Can semantic theories be tested experimentally?

The case of aspectual coercion

Oliver Bott

SFB 833, Eberhard Karls Universität, Tübingen

Abstract

This paper reflects upon the relation between semantic/pragmatic theory and psycholinguistic experiments within the realm of aspectual coercion. What predictions can be derived from standard and not so standard theories for incremental aspectual interpretation? We contrast operator-based and underspecification accounts of aspectual coercion with the Event Calculus of Hamm & van Lambalgen (2005), a pragmatic theory of tense and aspect, which proposes a classification of coercion into different types. We present the results of an ERP study on one particular kind of coercion (adding an eventuality = “additive coercion”) showing that aspectual coercion is not necessarily triggered by an aspectual mismatch. We will then present a self-paced reading experiment on contextual influences in additive coercion which suggests that aspectual processing immediately takes into account the larger discourse context. Taken together, the first two experiments provide clear evidence against a semantic operator-based account but indicate that aspectual coercion heavily relies on pragmatic inference. Two further self-paced reading experiments compared different subtypes of aspectual coercion with each other (deletion of an eventuality = “subtractive coercion” and

*I would like to thank the audiences of the TRAIT workshop How categorical are categories? in Wroclaw 2013 and an invited talk at the German department in Cologne 2013 for their comments on the work presented in this paper. I would also like to thank Johannes Dölling, Fritz Hamm, Janina Radó, Edith Scheifele, Matthias Schlesewsky and two anonymous reviewers for valuable comments on the research reported in this paper. The usual disclaimers apply, any remaining errors are mine. The research was funded by the German Science Foundation within projects C1 and B1 in the Collaborative Research Centers 441 and 833 at the University of Tübingen.
two types of iteration of an eventuality = “abstract type shift”). The results indicate that different kinds of coercion do in fact lead to differences in how they are processed. Most strikingly, different coercion subtypes which should involve the same operator – an iterative operator – revealed different coercion costs. Taken together, the experiments demonstrate that a operator-based coercion theory or underspecification account cannot fully account for the data. The predictions of Event Calculus, however, were fully confirmed by the experiments. Event Calculus thus allows us to derive empirically valid predictions about the underlying cognitive processes while incrementally constructing a temporal model of the unfolding discourse.

1 Introduction

Events are key components of how humans perceive the world and talk about it. The cognitive notion of events as reflected in our linguistic categories is, however, only indirectly connected to physical time in its Newtonian representation as a dimension comprising an infinite number of instants corresponding to the real numbers. In contrast, mental time involves the construction of a cohesive representation of ontological primitives like processes, states and different kinds of events by means of a whole range of relations like precedence, causality, and so forth (Vendler, 1957) which seem to be universal (Dahl, 1985; Smith, 1991). Psychological studies have shown that event perception can be conceived of as a fundamental cognitive capacity analogous to object perception (Zacks & Tversky, 2001).

This paper is about the cognitive functions underlying the construction of event representations during language perception. We will review a series of online experiments on aspectual coercion, i.e. implicit shifts from one lexical aspectual class to another, which may occur during aspectual composition with temporal modifiers. The findings will be discussed against the background of different coercion accounts from the semantic and pragmatic literature. Our aim is to show that experimental findings can, in fact, constrain the range of viable theoretical alternatives.
1.1 Lexical aspect, aspectual coercion and coercion theories

In the following we will use the event nucleus of Moens and Steedman (1988) in its slightly modified version (1) by van Lambalgen and Hamm (2005) to represent the five basic situation types or lexical aspectual classes found across languages (Smith, 1991).

(1) An eventuality is a structure \( \langle f_1, f_2, e, f_3 \rangle \), where

- \( f_1 \) is a fluent\(^1\), which represents a preparatory process, something which exerts a force,
- \( f_2 \) is a parametrized fluent which is driven by \( f_1 \), i.e. the incremental theme in the sense of Dowty (1979),
- \( e \) is the culminating event and
- \( f_3 \) is the resultant state.

Accomplishments like build a house or drink a glass of wine describe dynamic situations with a canonical goal. They are represented by a complete event nucleus \( \langle f_1, f_2, e, f_3 \rangle \). I cannot go into further detail here, but the semantic properties of \( f_1, f_2, e \) and \( f_3 \) are well defined in van Lambalgen and Hamm (2005, ch. 7)’s formalization of different event nuclei by their roles in the corresponding scenario. In the case of build a house, for instance, \( f_1 \) is a build-activity which drives a partial object \( f_2 \), a house gradually coming into existence, which culminates in a finish event \( e \); the resultant state corresponds to the complete house \( f_3 \). Its rich internal structure makes an accomplishment accessible for adverbial modification by in-adverbials (2-d) and as complement of finish (2-e), whereas modification by a for-adverbial (2-c) is semantically marked\(^2\).

\(^1\)The basic ontology used in the present paper is that of the Event Calculus of (van Lambalgen & Hamm, 2005). It employs a many sorted logic comprising fluents, i.e. any property that can change over time, and events as well as entities and real numbers to represent time instants.

\(^2\)Obviously, this interacts with the grammatical aspect of the verb since the example would be fully acceptable if the accomplishment were in the progressive instead of the simple past. We will largely ignore the interaction between lexical and grammatical aspect here and refer the reader to Bott and Hamm (2014) for further discussion.
The other four basic lexical aspectual classes are characterized by event nuclei that lack some of the meaning components present in accomplishments. Achievements like *win the race or find the keys* lack a preparatory process and an incremental theme. Instead, they express instantaneous change corresponding to an event nucleus of type \(<-,-,e,f_3>\). That they are unlike accomplishments is shown by their different behavior in combination with *for-* and *in-* adverbials (3-c)/(3-d) and their inacceptability as complement of *finish* (3-e).^3^  

Point action verbs like *sneeze* or *cough* express single events. Unlike achievements, points do not lead to a change of affairs and correspond to simple nuclei of type \(<-,-,e,->\). Modification by a *for-*adverbial automatically leads to an iterative interpretation (4-c). In the remainder of the paper, point action verbs will also be referred to as semelfactives (Smith, 1991).  

---  

^3^I deliberately chose variable amounts of time (*x time*) in order to show that the observed difference in acceptability isn’t due to differences in the typical duration of the denoted events. That is, no matter how short or long the interval denoted by the *for-*adverbial is, it can never modify an achievement. The only exception may be generic statements of the kind *for five years this club won the Champions League*. To me, these still sound marginal and seem to require overt quantification over events, e.g. by adding *every year.*
b. *John was sneezing.

c. John sneezed for x time.

d. *John sneezed in x time.

e. *John finished to sneeze.

Activities like *push a cart or *jog are dynamic eventualities of type \( \langle f_1, f_2, -, - \rangle \), i.e. activities with an incremental theme like *push a cart, or \( \langle f_1, -, -, - \rangle \), i.e. activities in the narrow sense like *jog. Activities lack a culmination. This explains why they are generally bad with in-adverbials (5-d) and cannot serve as complement of *finish (5-e).

(5) a. John pushed a cart.
   b. John was pushing a cart.
   c. John pushed a cart for x time.
   d. *John pushed a cart in x time.
   e. *John finished to push a cart.

States such as *know the answer express non-dynamic situations and – unlike the other aspectual classes – are generally bad in the progressive. This is shown in (6-b).

(6) a. John knew the answer.
   b. *John was knowing the answer.

The standard tests (see Dowty (1979, p. 60)) applied in (2)–(6) illustrate the special semantic properties of the five aspectual classes. Even though their different behavior suggests that they have to be conceptually distinguished, they are intimately connected to each other. An accomplishment, for instance, consists of an activity part, the preparatory process, and an achievement, i.e. its culmination. It is, however, more than a mere amalgam of these two ingredients since it establishes a causal connection between them. It is a primitive but has internal structure. The five lexical aspectual classes constitute the primitive building blocks of more complex event representations (see Bott (2010, ch. 2) for further discussion and motivation from psychological stud-
ies).

The examples (2)–(6) also illustrate the phenomenon of aspectual coercion, that is, shifts from one class to another in the absence of an overt operator (Moens & Steedman, 1988). These shifts are, for instance, triggered by the semantic input requirements of in- and for-adverbials. In the remainder of this paper, we will be concerned with aspectual coercion of types (7-a)–(7-e). A summary of the required coercion operations is provided in Table 1.

(7) a. Johann fand seinen Schlüssel in drei Stunden.
   John found his keys in three hours.

b. Heute joggte Johann in (nur) zwanzig Minuten
   Today, John jogged in (only) twenty minutes.

c. Johann belud die Schubkarre zehn Minuten lang.
   John was loading the wheelbarrow for ten minutes.

d. Johann belud die Schubkarre zwei Monate lang.
   John was loading the wheelbarrow for two months.

e. Johann nieste eine Stunde lang.
   John sneezed for an hour.

[Table 1 about here.]

The first two examples involve what we will refer to as additive coercion (c.f. the typology proposed by van Lambalgen and Hamm (2005, ch. 7) and taken up by Bott (2010)). Both the achievement in (7-a) and the activity in (7-b) are missing some eventuality to meet the input requirements of the in-adverbial which modifies an accomplishment. Therefore, in the case of (7-a) a preparatory process has to be added, whereas (7-b) requires adding a culmination. (7-c) exemplifies the opposite of the additive cases. The accomplishment is too rich to serve as input of the for-adverbial which modifies an activity. Therefore, the culmination has to be eliminated leaving only the preparatory process. While John loaded the wheelbarrow licenses the inference that the wheelbarrow was eventually loaded, sentence (7-c) does not trigger this inference. The iteration examples embody yet another coercion type. Here, the complete nucleus
of an accomplishment (7-d) and a semelfactive (7-e) are packed into a new $f_1$ fluent consisting of a series of loading or sneezing events, respectively.  

Semanticists working on aspectual coercion have proposed alternative theories to account for these coerced meanings. In the following, I will contrast three types of theories and lay out their predictions concerning the online interpretation of coercion types (7-a)–(7-e).

**Operator-based accounts** assume that coercion is triggered by a type-mismatch between two expressions A and B that have to be composed (e.g. de Swart, 1998; Rothstein, 2004). Let expression A be a functor of type $⟨\beta, \gamma⟩$ meaning that A takes an argument of type $\beta$ and outputs a value of type $\gamma$ and let expression B be an argument of type $⟨\alpha⟩$ with $\alpha \neq \beta$. A type-mismatch is a constellation $[A_{⟨\beta, \gamma⟩}][B_{⟨\alpha⟩}]$. It can be resolved by inserting a type-shifting operator $OP_{⟨\alpha, \beta⟩}$ from type $⟨\alpha⟩$ to type $⟨\beta⟩$ into the representation yielding $[A_{⟨\beta, \gamma⟩}][OP_{⟨\alpha, \beta⟩}][B_{⟨\alpha⟩}]$. The latter can be interpreted compositionally.

Operator-based accounts thus predict by and large uniform processing of the coercion types (7-a)–(7-e). All cases should initially lead to type mismatch. This mismatch is subsequently repaired by introducing a coercion operator into an enriched semantic representation. Obviously, the inserted coercion operators may differ across coercion types and this may be reflected by differences in coercion costs, respectively. However, the iterative cases (7-d) and (7-e) both require the insertion of an iterative operator and should thus be processed alike.

**Underspecification accounts** have been developed to represent ambiguities in an economical and elegant way (e.g. Dölling, 2003, 2014; Egg, 2005; Pulman, 1997).

In this paper I will not distinguish between iterative and habitual interpretations but will treat both, (7-d) and (7-e), uniformly as cases of iterative coercion. The difference between the two is that iterative coercion outputs an activity while habitual coercion outputs a habitual state. Obviously, this could make a difference when it comes to processing difficulty, a point raised by Johannes Dölling (personal communication). To see whether this is the case it would be relevant to compare processing costs of conditions such as (8-a) versus (8-b) where the former clearly triggers an iterative and the latter a habitual interpretation.

(8) a. John played the Minute Waltz for two hours.
   b. John played the Minute Waltz for two years.
They use a two-step procedure. First, an underspecified representation is computed. This representation - the semantic representation proper - leaves everything open which needs further specification. In a second step, the underspecified representation is transformed into a fully specified representation using pragmatic information from discourse context, conceptual knowledge et cetera.

A coerced sentence and its non-coerced counterpart share the same underspecified representation. Only the second step is different. While the non-coerced sentence is disambiguated by eliminating all potential coercion sites in the semantic representation, the coerced sentence is disambiguated by inserting coercion operators to fully specify the meaning. Under these accounts no semantic mismatch is expected to occur. Thus, if coercion should be difficult at all, the difficulty is solely due to selecting a pragmatically appropriate type-shift operator (or choosing the most plausible setting of parameters, Dölling (cf. 2014)). Assuming a single iteration-operator that can basically shift any Aktionsart into an iterative reading, (7-d) and (7-e) should be equally complex to process. Similar to operator-based accounts, the other coercion types may be different because here other types of operators are needed.

Planning accounts (Steedman, 2002; van Lambalgen & Hamm, 2005) make use of planning formalisms developed in artificial intelligence. According to these theories, the computation of the temporal profile for a piece of discourse can be viewed as coming up with a plan that permits a non-monotonic derivation of a goal state from an initial state. Coercion is characterized by shifts in meaning that result from the update of a plan when integrating new information. Depending on the required update, coercion can take very different forms.

1.1.1 Aspectual coercion in the Event Calculus

I will focus on the Event Calculus (EC), a planning account presented in van Lambalgen and Hamm (2005) for which an incremental version has been developed in Bott (2010). It is formalized in constraint logic programming. A sentence is under-

\footnote{See for instance Egg (2005, p. 97).}
stood as an instruction to update the situation or discourse model in a way that makes the sentence true. Updates can differ with respect to the number of steps that are required\(^6\). The computed model is minimal in the sense that everything one doesn’t have information about is assumed to be false. This is the *closed world assumption*. Every incoming piece of information is immediately incorporated into the model. Whenever possible, variables are unified with constants of the right type. I make the assumption that plausibility can only be evaluated after a model has been computed.

Discourse (9) illustrates the basic idea behind EC and how it accounts for closed world reasoning during incremental interpretation.

\[(9) \quad s_0 \{ \text{John was walking to the store} \} s_1 \{ \text{when he broke his leg} \} s_2\]

At the very beginning of the discourse, that is the initial state \(s_0\), the comprehender will start with an empty discourse model. The incoming discourse information is then taken to incrementally update the model with additional eventualities. Thus, at state \(s_1\) the reader will have to update it with the information available up to this point. In EC this corresponds to the instruction to make query (10) succeed which, formally, corresponds to an integrity constraint in constraint logic programming.

\[(10) \quad \text{There is a walk-to-the-store activity by John that holds at time } t_1 \text{ and } t_1 < \ now\]

The output of the computation is a minimal model in which there was a time \(t_1\) at which John was in the process of walking to the store. This process culminates, at some later time \(t_2\), in an event of arriving at the store. Minimality is guaranteed by building the closed world assumption into the semantics: everything the semantic processor doesn’t have information about is assumed to be false. Since no event is mentioned that would stop John’s walking activity, it is assumed that no such event happens and the activity

---

\(^6\) *Number of steps* is a notion that is exactly defined in logic programming since it corresponds to the number of applications of the fixed point operator until the least fixed point, that is the minimal model, is reached.
is assumed to go on until the culmination. This is different at state $s_2$. At this point, the discourse has to be updated with query (11-a).

(11)  
   a. A breaking event of John’s leg at $t_3$ and $t_3 = t_1$
   
   b. world knowledge: breaking a leg stops a person from walking

The resulting discourse model includes a breaking event that happens at the same time as walking. Given that world knowledge includes a rule like (11-b), the discourse model at state $s_2$ doesn’t include the culmination anymore. Discourse (9) thus exemplifies EC’s solution (c.f. Baggio, van Lambalgen, and Hagoort (2008); van Lambalgen and Hamm (2005)) to the imperfective paradox (Dowty, 1979), namely as recomputation of a default interpretation employing non-monotonic logic. Note that coercion type (7-c) requires a similar update.

In the following, I will give an informal description of the underlying computational mechanisms of coercion types (7-a)–(7-e). For a worked out formal analysis the interested reader is referred to Bott (2010, ch. 3).

**Additive coercion** is illustrated in (7-a) and (7-b). In both cases the situation model has to be updated with an eventuality not explicitly mentioned in the sentence. Here is what EC predicts to happen during incremental interpretation of (7-a). *John found his keys* will lead to an update of the discourse model with a finding event $e$ at some time $t$ that happened before now. In turn, $e$ brings about a change of state of affairs. Before $t$ John is in the state of not having his keys and not knowing where they are, but starting from $t$ this changes such that John now is in the state of being consciously aware of having them (= fluent $f_3$). Further updating the model with *in three hours* introduces an accomplishment nucleus with variables that have to be unified with particular event tokens and fluents of the situation model. The culmination and the resultant state can immediately be unified with $e$ and $f_3$. The fluents $f_1$ and $f_2$ corresponding to a preparatory process that lasts for three hours are missing, though. This triggers an abductive inference of a plausible process - searching for the keys, for instance -, an inference which heavily relies on world knowledge. The required abductive reasoning is right at
the heart of constraint logic programming. Making a particular query succeed consists in searching for a default rule like (12-a) which has the queried information in the consequent. In order to make the query succeed, the discourse model can be updated with the antecedent of the rule. The preparatory process has, in fact, to be inferred on the basis of world knowledge. This is illustrated by the alternative rule in (12-b) which is less plausible and should therefore be dispreferred.

(12)  
a. Searching for your keys results, per default, in finding them.  
b. Getting a hint about where your keys are, per default, results in finding them.

Since EC is an essentially pragmatic theory, it is expected that any difficulty will disappear if the preceding discourse contains the required eventualities. For instance, the coercion sentence embedded in discourse (13) is expected to be as easy as a sentence without coercion.

(13) After John had lost his keys he started to search the entire flat. He found them within three hours in the pocket of his coat.

Subtractive coercion in (7-c) proceeds along the lines outlined above when discussing non-monotonic inference in incrementally processing discourse (9). John loaded the wheelbarrow is interpreted as an accomplishment consisting of a preparatory process load ($f_1 = a$ loading activity of the wheelbarrow by John, $f_2 = the$ changing quantity of stuff in the wheelbarrow) followed by the culminating event finish. The resultant state $f_3$ is the fully laden wheelbarrow. The for-adverbial contains a process-variable that has to be unified with a process constant from the discourse representation. The only available process is load. The adverbial further introduces a stop-event that will terminate load before finish happens. Hence, the default interpretation of the accomplishment is shifted non-monotonically into an activity.

What’s the expected difficulty of subtractive coercion? German has no grammat-
ical aspect marking and the accomplishment in (7-c) therefore can easily receive an imperfective interpretation (see von Stutterheim, Andermann, Carroll, Flecken, and Schmiedtová (2012) and Bott and Hamm (2014)). This construction type should therefore be similar to the English example (14).

(14) John was building a house for two years.

Interestingly, the event-related potentials study by Baggio et al. (2008) suggests that processing examples like these should lead to measurable coercion costs. They investigated whether, in Dutch sentences of the type *The girl was writing a letter when her friend spilled coffee on the tablecloth/paper*, the resultant state (a complete letter) was computed online and canceled again. The critical words were *paper* compared to *tablecloth*. Spilling coffee on the paper implies that the resultant state was not attained. Thus, the system should be forced to revise the earlier commitment to the culmination. Spilling coffee on the tablecloth, however, does not have this implication. The former condition did result in a larger sustained anterior negativity compared to the latter condition. We therefore expected German subtractive cases to lead to processing difficulty, too.

Finally, coercion types (7-d) and (7-e) involve *iterative coercion*. For (7-d) to be true one needs to derive an activity predicate from a simple event. EC employs reification to model this type of shift (event $\rightarrow$ fluent). To give an impression of how this works it is insightful to look at imperfect nominalizations (see van Lambalgen and Hamm (2005, sec. 6.1 & ch. 12)). As an illustration, in (15-a) the gerund *John’s reaching the summit* has to be interpreted as a fluent and not as event. Note that *took place at ten* is a narrow container selecting for an event and not a fluent (Vendler, 1967). The example demonstrates that language has grammatical means to code predicates like *reach the summit* as fluents. (15-b) sketches how the required fluent term is constructed. It is the set of times $t$ at which *John reaches the summit at t* is true.

(15) a. John’s reaching the summit surprised us/*took place at ten.*
EC incorporates two alternative ways of how terms can be constructed from predicates with a time parameter $t$. In contrast to the construction of a fluent term, a predicate can also be constructed as an event term. This is illustrated in (16), i.e. a perfect nominalization.

(16)  

(a) John’s reaching of the summit took place at ten.  

(b) $e_{reach}(john, summit) = \exists t. reach(john, summit, t)$

Iterative coercion is the operation of reifying an event as a fluent (17). The constructed fluent term corresponds to the set of times at which the corresponding event tokens happened\(^7\). If this set comprises at least two instants of time, it trivially consists of an iteration of reaching events.

(17)  

$p_{reach}(john, summit) = \{t \mid \text{there is an event token of John reaching the summit at } t\}$

The operation of iterative coercion can turn any kind of event into a fluent, i.e. an iterative process. In contrast to operator-based accounts and underspecification, however, EC predicts clear processing differences between iterative coercion of semelfactives and accomplishments. Let’s come back to examples (7-d) and (7-e), respectively.

In (7-e) the semelfactive *John sneezed* enters a sneezing event in the aspectual representation. The for-adverbial requires a process fluent of type $\langle f_1, -, -, - \rangle$ that lasted for two hours before it was stopped. The representation of *John sneezed*, however, lacks any process the adverbial could be composed with. As a result, the only way to go is to shift the sneezing event into a process of iterative events via iterative coercion.

In (7-d), the situation is different. By definition, an accomplishment involves a

\(^7\)For a formal version of this operation the reader is referred to van Lambalgen and Hamm (2005, sec. 11.3.3).
process fluent $f_1$. This fluent is predicted to be unified with the process variable of the for-adverbial. Initially, the sentence should thus be processed exactly like (7-c) resulting in a subtractive coercion interpretation in which John was engaged in an activity of loading the wheelbarrow that lasted for two months. However, this reading turns out to be implausible and calls for further revision. If subtractive coercion is ruled out, the only way to go is an iterative interpretation\(^8\). This step involves the operation of iterative coercion as in (7-e)\(^9\). Taking everything together, (7-d) should thus be computationally far more complex than (7-c) or (7-e): while these require just one coercion step, (7-d) is predicted to require two coercions: subtractive coercion, rejection of the resulting interpretation due to pragmatic infelicity and computation of the iterative interpretation.

1.2 Existing experimental studies on aspectual coercion

Coercion has been on the agenda in psycholinguistic research for approximately fifteen years. The existing research in psycho- and neurolinguists has mainly concentrated on two types of phenomena: aspectual coercion from a punctual event into a series of events of type (7-e) (starting with Piñango, Zurif, and Jackendoff (1999)), and complement coercion from an object into an event, (18) (starting with McElree, Traxler, Pickering, Seely, and Jackendoff (2001)).

(18) John began the book.

Whereas early studies found coercion to be difficult in a number of tasks, Pickering, McElree, Frisson, Chen, and Traxler (2006) conducted two self-paced reading and two eyetracking experiments in which they didn’t find any coercion costs in iterative coercion. They took this lack of effect as evidence for aspectual underspecification

\(^8\)Additive coercion is impossible since accomplishments already correspond to maximal event nuclei.

\(^9\)Actually, in order to be able to perform iterative coercion, it is required to pack the complete scenario of the accomplishment into a single, new event, i.e. perfectivization. EC uses hierarchical planning to model perfective interpretations of complex events. Strictly speaking this adds another step which will be ignored here. But see Blaszczak and Klimek-Jankowska (this volume) for perfective versus imperfective grammatical aspect in Polish.
during ordinary reading. Brennan and Pylkkänen (2008) argued against aspectual underspecification and provided evidence that once experimental items are controlled for aspectual class, there is clear indication of aspectual coercion cost during self-paced reading and in magnetoencephalography. Recently, Paczynski, Jackendoff, and Kuperberg (2014) presented evidence from event-related potentials also showing that iterative aspectual coercion is costly. In their study, iterative coercion of very brief events was accompanied by a late, sustained anterior negativity relative to non-coercing control conditions and to explicitly frequentative sentences such as several times the cat pounced on the rubber mouse. However, the authors did not restrict the materials in the iterative coercion condition to a single lexical aspectual class, but tested ‘punctive verbs, describing a short, near-instantaneous action’ (Paczynski et al., 2014, p. 1909). In principle, the tested verbs could thus comprise semelfactives, achievements and accomplishments with brief preparatory processes. This makes it difficult to tell which aspectual transitions were responsible for the observed effect.

The phenomena that have been investigated experimentally only cover a small part of what has been discussed in the semantic literature. An exception are Pylkkänen, Brennan, and Bernis (2011) who compared different kinds of complement coercion with aspectual reinterpretation. However, they also didn’t look systematically at different subtypes of aspectual coercion. We have demonstrated above that in order to be able to decide between alternative theoretical proposals it is crucial to compare different coercion subtypes and, ideally, to do so in the same experiment. For instance, just knowing that (7-e) leads to difficulty during composition is not enough because all three classes of theories would led us to expect this result. In the following, I will briefly review experiments that have been reported in Bott (2010). It will be argued that a comparison of different subtypes of aspectual coercion in fact puts us in the position to relate psycholinguistic evidence to semantic and pragmatic theory.
2 Additive coercion – underlying processes and the role of context

The two experiments in this section serve to decide between operator-based coercion theories and pragmatic frameworks of coercion like underspecification or planning accounts. Both experiments investigated whether additive coercion involves a temporary type mismatch during semantic composition which is then resolved by introducing the right kind of coercion operator. In particular, we tested the following hypothesis which is central to operator-based accounts:

Reject and recompute: “Coercion operators are only inserted when they are triggered by a mismatch.” (de Swart, 1998, p. 8)

By contrast, planning accounts like EC predict that this type of coercion leads to an enrichment of the aspectual representation without revising any part of it. This is stated in the Smooth Update Hypothesis (c.f. Bott (2010, ch. 3)).

Smooth Update: Additive coercion consists in elaborating a given aspectual representation. The coercing stimulus adds an eventuality that has to be substituted with a concrete event from discourse context or world knowledge.

Why should there be no temporary mismatch in aspectual enrichment according to EC? Let us briefly reconsider the enrichment of an achievement into an accomplishment as illustrated in the mountaineer reached the top within three hours. The time span adverbial within three hours introduces a yet abstract accomplishment nucleus with variables that have to be unified with particular event tokens and fluents from the sentence. The culmination e and the resultant state f₃ can immediately be unified with the information introduced by the achievement the mountaineer reached the top. The fluents f₁ and f₂ corresponding to a preparatory process that lasts for three hours with an appropriate incremental theme are missing from the sentence, though. They may be thought of as free variables in the discourse representation which have to be instantiated by concrete fluents, respectively. However, except from this lack of concrete,
plausible \( f_1 \) and \( f_2 \), nothing is ill-formed about the event representation. Nevertheless, we may still expect enrichment to be costly because – without further context – the missing eventualities have to be inferred abductively. Importantly, the predicted cost should result from completely different underlying processes than those involved in the processing of aspectual mismatch.

As for underspecification accounts, it is not so clear what to expect. Semantic composition should work smoothly, but during pragmatic specification the right kind of type-shifting operator has to be introduced. Which computational mechanisms are yet involved in the specification stage is not spelled out in enough detail to be able to derive clear predictions. Without further qualification, I will, however, assume the *Smooth Update Hypothesis* for underspecification theories, too.\(^{10}\)

### 2.1 An ERP study on additive coercion (Experiment 1)

The first experiment to be reviewed here was an event-related potentials (ERP) study on additive coercion in German (see Bott (2010, ch. 7/Exp. 9) for further details). Event-related potentials are changes in voltage deflection at the scalp that correspond to that part of the spontaneous EEG which is a stereotyped electrophysiological response to a stimulus (for an introduction see Luck (2005)). ERPs have very good temporal resolution and provide a multidimensional measure which can be characterized along the following four dimensions: *polarity* (positive vs. negative), *latency* (the time involving onset, peak and duration at which the effect is visible), *amplitude* (the strength of the effect) and *topography* (the distribution over the scalp). In the neu-

---

\(^{10}\)But see the discussion in Brennan and Pykkänen (2008) (and to some degree in Dölling (2014) who remains sceptical about the relevance of experimental data for coercion theories) who took underspecification accounts to predict a two stage coercion procedure for aspectual coercion. In a first step, the default interpretation is computed, i.e. the coercion site is closed by applying the (default) identity function to the predication. This results in an aspectual mismatch, which is then repaired pragmatically in a second step by aspectual reanalysis. Of course, if this is how aspectual coercion is processed, we should assume the *Reject and Recompute Hypothesis* for underspecification theories, as well. As noted above, existing underspecification accounts lack a procedural implementation making it extremely hard to test them experimentally. In the reviewed experiments, we will therefore always assume that kind of procedural implementation of the theory which is consistent with our experimental findings. As we will see, the complete set of data nevertheless turns out to be problematic for underspecification accounts, even though we tried to be maximally generous with respect to the procedural implementation at a not yet worked out algorithmic level.
rolinguistic literature a number of ERP signatures have been identified that are relevant for language processing but there is still debate about the exact cognitive functions that underlies these components (for a review see e.g. Bornkessel-Schlesewsky and Schlesewsky (2009)). In the present experiment we used the ERP methodology to find out whether additive coercion and aspectual mismatch lead to the same brain signatures or are reflected by functionally different processes (see also Blaszczyk and KlimekJankowska (this volume)).

The experiment comprised three conditions. The additive coercion condition is illustrated in (19-a) while (19-b) involves aspectual mismatch. Both, coercion and mismatch, were compared to a control condition (19-c) which could be interpreted fully compositional. The three conditions only differed in the first word of the sentence. The critical word was the underlined past participle which was derived from an unambiguous achievement verb such as discover, find, reach, win and so forth. Any ERP effects must thus be due to (failed) composition of the achievement with the adverbial phrase.

(19) a. In zwei Stunden hatte der Förster die Falle entdeckt obwohl sie gut versteckt war. 'In two hours the ranger had discovered the trap although it was well hidden'

   b. Ganze zwei Stunden hatte der Förster die Falle entdeckt | . . . 'For two hours the ranger had discovered the trap . . .'

   c. Vor zwei Stunden hatte der Förster die Falle entdeckt | . . . 'Two hours ago the ranger had discovered the trap . . .'

2.1.1 Methods

24 right-handed, native German participants read 120 items (40 per condition) plus 180 fillers in a latin square design. The stimuli were presented visually in the center of the screen with a presentation rate of 500ms per word plus a 300ms inter-stimulus interval.
After each trial participants were asked for a sensicality judgment. Overall, 50% of all sentences in the experiment made sense assuming that the coercion condition should be judged sensible whereas the mismatch condition should be rejected.

The EEG-activity was measured with 30 Ag/AgCl-electrodes. To investigate the topographical characteristics of ERP effects, four topographical regions were defined: left anterior electrode sites (containing FP1, F7, F3, FC5 and FC1), right anterior sites (FP2, F4, F8, FC2 and FC6), left posterior sites (CP5, CP1, P7, P3 and O1) and right posterior sites (CP2, CP6, P4, P8 and O2). The mean amplitudes of these four regions were used to compute repeated measures ANOVAs including the factors anteriority, hemisphere and adverbial.

### 2.1.2 Results and discussion

Participants judged aspectual coercion as sensible in 75% of the cases compared to 87% in the control condition and 15% in the mismatch condition. Thus, although coercion was judged nonsensical more often than control, subjects chose a coerced interpretation in the majority of the cases.

![Figure 1 about here.](image)

Grand averages on the participle are shown in Figure 1. In the time window from 500 to 900 ms post onset the conditions started to differ. While mismatch revealed a positive ERP signature compared to both coercion and control, coercion led to an anterior negativity compared to mismatch and control. The positivity was strongest over the central and parietal midline electrode sites like it is typical for the P600.

Two ANOVAs with the within factors anteriority, hemisphere and aspect were computed. The first compared mismatch to control and the second coercion to control. In the former analysis there was both a significant interaction between aspect (mismatch vs. control) and anteriority \( (F(1, 23) = 7.62, p < .01) \) resulting from the

\[11\]Statistical analyses of earlier time windows didn’t reveal any significant differences between conditions.
more posterior positivity. The latter analysis also revealed both a significant interaction between aspect (coercion vs. control) and anteriority \( (F(1, 23) = 6.28, p < .05) \) due to the anterior negativity. No other effects involving the factor aspect were significant. Figure 2 shows the mean amplitudes of anterior and posterior electrode sites with respect to both hemispheres.

To further investigate the anterior negativity the time window from 900 ms to 1500 ms was analyzed. An ANOVA with the within factors anteriority, hemisphere and adverbial (coercion vs. control) resulted in a significant interaction between adverbial and anteriority \( (F(1, 23) = 8.70, p < .01) \). Besides a main effect of hemisphere no other effects were significant. This interaction is due to the negativity extending right to the following word which started 800 ms post onset. The mismatch condition was not analyzed in this ROI because of the preceding P600 effect.

The qualitatively different ERP effects in the coercion and the mismatch condition show that the processing of additive coercion is not triggered by a temporary mismatch. Instead, the findings of the ERP study provide evidence for a smooth update process in aspecual reanalysis. This pattern of results is fully compatible with the outlined EC analysis in section 1.1.1 where we assumed that additive coercion consists in adding an eventuality to the existing representation without revising it first. In order to coerce an achievement into an accomplishment the achievement scenario has to be updated with abductively inferred fluents \( f_1 \) and \( f_2 \) which encode a plausible preparatory process. The culmination and the resultant state are not affected by this operation but remain as they are.

If this analysis is on the right track, the anterior negativity would be due to an abductive inference of a preparatory process and enriching the situation model with it. Since the negativity emerged rather lately and persisted for at least a second it may be related to what has been referred to as working memory LAN in the neurolinguistic literature. This kind of negativity is generally taken to reflect increased working memory
demands (see e.g. Münte, Schiltz, and Kutas (1998)) and would therefore nicely fit the enrichment analysis outlined above. Whether the observed negativity actually is a kind of working memory LAN has, however, to be left for future research.

2.2 Immediate context effects in additive coercion (Experiment 2)

If the proposed pragmatic enrichment analysis of additive coercion is on the right track, we may expect coercion costs to disappear in case the preceding discourse context contains the required eventuality. This was tested in a self-paced reading study in which readers had to additively coerce activities into accomplishments. We paired telic (20-a) vs. atelic (20-b) contexts with target sentences with unambiguous activity verbs like jog which were modified by in- (20-c) or for-adverbials (20-d) according to a $2 \times 2$ within-design. The actual experiment was carried out in German, but for space limitations I will only provide the English translations of the materials (for a detailed description see Bott (2010, ch. 4/Exp. 3)). Vertical lines indicate segmentation of the target sentences for self-paced reading.

(20)  

a. Half a year ago, John started to jog about ten miles every day. When he began he was quite slow, but now he is really fast.

b. Half a year ago, John started jogging every day. When he began he had to stop after a short time, but now he can run for a long time.

c. Today | he jogged | in one and a half hours.

d. Today | he jogged | for one and a half hours.

Note that without supportive context, (20-c) is hardly interpretable. The activity jog does neither include an incremental theme $f_2$, nor a culmination $e$ or a resultant state $f_3$ and thus doesn’t fit the input requirements of an in-adverbial which modifies an accomplishment $\langle f_1, f_2, e, f_3 \rangle$.

The telic context (20-a) introduces a path argument (ten miles). From the context it is clear that whenever John engages in a jogging activity it stops when the contextually introduced endpoint has been reached. If the target sentence can be directly connected
to the discourse context it is expected that jog will be immediately interpreted as an accomplishment. Thus, a target sentence containing an in-adverbial should be as easy to interpret as a target sentence containing a for-adverbial. If, however, coercion is triggered by a type conflict during semantic composition, a local mismatch should occur which can only be repaired once the information from the preceding discourse context is integrated.\(^\text{12}\)

By contrast, the atelic context (20-b) doesn’t introduce a path argument, but establishes jogging as an atelic activity. Jog in the target sentence can only be interpreted as an activity and therefore easily combines with a for-adverbial. However, modification by an in-adverbial is impossible since there is no culminating event with which the adverbial can be unified.

### 2.2.1 Methods, results and discussion

32 native German participants read 24 items plus 64 filler discourses in a self-paced reading experiment with moving window presentation (see e.g. Haberlandt (1994)). Discourses were distributed over four lists according to a latin square design.

[Figure 3 about here.]

Figure 3 shows the mean reading times of the target sentences for each condition.\(^\text{13}\) Up to the adverbial, reading pace did not differ across conditions (all F's < 1). When readers encountered the adverbial phrase they slowed down in case it was aspectually inconsistent with the preceding context (telic for: 102.5 ms/char.; atelic in: 106.9 ms/char.) compared to the two aspectually matching conditions (atelic for: 86.9 ms/char.; telic in: 90.1 ms/char.). ANOVAs computed on reading times of the adver-

---

\(^{12}\)How early this is allowed to happen obviously depends on the particular theory of aspectual coercion and on the assumed relative time course of semantic and pragmatic processing. The point I would like to make is that, in order to account for immediate context effects, a cognitively plausible coercion theory should be formulated at the level of discourse representations.

\(^{13}\)An additional analysis of residual reading times instead of reading times per character revealed the same pattern. Again, there was a highly significant interaction between context and adverbial \((F_1(1,31) = 19.05, p < .01; F_2(1, 23) = 29.34, p < .01)\). A pairwise comparison between the telic in and the atelic for condition confirmed that there was no significant difference between the two conditions \((t_1(31) = 1.52, p = .14; t_2(23) = 1.18, p = .25)\).
bial revealed a significant interaction between context and adverbial ($F_1(1, 31) = 23.67, p < .01$; $F_2(1, 23) = 11.48, p < .01$). But neither the main effect of context nor the main effect of adverbial reached significance (all $Fs < 1.5$). A paired t-test revealed that the reading times of the adverbial in the telic in condition did not reliably differ from the reading times of the atelic for condition ($t_{1/2} < 1$).

This experiment shows that context information has an important influence on the interpretation of lexical aspect. The activity sentences modified by an in-adverbial are semantically ill-formed and require a fair amount of coercion when uttered out of the blue. However, if the context includes a path-argument which provides an upper bound to the activity, an in-adverbial can easily be combined with it. In fact, processing proceeded as smoothly as in the case of the atelic for condition. To properly deal with aspectual coercion and coercion costs the phenomenon should therefore be modeled at the discourse level.

3  A comparison between different types of aspectual coercion (Experiments 3a/b)

The first two experiments demonstrated that aspectual coercion of the additive kind is not triggered by a type mismatch during composition but consists in elaborating the situation model. In addition, we have demonstrated that this elaboration crucially depends on pragmatic information. In the remainder of the paper I will briefly review two further reading time experiments which directly compared different kinds of aspectual coercion, namely coercion types (7-c)–(7-e). Readers interested in the details of the experiments are referred to Bott (2008) and Bott (2010, Exp. 10a/b).

EC predicted a clear processing difference between iterative coercion of semelfactives as in (7-e) and accomplishments as in (7-d). Table 2 summarizes the required steps. Operator-based accounts and underspecification frameworks do not readily predict a difference in processing difficulty, though, since the same kind of operator has to be selected in the two iterative coercion cases.
Table 2 about here.

(21) a. Hans durchtauchte das Schwimmbecken dreißig Sekunden lang, dann wurde ihm schwarz vor Augen. 'Hans dived through the pool for thirty seconds before he fainted'

b. Hans durchtauchte das Schwimmbecken in dreißig Sekunden, dann wurde ihm schwarz vor Augen. 'Hans dived through the pool in thirty seconds . . .'

c. Hans durchtauchte das Schwimmbecken dreißig Jahre lang, dann wurde ihm das Tauchen vom Arzt verboten. 'Hans dived through the pool for thirty years before his doctor forbade him to continue'

d. Hans durchtauchte das Schwimmbecken in dreißig Jahren, dann wurde ihm das Tauchen vom Arzt verboten. 'Hans dived through the pool in thirty years . . .'

40 sets of German accomplishment sentences were constructed in four conditions like the sample item in (21). (21-a) requires subtractive coercion; the short in-adverbial in (21-b) fits the aspectual class of the accomplishment and serves as aspectual control; (21-c) has to be iteratively coerced into a habitual reading; and the long in-adverbial in (21-d) yields an implausible sentence. The latter was included to compare coercion with aspectual mismatch.

The experimental sentences were always followed by a then-clause. There were two types of then-clauses depending on the adverbial. Following short adverbials the second clause introduced a concrete event which can plausibly interrupt the preparation. Participants are thus expected to understand subtractive coercion in (21-a) as implying that the culmination did not happen. By contrast, in (21-b) the event introduced by then is expected to be interpreted as abutting the culmination. Following long adverbials the event was rather abstract and shifted the focus to higher order events. This was done to make it coherent with the habitual interpretation of the iterative accom-
plishment\textsuperscript{14}. Crucially, the two continuations only differed after the first three words of the second clause whereas the region of interest was at the end of the first clause. The short and long adverbials were matched for length.

In addition, 20 sets of semelfactive sentences were constructed in two conditions. (22-a) is like the English example (7-e) and requires iterative coercion. (22-b) is a non-coercing control. The experimental design was as close as possible to the one used in the study by Brennan and Pylkkänen (2008) on aspectual coercion in English. As mentioned above, Brennan and Pylkkänen observed coercion costs due to iterative coercion in iterated semelfactives both in self-paced reading and in magnetoencephalography.

(22) a. Vor wenigen Minuten \textit{nies}te |der Junge \textit{recht laut, dann} verließ er das \textit{Klassenzimmer.} Few minutes ago \textit{sneezed} \textit{the boy} \textit{rather loudly} then \textit{he left} \textit{the classroom.}\n
   'Few minutes ago, the boy sneezed rather loudly then he left the classroom’

b. Den ganzen Morgen \textit{nies}te |der Junge \textit{recht laut, …} The whole morning \textit{sneezed} \textit{the boy} \textit{rather loudly} …

   'The whole morning, the boy sneezed rather loudly …’

The materials were pretested with respect to the interpretations participants assigned to them. The initial part of the target sentences (accomplishment: \textit{the worker loaded the cart}; semelfactive: \textit{the boy sneezed loudly}) were presented to 20 German native speakers to find out whether the chosen sentences were clear instances of accomplishments and semelfactives, respectively. The accomplishments were interpreted as complete single events 89\% of the time. Only 10\% of all judgments were “no, not completed”. Without adverbial modification they did not allow for an activity interpretation. The semelfactives received 100\% “yes” responses. Of these, 98\% were single event judgments. This shows that the semelfactives, in fact, denote single events and do not have a process interpretation.

\textsuperscript{14}The time span expressed by the adverbial was chosen extraordinary long because I wanted to trigger clearly iterative interpretations. Currently, a study is under preparation which will test non-habitual, iterative accomplishments of the type (7-d) to see whether the effects to be reported below generalize to these, too.
We then pretested the effects of adverbial modification in a second rating study with 24 participants. For each sentence participants had to provide two judgments. First, they had to decide whether the sentence expressed that the (with accomplishments: culminating) event happened and if so, whether it happened only once or repeatedly. Accomplishments modified by short in-adverbials and long for-adverbials overwhelmingly received “yes, the culminating event happened” responses (short in: 77%; long for: 82%) but the short for-adverbials received “yes” responses only 16% of the time. This shows that the short for condition led to subtractive coercion.

The long for condition triggered the computation of iterative readings: 96% of the “yes” answers were “more than once”. In comparison, in the short in condition 95.1% of the “yes” answers were “only once”.

Semelfactives modified by ago-adverbials had “only once” judgments 94% of the time, whereas semelfactives modified by for were 95% of the time judged to express repeated events indicating the computation of iterative readings. To conclude, the comparison of the sentences without modification and the modified sentences shows that the tested constructions (21-a), (21-c) and (22-b) in fact exhibit the aspectual shifts we have proposed above.

3.1 Methods, results and discussion

40 native German participants read the 40 (Exp. 3a) + 20 (Exp. 3b) experimental items plus 90 fillers in a self-paced reading experiment with a sensicality judgment after each trial. Segmentation of the sentences is illustrated by vertical lines in (21) and (22).

[Figure 4 about here.]

The reading times in the four accomplishment conditions are shown in Figure 4. At the adverbial, short in was read fastest with a mean of 904ms, short for had a mean RT of 936ms, long for 1000ms and long in was read slowest with a mean RT of 1042ms. ANOVAs revealed a significant main effect of duration ($F_1(1, 39) = 18.35, p < .01; F_2(1, 39) = 14.52, p < .01$) which is due to the fact that the long
adverbials took longer to read than the short adverbials. The interaction between duration and adverbial was marginally significant in the subjects analysis but wasn’t reliable by items ($F_1(1,39) = 3.84, p = .06; F_2(1, 39) = 2.13, p = .15$). Also, there was no reliable main effect of adverbial ($F_{1/2} < 1$). Planned comparisons revealed that short for-adverbials didn’t reliably differ from control ($t_1(39) = 1.25, p = .22; t_2(39) = .96, p = .34$) but that long for-adverbials took longer to read than control ($t_1(39) = 3.44, p < .01; t_2(39) = 2.92, p < .01$). Long for-adverbials didn’t differ significantly from the implausible long in-adverbials ($t_1(39) = 1.73, p = .09; t_2(39) = 1.03, p = .31$).

At the following segment, the implausible long in condition had slower reading times (564ms) than the other conditions (short for: 485ms, short in: 472ms and long for: 509ms). This difference was reflected in a significant interaction between duration and adverbial ($F_1(1,39) = 8.66, p < .01; F_2(1, 39) = 9.16, p < .01$), a significant main effect of duration ($F_1(1,39) = 38.34, p < .01; F_2(1, 39) = 36.67, p < .01$) and a marginal effect of adverbial ($F_1(1,39) = 3.36, p = .08; F_2(1, 39) = 3.21, p = .08$). Paired t-tests revealed that long for was slower than short in ($t_1(39) = 2.58, p < .05; t_2(39) = 2.39, p < .05$) and long in was slower than short in ($t_1(39) = 5.56, p < .01; t_2(39) = 6.87, p < .01$). The numerical difference between short for and short in wasn’t reliable ($t_1(39) = 1.17, p = .25; t_2(39) = .92, p = .36$). At later segments there were no significant differences between conditions (all $F$s < 1).

The reading times of coerced and non-coerced semelfactives are shown in Figure 5. Paired t-tests revealed that the two conditions didn’t differ in reading time at any segment (all $t_{1/2} < .5$). At the critical verb region, coercion was numerically even read faster than control.

The findings provide evidence against operator-based accounts. Under this type of theory, aspectual mismatch is expected to occur in all three coercion types. However, neither subtractive coercion nor iterated semelfactives were harder to comprehend than
their aspectual controls. This shows that some kinds of aspectual coercion do not cause processing difficulty at all. The lack of effect cannot be due to aspectual ambiguity of the accomplishments and semelfactives because in the first pretest the unmodified accomplishments were judged to unambiguously express perfective telic events and the unmodified semelfactives were judged to denote single events. Our findings are in line with the experimental results of Pickering et al. (2006) who also didn’t find difficulty in iterated semelfactives either in self-paced reading or in eyetracking.

Unlike the data from Pickering et al. (2006), the results of the present experiments are not compatible with underspecification accounts. Iterated accomplishments immediately slowed down reading pace. This indicates that this coercion type is taxing. According to underspecification accounts, both iteration types should be resolved by plugging an iterative operator into the underspecified representation. Since specification is carried out on pragmatic grounds the processor should never consider the implausible subtractive coercion reading in accomplishments modified by long for-adverbials. Instead, it should select the iterative reading right from the start. It is thus completely unexpected why one iterative coercion type should be harder to process than the other.

Our experimental findings fully support the predictions of EC. Iterative accomplishments were much harder to coerce than subtractive coercion or iterative semelfactives. I proposed that difficulty in the first coercion type stems from first computing an implausible subtractive coercion reading then revising it and giving it an iterative interpretation instead. Of these three steps the first and the last do not cause difficulty when they appear in isolation as was shown by the subtractive coercion condition and the iterative semelfactive condition. Only when a coerced meaning has to be reanalyzed, is difficulty enhanced. The situation is similar to syntactic garden path sentences where readers also have great difficulty recovering from local misinterpretation.

A possible concern may be that no distinction was made between different kinds of iterative interpretations. In particular, iterative processes were equated with habitual, i.e. stative, interpretations. This may, however, be oversimplistic. It is possible that
the results reflect difficulty in processing habitual readings rather than revision of an implausible interpretation. I do not think that this alternative explanation is likely to be correct because, intuitively, habitual readings do not seem to be difficult in general (consider, for instance, *in the morning, John brushes his teeth*). Further research is needed to find out whether this intuition is correct.

Why did Brennan and Pylkkänen (2008) find iterative coercion of unambiguous semelfactives to be costly in contrast to the lack of effect in our Experiment 3b? A likely explanation can be given once we take into account a crucial grammatical difference between English and German (see also Blaszczak and Klimek-Jankowska (this volume) for their closely related blocking hypothesis). English has a grammatical progressive form but German has no grammatical means to mark the ongoingness of a situation. Bott and Hamm (2014) have provided experimental evidence that the competition between different forms is reflected by clear processing differences. In contrast to German, English subtractive coercion of an accomplishment in the simple past as in (23-a) causes a slow-down in reading. They showed that in order to express non-completedness of the event, the grammatical alternative (23-b) would be more appropriate.

(23) a. John built a house for two years.
    b. John was building a house for two years.

The same point can be made for iterative semelfactives\textsuperscript{15}. Again, there are two grammatical alternatives to describe an iterative semelfactive in English (c.f. (24-a) vs. (24-b)) but only one construction (24-c) in German. Hence, in English we may expect competition between these alternative forms which should result in a penalty for the simple form. Note that if this explanation turns out to be correct, it would call for a rather different interpretation of the aspectual coercion effect observed by Brennan and Pylkkänen (2008) who attributed it to difficulty during semantic composition.

\textsuperscript{15}This has been pointed out by Beatrice Primus (p.c.) to me.
4 Conclusions

This paper started with an overview over three broad classes of semantic or pragmatic theories on aspectual coercion. On the basis of experimental evidence from psycholinguistic experiments, we have argued that planning accounts such as EC (van Lambalgen & Hamm, 2005) fare considerably better than their competitors, operator-based accounts and underspecification theories.

The first piece of evidence was an ERP study in which we compared the processing of additive coercion achievement $\rightsquigarrow$ accomplishment with ERP correlates of aspectual mismatch. The experiment revealed a double dissociation between coercion and mismatch showing that the event representation can be smoothly updated without being initiated by temporary aspectual mismatch. A self-paced reading experiment investigated whether the preceding discourse context can eliminate coercion costs in case it contains the required eventuality. The results indicate that aspectual processing of an incoming sentence immediately accesses the discourse representation. We took this as evidence that aspectual coercion in fact is a discourse phenomenon and requires a pragmatic treatment. Finally, Experiments 3a/b demonstrated coercion costs to differ across sub-types of aspectual coercion. Most relevant for the purposes of this paper was the finding that the same coercion operation – iteration of an event – is open to large variation with respect to coercion costs. EC predicted iterative coercion of an accomplishment into a series of events reading to be much harder than iteration of a semelfactive. As far as I can see, neither an operator-based account nor an underspecification theory would have predicted this result.

An independent piece of evidence against aspectual underspecification comes from a self-paced reading experiment reported in Bott (2010, Exp. 2). In this experiment,
verbs that predominantly appear in activity sentences like *jog* were contrasted with verbs like *write* which are biased towards an accomplishment use. This was established by way of a corpus study. Nevertheless, both types of verbs can appear in activity sentences (25-a)/(25-c) as well as in accomplishment sentences (25-b)/(25-d) depending on, for instance, the semantic properties of the arguments (Krifka, 1992). The reading time analysis of the underlined adverbial region revealed an interaction between *verb type* and *adverbial*. Following an activity verb *in*-adverbials were read significantly slower than *for*-adverbials whereas following an accomplishment verb there was no reliable difference in reading time between the two adverbials. This suggests that readers immediately assigned an aspectual interpretation to the verbs as they were encountered, in opposition to what would be expected if aspectual properties were underspecified.

(25)  

|   | Peter joggte **eine Stunde lang** im Park.  
|---|Peter jogged **for an hour** in the park. |
| a. | Peter joggte **in einer Stunde** bis ans Ende des Parks.  
|   | Peter jogged **in an hour** to the end of the park. |
| b. | Peter schrieb **eine Stunde lang** an einem Brief.  
|   | Peter was writing **for an hour** at a **letter**. |
| c. | Peter schrieb **in einer Stunde** einen Brief.  
|   | Peter wrote **in an hour** a **letter**. |
| d. |

I would like to point out that the reported work can only serve as a starting point. Many issues remain unresolved within the EC framework. As, for instance, one of our reviewers pointed out, it remains completely mysterious why accomplishments allow *for*-modification of the resultant state illustrated in (26-a), whereas achievements do not (26-b). This is rather surprising given the fact that accomplishments and achievements share the relevant part of the event nucleus and have completely parallel scenarios. Thus, further work needs to be done to formulate a procedural theory that can fully account for the possible and impossible coercion readings that are intuitively accessible. This is, however, well beyond the scope of the present paper.
On the empirical side, we need to consider more coercion cases than the ones reviewed in this paper. One case in question is the repeatedly mentioned comparison between iterative interpretations and habitual interpretations. Here, a comparison between minimal pairs such as (8-a) versus (8-b) will be crucial. Another open question concerns the exact interdependencies between lexical aspectual class and typical event duration. A plausible theoretical alternative to the account advocated here is that coercion costs in iterative coercion may depend to a large degree on how variable the typical event duration of an event predicate is. For instance, when processing the construction worker loaded the wheelbarrow for three hours coercion difficulty may - at least in part – be caused by insufficient information about the typical duration of a single load event. In particular, the semantic processor may face a decision problem which coercion type is the most plausible and, in case of iterative coercion, how to partition the given interval of three hours into plausible subintervals. This decision problem may be absent in the boy sneezed for three hours where the typical duration of a single sneeze event is much more constrained by our world knowledge (see also the discussion in Deo and Piñango (2011) and Champollion (2013)). Obviously, more research is required to properly control for effects of coercion type and event duration, respectively.\textsuperscript{16}

To conclude, I hope to have shown that processing data are a rich resource which can help to decide between alternative semantic and pragmatic proposals. It is only recently that semanticists have begun to consider processing data in addition to intuitive judgments. The young and rapidly developing fields of experimental semantics and pragmatics, however, demonstrate how valuable experimental evidence can be and that processing data can impose important constraints on semantic and pragmatic theories (see among others Jackendoff (2002), McMillan, Clark, Moore, Devita, and Grossman (2005), Pietroski, Lidz, Hunter, and Halberda (2009), Pylkkänen et al. (2011), Bag-

\textsuperscript{16}I would like to thank Johannes Dölling (personal communication) for pointing out to me these factors and their potential importance for processing.
gio, van Lambalgen, and Hagoort (2012), Hackl, Koster-Hale, and Varvoutis (2012) and Noveck and Sperber (2012). Psycho- and neurolinguistics, on the other hand, depend on semantic and pragmatic theories that are spelled out in such way that they can be easily linked to language processing. Theoretical semantics and pragmatics is a prerequisite for processing theories at the algorithmic or even implementational level (Marr, 1982) because what we need first and foremost is a proper understanding of the computational level and the general complexity of the computations involved.

References


Luck, S. (2005). *An introduction to the event-related potentials technique*. Cambridge,
MA: MIT Press.


Malden, MA: Blackwell.


## List of Figures

1. Grand averages of additive coercion, aspectual mismatch and aspectual control from the onset of the participle. 38
2. Mean amplitude for the three aspectual conditions 500 - 900 ms post stimulus with respect to anteriority and hemisphere (+ standard errors of the mean computed per participant). 39
3. Mean reading times in Experiment 2 (confidence intervals computed per participant). 40
4. Mean reading times in Experiment 3a (confidence intervals computed per participants). Panel (a) mean reading times across sentence regions; panel (b) mean reading times of the adverbial region. 41
5. Mean reading times in Experiment 3b (confidence intervals computed per participants). 42
Figure 1: Grand averages of additive coercion, aspectual mismatch and aspectual control from the onset of the participle.
Figure 2: Mean amplitude for the three aspectual conditions 500 - 900 ms post stimulus with respect to anteriority and hemisphere (+ standard errors of the mean computed per participant).
Figure 3: Mean reading times in Experiment 2 (confidence intervals computed per participant).
Figure 4: Mean reading times in Experiment 3a (confidence intervals computed per participants). Panel (a) mean reading times across sentence regions; panel (b) mean reading times of the adverbial region.
Figure 5: Mean reading times in Experiment 3b (confidence intervals computed per participants).
List of Tables

1  Coercion types investigated in the present paper . . . . . . . . . . 44
2  Predicted processing steps for iterative semelfactives and iterative ac-
    complishments by EC. . . . . . . . . . . . . . . . . . . . . . . . 45
Example | original asp. class | asp. class required by adverbial | coercion type
--- | --- | --- | ---
(7-a) | \((−, −, e, f_3)\) | \((f_1, f_2, e, f_3)\) | additive coercion
(7-b) | \((f_1, −, −, −)\) | \((f_1, f_2, e, f_3)\) | additive coercion
(7-c) | \((f_1, f_2, e, f_3)\) | \((f_1, −, −, −)\) | subtractive coercion
(7-d) | \((f_1, f_2, e, f_3)\) | \((f_1, −, −, −)\) | iterative coercion
(7-e) | \((−, −, e, −)\) | \((f_1, −, −, −)\) | iterative coercion

Table 1: Coercion types investigated in the present paper
Table 2: Predicted processing steps for iterative semelfactives and iterative accomplishments by EC.

<table>
<thead>
<tr>
<th>Semelfactive modified by <em>for</em></th>
<th>Accomplishment modified by <em>for</em></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>step 1)</em> iterative coercion</td>
<td><em>step 1)</em> subtractive coercion</td>
</tr>
<tr>
<td></td>
<td><em>step 2)</em> plausibility check triggers reanalysis</td>
</tr>
<tr>
<td></td>
<td><em>step 3)</em> iterative coercion</td>
</tr>
</tbody>
</table>