%), but not by much, which suggests that Exp. 1's high rate was largely not due to its use of Likert-item attention checks but rather due to, e.g., the length/complexity of the task. Whatever the cause of these high exclusion rates, I believe the results suggest that my manipulations were effective for attentive paticipants.

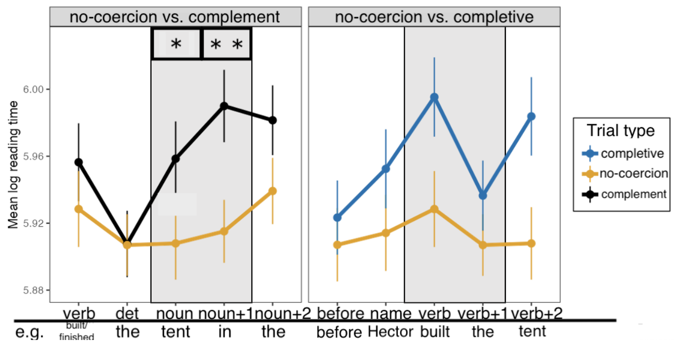

Figure 2 Mean logRTs in Experiment 2. Error bars are SEM.

effects as Experiment 1. The two models, which were fit on direct object noun and noun+1 logRTs, respectively, found that complement trials had longer logRTs than no-coercion trials at noun (β = 0.058, SE = 0.24, p = 0.03) and noun+1 (β = 0.057, SE = 0.017, p = 0.001). In addition, I fit two more mixed-effects models with the same structure on the data from completive and no-coercion trials. The two models, which were fit on verb and verb+1 logRTs for these trial types, respectively, found no logRT differences between completive vs. no-coercion at verb (β = 0.036, SE = 0.022, p = 0.12) or verb+1 (β = 0.008, SE = 0.017, p = 0.64). Note that while the contrast between no-coercion vs. completive at verb looks sizable, the verb model tells us that this effect is not robust when controlling for, e.g., previous-word logRT.

In addition to the above analyses, which tested the hypotheses I went into the experiment with, I ran an exploratory analysis to see if there was within-experiment learning that diluted the completive coercion effect. For example, maybe participants were processing sentences incrementally early on in the experiment, but as they became acclimated to the experiment and got more fatigued, they started reading sentences more quickly and processing them less incrementally. If this hypothesis is correct, we would expect to find a strong completive coercion effect localized to the verb on early trials, an effect that weakens/becomes less localized in later trials.

To test this hypothesis, I compared how participants acted on the first no-coercion trial and completive trial they saw (Figure 3, left panel) to how they acted on the remainder of no-coercion and completive trials (right panel). In the left panel, logRTs for completive vs. no-coercion trials are much more different at verb than name. Accordingly, the completive vs. no-coercion difference at verb in Figure 3's left panel is significant (β = 0.101, SE = 0.05, p = 0.04, similar model structure as above). In the right panel, not only do we see faster logRTs across-the-board and a weaker contrast (suggestive of shallower processing), but logRTs in the right panel look about the same at verb and name (indicating a lack of local slowdown at verb).

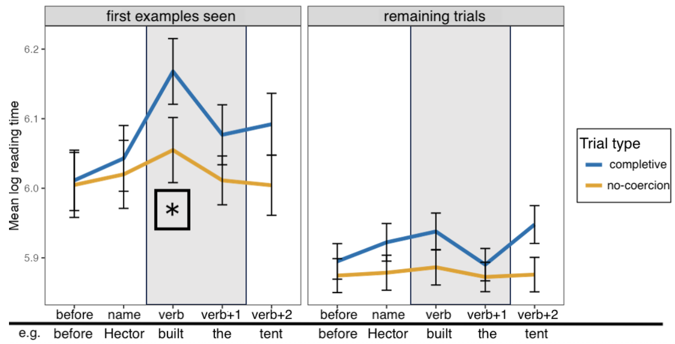

Figure 3 LogRTs on the first completive and no-coercion trial that each participant saw vs. logRTs on the remaining completive and no-coercion trials.

Finally, the mean ratings (out of 5) and SEM for no-coercion, completive, and complement trials were 4.19 ($\pm$0.06), 3.96 ($\pm$0.06), and 3.76 ($\pm$0.07), respectively. To analyze the ratings data, I fit a mixed-effects ordinal logistic regression on the three critical trial types; it predicted rating from trial type, with by-item and by-participant random intercepts and slopes. I dummy-coded "trial type" with "no-coercion" as reference level, and thus the model returned two contrasts: no-coercion vs. completive and no-coercion vs. complement. The former contrast was significant (β = -0.72, SE = 0.23, p = 0.001), as was the latter (β = -1, SE = 0.24, $p < 0.001$).

5.3 Discussion

Experiment 2 replicated past findings about the processing of complement coercion, verifying that we can detect coercion effects using my paradigm. In addition, I found that completive trials and complement trials, like some other cases of coercion (see section 4.3), are less natural than controls. Finally, while the predicted slowdown in completive trials did not occur in the full sample, it did occur in the subset of the data corresponding to participants' first encounter with each trial type. In other words, participants exhibited a completive coercion effect early on, but as they saw more trials, the effect weakened and became diffused. This finding is consistent with the hypothesis that processing became shallower as the experiment went on.

This hypothesis nicely explains the lack of completive coercion effect in the full sample, but one question remains: if the completive coercion effect was non-robust in the full sample because of shallow processing on later trials, why did the complement coercion effect not also fail to be robust in the full sample? *Prima facie*, shallow processing on later trials would influence both effects equally.

I can think of two explanations for this completive vs. complement asymmetry, both of which exploit the fact that completive coercion in *before*-clauses must be thought of as pragmatic (i.e. as repairing a clash between a sentence's truth-conditions and the context), while complement coercion must be thought of as semantic (i.e. as repairing a sentence-internal mismatch). First, perhaps participants' fatigue on later trials only caused "pragmatic" processing (i.e. verifying the target sentence's compatibility with the context) to become shallow, while lower-level "semantic" processing (e.g. verifying that there is no type clash between *finish* and its sister) remained incremental. Second, perhaps "pragmatic" coercions (whose processing has not been studied before, see section 3) become easier upon repeated exposure to coercion-forcing contexts, while "semantic" coercions remain hard.

Regardless of which explanation is correct, Experiment 2 clearly shows that vis-à-vis before-start readings of telic *before*-clauses, before-finish readings incur a cost in online processing and in offline ratings. While it would be too hasty to claim that Rett 2020 is the only imaginable theory that can capture these results, Rett 2020 is the only existing theory of that predicts any asymmetry between the two readings; she hence has a leg up on competitors when it comes to explaining the results.

6 General discussion and conclusion

This paper has investigated whether (1a)'s before-finish reading and (2) incur processing costs, as predicted by Rett's (2020) completive coercion theory. My results suggest that (1a)'s before-finish reading and (2) are indeed both costly, although the cost of the former weakens over time in a way that the cost of the latter does not. These results are more compatible with Rett 2020 than with competing underspecification theories, although the results of Experiment 2 require Rett (2020) to make further assumptions about, e.g., the processing of "semantic" vs. "pragmatic" coercion. By providing moderate support for Rett's idea that there is a covert operator COMPLET that does similar semantic work as overt morphemes in other languages, these results contribute to building a typology of covert aspectual operators.

The main direction for future research is to test a set of predictions that Rett's theory makes about **inchoative coercion**, an operation she posits in some atelic *after*-clauses. In a nutshell, Rett's theory of atelic *after*-clauses is a mirror image of her theory of telic *before*-clauses. As the reader can verify, Rett (2020) predicts that *Ben was cold after Sue was cold* is true iff Ben was cold after Sue stopped being cold; however, the sentence can also be judged true if Ben was merely cold after Sue started being cold (Anscombe 1964). To deal with these "after-start" readings, Rett (2020) claims that such readings involve coercing *Sue was cold* into meaning *Sue started being cold* (inchoative coercion). Following the logic from section 3, we see that this theory predicts that after-start readings of atelic *after*-clauses should incur a cost relative to "after-finish" readings. Testing this prediction is beyond the scope of this short paper, but doing so would contribute greatly to the theory of *before*/*after*.

References

Anscombe, Elizabeth. 1964. Before and after. *The Philosophical Review* 73(1). 3–24. doi:10.2307/2183199.

Balota, David A, Melvin J Yap, Keith A Hutchison, Michael J Cortese, Brett Kessler, Bjorn Loftis, James H Neely, Douglas L Nelson, Greg B Simpson & Rebecca Treiman. 2007. The English Lexicon Project. *Behavior Research Methods* 39. 445–459. doi:10.3758/BF03193014.

Beaver, David & Cleo Condoravdi. 2003. A uniform analysis of *before* and *after*. In Robert B. Young & Yuping Zhou (eds.), *Semantics and Linguistic Theory (SALT) 13*, 37–54. doi:10.3765/salt.v13i0.2899.

Brennan, Jonathan & Liina Pylkkänen. 2008. Processing events: Behavioral and neuromagnetic correlates of aspectual coercion. *Brain & Language* 106. 132–143. doi:10.1016/j.bandl.2008.04.003.

Brennan, Jonathan & Liina Pylkkänen. 2010. Processing psych verbs: Behavioral and MEG measures of two different types of semantic complexity. *Language and Cognitive Processes* 25(6). 777–807. doi:10.1080/01690961003616840.

Christensen, Rune H. B. 2023. *ordinal—Regression Models for Ordinal Data*. https://CRAN.R-project.org/package=ordinal. R package version 2023.12-4.1.

Condoravdi, Cleo. 2010. NPI licensing in temporal clauses. *Natural Language and Linguistic Theory* 28. 877–910. doi:10.1007/s11049-010-9115-z.

DeBruine, Lisa. 2025. *faux: Simulation for Factorial Designs*. doi:10.5281/zenodo.2669586. R package version 1.2.2.

Dölling, Johannes. 2014. Aspectual coercion and eventuality structure. In Klaus Robering (ed.), *Aspects, Phases, and Arguments: Topics in the Semantics of Verbs*, 189–226. doi:10.1075/slcs.152.05dol.

Dowty, David. 1979. *Word meaning in Montague Grammar*. Dordrecht: Reidel. doi:10.1007/978-94-009-9473-7.

Heinamäki, Orvokki. 1974. *Semantics of English temporal connectives*: University of Indiana PhD dissertation.

Husband, E Matthew, Alan Beretta & Linnaea Stockall. 2006. Aspectual computation: Evidence for immediate commitment. Talk given at AMLaP.

Jackendoff, Ray. 1997. *The architecture of the language faculty*. MIT Press.

Krifka, Manfred. 2010. *Before* and *after* without coercion: Comment on the paper by Cleo Condoravdi. *Natural Language and Linguistic Theory* 28. 911–929. doi:10.1007/s11049-010-9116-y.

Kuznetsova, Alexandra, Per B. Brockhoff & Rune H. B. Christensen. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13). 1–26. doi:10.18637/jss.v082.i13.

Lukassek, Julia, Prysłopska, Robin Hörnig & Claudia Maienborn. 2017. The semantic processing of motion verbs: Coercion or underspecification? *Journal of Psycholinguistic Research* 46. 805–825. doi:10.1007/s10936-016-9466-7.

McElree, Brian, Matthew J Traxler, Martin J Pickering, Rachel E Seely & Ray Jackendoff. 2001. Reading time evidence for enriched composition. *Cognition* 78(1). B17–B25. doi:10.1016/S0010-0277(00)00113-X.

Moens, Marc & Mark Steedman. 1988. Temporal ontology and temporal reference. *Computational Linguistics* 14(2). 15–28. doi:10.1093/oso/9780199268535.003.0007.

Piñango, Maria M, Edgar Zurif & Ray Jackendoff. 1999. Real-time processing implications of enriched composition at the syntax–semantics interface. *Journal of Psycholinguistic Research* 28. 395–414. doi:10.1023/A:1023241115818.

Pustejovsky, James. 1991. The syntax of event structure. *Cognition* 41. 47–81. doi:10.1016/0010-0277(91)90032-Y.

R Core Team. 2025. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria. https://www.R-project.org/.

Rett, Jessica. 2020. Eliminating *EARLIEST*: A general semantics for *before* and *after*. In Michael Franke, Nikola Kompa, Mingya Liu, Jutta L. Mueller & Juliane Schwab (eds.), *Sinn und Bedeutung (SuB) 24*, 201–218. doi:10.18148/sub/2020.v24i2.893.

Schwarz, Florian & Jeremy Zehr. 2021. Tutorial: Introduction to PCIbex – an open-science platform for online experiments: Design, data-collection, and code-sharing. In *Cognitive Science Society (CogSci) 43*, 15–16. https://escholarship.org/uc/item/1ng1q4c6.

de Swart, Henriëtte. 1998. Aspect shift and coercion. *Natural Language & Linguistic Theory* 16. 347–385. doi:10.1023/A:1005916004600.

Todorova, Marina, Kathy Straub, William Badecker & Robert Frank. 2000. Aspectual coercion and the online computation of sentential aspect. In *Cognitive Science Society (CogSci) 22*, 1–6. https://escholarship.org/uc/item/14x15786.

Traxler, Matthew J, Martin J Pickering & Brian McElree. 2002. Coercion in sentence processing: Evidence from eye-movements and self-paced reading. *Journal of Memory and Language* 47(4). 530–547. doi:10.1016/S0749-596X(02)00021-9.

Johanna Alstott
Department of Linguistics and Philosophy
Massachusetts Institute of Technology
32 Vassar Street
Cambridge, MA 02139
jalstott@mit.edu