

Chapter 8

The Processing Domain of Aspectual Interpretation

Oliver Bott

8.1 Introduction

In theories of language processing it is commonly assumed that interpretation proceeds incrementally, that is on a word by word basis. An open question is whether this holds for aspectual semantic processing and for semantic processing in general as well. Crocker (1996, p. 251) formulated the principle of incrementality (the psycholinguistic perspective on syntactic processing) in the following way:

“The sentence processor operates in such a way as to maximize the interpretation and comprehension of the sentence at each stage of processing (i.e., as each lexical item is encountered).”

By contrast, in semantic theory lexical aspect is often treated as a property of whole VPs or even whole sentences. This is what I call the semantic perspective (Dowty 1979, p. 62):

“Not just verbs, but in fact whole verb phrases must be taken into account to distinguish activities from accomplishments. (In a certain sense, even whole sentences are involved. . .)”

According to the semantic perspective we should expect that a transitive verb on its own has no lexical aspect until it is composed with (at least) its internal argument. As a consequence, effects due to aspectual violations can only arise when the verb has received all or at least some of its arguments. Using an analogy from chemistry, the event type can thus be viewed as an atomic property which supervenes on the properties of its constituents. Building on this analogy, a lexical aspectual class is a higher order concept similar to the concept of a noble gas in chemistry. It should be clear right from the start that investigating the domain size of lexical aspect

O. Bott (✉)

Sonderforschungsbereich 833, University of Tübingen, Nauklerstr. 35,
Tübingen 72074, Germany
e-mail: oliver.bott@uni-tuebingen.de

is independent of investigating the interplay of the verb and its arguments as for example on the thematic level (see eg. [Ferretti et al. \(2007\)](#), [Malaia et al. \(this volume\)](#)) for very early aspectual effects in this respect).

Consider the examples in (1-b) to (1-d) which are all legal word order variants of (1-a). Note that *erreichen* (*reach*) is an unambiguous transitive German achievement verb. Like accomplishments, achievements are telic, but they express an instantaneous change of state and therefore lack a preparatory process (cf. [Moens and Steedman \(1988\)](#)). This explains why they don't allow modification by a *for*-adverbial rendering all three word order variants ungrammatical¹ whereas accomplishments can be coerced into an activity reading (see eg. [Bott \(2010\)](#)).

- (1) a. *Der Bergsteiger erreichte den Gipfel zwei Stunden lang.
The mountaineer_{nom} reached the summit_{acc} two hours long.
*The mountaineer reached the summit for two hours.
- b. *Den Gipfel erreichte zwei Stunden lang ein Bergsteiger.
The summit_{acc} reached two hours long a mountaineer_{nom}.
*The mountaineer reached the summit for two hours.
- c. *Der Bergsteiger erreichte zwei Stunden lang den Gipfel.
The mountaineer_{nom} reached two hours long the summit_{acc}.
*The mountaineer reached the summit *for two hours.
- d. *Zwei Stunden lang erreichte der Bergsteiger den Gipfel.
Two hours long reached the mountaineer_{nom} the summit_{acc}.
*The mountaineer reached the summit for two hours.

What makes the three examples interesting is the point at which the aspectually mismatching information comes into play: In (1-a) the verb-argument structure is complete when the adverbial enters the sentence. In (1-b) the verb has already received the direct object, but the subject is still missing. This means we are dealing with a complete VP. Finally, in (1-c) the VP is actually not complete yet. At this point, the adverbial has to modify the bare verb. The same point is exemplified even more clearly in (1-d) where both the subject and the object enter the sentence only after the *for*-adverbial and the achievement verb.

Whereas the aspectually mismatching adverbial in (1-a)–(1-d) leads to a nonsensical sentence, cases of so called coercion provide examples where an aspectual mismatch emerges only locally and can somehow be repaired (see [Moens and Steedman \(1988\)](#) for a systematic overview over different kinds of coercion). Consider the following example.

- (2) Der Bergsteiger erreichte den Gipfel in drei Tagen.
The mountaineer reached the summit in three days.

¹It may be objected that the examples become sensical when coerced into an iteration of reaching events. Acceptability ratings for examples like these, however, indicate that naive informants are not aware of this possibility ([Bott, 2010](#)). On the basis of these findings I make a categorical distinction between cases of aspectual mismatch and coercion.

(2) is an instance of what Hamm and van Lambalgen (2005) called additive coercion. *Erreichen* (*reach*) is an unambiguous achievement verb which introduces a culmination (*the mountaineer reaches the summit*) and a consequent state (*the climber now being on top*). When the achievement is, however, combined with an *in*-adverbial, it has to be coerced into an accomplishment. For this, new semantic structure (a preparatory phase) has to be added to the aspectual representation. In our example world knowledge probably suggests that this was a *climbing* activity. But, the mountaineer also could have reached the top using a helicopter. This demonstrates that additive coercion requires an abductive inference to what preparatory process may have lead to the culmination event.

The present paper investigates whether an aspectual violation can be detected immediately at the mismatching adverbial irrespective of its structural position in the sentence. Since aspectual coercion may require more contextual information than mere mismatch detection, even more deferred processing may be expected in coercion than in mismatch cases. The time course of aspectual violation and reanalysis were investigated with word order variants of German transitive achievement verbs. These were modified by mismatching or coercing adverbial phrases and their processing was compared to an aspectual control condition using aspectually matching adverbials.

8.1.1 Previous Studies on Aspectual Coercion

It may be worth looking at the existing studies on the processing of lexical aspect. Without exception, all of them focus on aspectual coercion and none compares coercion effects to effects of aspectual mismatch. Moreover, as things stand, it is still an open question whether aspectual coercion leads to processing difficulty at all.² A reason for this somewhat unsatisfactory situation may be that the research has almost exclusively limited itself to one type of aspectual coercion, ie. the iteration of point action verbs. Furthermore, all existing studies on aspectual coercion used English materials. Because English has fixed word order it cannot be used to systematically investigate the processing of lexical aspect at various hierarchical levels. For instance, to test the VP as processing domain, the most natural choice is to use a transitive verb in a sentence with object before subject word order where the mismatching or coercing stimulus intervenes between the VP (= verb + direct object) and the subject. Unfortunately, this word order is ungrammatical in English. Thus, a language with relatively free word order like German is needed where all four construction types in (1-a)–(1-d) are grammatical.

²For complement coercion, things look different. Numerous studies have demonstrated that this type of coercion clearly enhances processing load (for an overview see eg. Pykkänen and McElree (2006)).

Not surprisingly, the processing domain of lexical aspect has not been explicitly mentioned in the psycholinguistic literature. Let's have a look at the materials used in these studies to see if there is any implicit evidence concerning the issue. The following examples present sample materials from the first studies reporting a coercion effect:

- (3) a. The insect glided effortlessly until ...
b. The insect hopped effortlessly until ...
- (4) a. Howard sent a large check to his daughter for many years ...
b. Howard sent a large check to his daughter last year ...

Sentences like (3-a) vs. (3-b) were used in the cross modal lexical decision studies by Piñango et al. (1999) and Piñango et al. (2006). The coercing adverbial (*until ...*) only appeared after a minimal sentence was complete. Similarly, the materials in (4-a) vs. (4-b) used in a stops-makes-sense-judgment experiment by Todorova et al. (2000) only reveal a coercion effect after a complete verb-argument structure had been presented.

To complicate matters, Pickering et al. (2006) used the same materials as in the experiments mentioned above, but tested a coerced meaning during ordinary reading without an additional task. In two self-paced reading and two eyetracking experiments, they found aspectual coercion to be no more difficult than their aspectual control conditions. This lack of effect led them to propose the aspectual underspecification hypothesis. This hypothesis states that the aspectual representation stays underspecified during normal reading. Brennan and Pykkänen (2008) challenged this view and reported a coercion effect of coercion sentences like (5-a) as compared to aspectual controls (5-b) both in self-paced reading (but see Bott (2010) for different findings in German) and in MEG. On the basis of a rating study they had carefully selected clear instances of point action verbs. However, the specific processing of aspectual coercion could be performed earliest at the verb, that is after readers were already dealing with a complete sentence.

- (5) a. Throughout the day, the student sneezed in the back of the classroom.
b. After twenty minutes, the student sneezed in the back of the classroom.

To sum up, all online effects that have been reported were measured rather late downstream of the sentence. The existing studies, therefore, do not let us decide between the incremental aspectual interpretation hypothesis in (6) and the late aspectual interpretation hypothesis in (7).

- (6) Incremental Aspectual Interpretation Hypothesis (IAIH)
Lexical aspect is computed incrementally, on a word-by-word basis.
- (7) Late Aspectual Interpretation Hypothesis (LAIH)
Lexical aspect is not computed before the verb has all its arguments.

The IAIH and its counterpart, the LAIH, are the two extremes with respect to incrementality. To be maximally clear, the LAIH is not intended to imply that aspectual processing is delayed until the comprehender crosses a sentence boundary, but only to depend on the verb plus all its arguments. That is, even under the LAIH we may expect to find effects of aspectual processing well before the end of the sentence. Arguments in favor of this hypothesis have, for instance, been provided by Verkuyl (1993) showing that both the internal (= undergoer) and the external argument (= agent) can lead to a change in aspectual class. Certainly, there is also an intermediate alternative to the IAIH and the LAIH. Not the complete verb-argument complex, but only the verb and its internal argument (the VP) may constitute the processing domain of lexical aspect. We will come back to the intermediate hypothesis later in this paper. We conducted a series of reading time experiments to determine the processing domain of aspectual interpretation.

8.1.2 *The Constructions Used in the Experiments*

The following experiments tested German transitive achievement verbs which were modified by three types of temporal adverbials. Here is a sample item in *subject-verb-object-adverbial* (SVOA) word order.

- (8) a. *Der Rentner fand den Schlüssel zwei Stunden lang in der Schublade.
 The pensioner found the key two hours long in the drawer.
 *For two hours, the pensioner found the key in the drawer.
- b. Der Rentner fand den Schlüssel in zwei Stunden in der Schublade.
 The pensioner found the key in two hours in the drawer.
 In two hours, the pensioner found the key in the drawer.
- c. Der Rentner fand den Schlüssel vor zwei Stunden in der Schublade.
 The pensioner found the key ago two hours in the drawer.
 Two hours ago, the pensioner found the key in the drawer.

Sentence (8-a) illustrates aspectual mismatch. The durative adverbial *for two hours* cannot modify the achievement denoting a punctual event. (8-b) exemplifies additive coercion (Hamm and van Lambalgen, 2005; Bott, 2010). Although the *in*-adverbial requires an accomplishment – one of the classic tests by Vendler (1957) – the sentence doesn't feel ill-formed. Obviously, comprehenders are able to infer the right kind of preparation (eg. *searching*) and implicitly shift the achievement into an accomplishment. (8-c) serves as control since the input requirements of the *ago*-adverbial perfectly match the achievement: composition yields a punctual event that is located 2 h before utterance time. We constructed 30 items in three conditions like (8-a)–(8-c). This set of experimental items was used in all of the following experiments except for the eyetracking study (Experiment. 4). The complete list of experimental sentences can be found in Bott (2010, Experiment. 4a/b and 8).

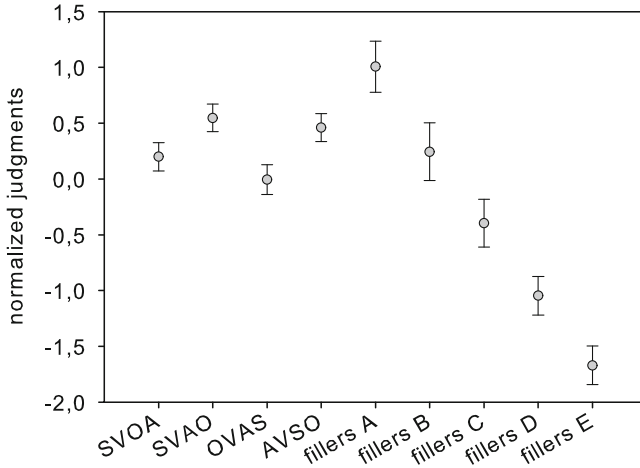


Fig. 8.1 Mean grammaticality judgments (+ 95% CI intervals) for the control condition in the four word orders. Also shown are mean judgments of five categories of normed filler sentences ranging from *perceived natural* (cat. A) to *strongly marked* (cat. E)

To test the incrementality of aspectual interpretation we manipulated the word order in these sentences. Besides SVOA sentences, we changed the position of the direct objects to SVAO word order. Furthermore, we manipulated the position of the subject yielding OVAS sentences. Finally, we constructed AVSO sentences in which the adverbial directly precedes the verb and the arguments come in only later.

German word order is relatively free, although not entirely free. Some word order variants may clearly be more marked than others. To compare aspectual processing among different syntactic configurations it is thus crucial that the constructions under study do not differ in grammaticality. For this purpose, we gathered judgments for all four word order variants using the thermometer judgement method (Featherston, 2008). The following orders were tested: SVOA, SVAO, OVAS and AVSO. All sentences were semantically well formed and used a transitive achievement modified by a *ago*-adverbial, ie. the control condition in the online studies. To find out if all four constructions are perceived as fully grammatical 20 normed distractors of five different levels of grammaticality were included. These were chosen from a pool of German example sentences which have been repeatedly tested in grammaticality surveys (see Featherston (2008)). Figure 8.1 depicts the mean judgments from 20 German native speakers. Two of the word orders, SVAO and AVSO, in which the adverbial preceded some of the arguments were rated even better than the canonical SVOA condition. Object topicalized sentences in the OVAS condition were rated slightly worse than the canonical SVOA sentences, but were still in the range of fully grammatical sentences. To compensate for this difference, the OVAS construction will be tested in an experiment (Experiment 3) that exclusively uses object initial sentences in the items and in the fillers.

8.2 Experiment 1: Providing a Continuation

Investigating adverbial modification in yet incomplete verb-argument structures raises an important question. Do readers automatically predict an argument that yields the *aktionsart* which is required by the adverbial? Consider (9-a) with the two continuations in (9-b) and (9-c).

- (9) a. Der Bergsteiger erreichte zwei Stunden lang ...
 The mountaineer reached two hours long ...
 For two hours, the mountaineer reached ...
 b. *den Gipfel. (*the top)
 c. niemanden am Telefon. (nobody on the phone)

As (9-c) shows, (9-a) can be continued in a meaningful way, although the most typical continuation of a yet incomplete achievement in (9-b) yields a semantically ill-formed sentence. When the processor encounters the sentence fragment in (9-a), it will predict material that is yet to come (see eg. Altmann and Kamide (1999)). Let's assume that the IAIH is correct. Then the predictive capabilities of the parser are absolutely crucial and lead to different expectations about when processing difficulty emerges in sentences like (9-a) with the semantically anomalous continuation (9-b).

Let's assume that aspectual processing is incremental and the complete range of possible arguments is considered by the sentence processor. It will then interpret the incomplete sentence in (9-a) with the expectation of a continuation like (9-c). As a result, including the adverbial, the sentence fragment is predicted to be well formed. Only when a continuation like (9-b) is encountered, is the expectation disconfirmed and processing difficulty emerges. Thus, we would expect delayed processing precisely because of incremental interpretation with an extremely high predictive power, that is able to predict a specific continuation (including negation, bare plurals etc.) making the sentence well-formed.

A theoretical alternative is that the processor expects a continuation that is highly associated with the lexical material encountered so far, but there is no "deep" analysis of it. Interestingly, although this second alternative requires less predictive power than the first option, it predicts earlier difficulty in aspectual processing. In the context of *the mountaineer reached* something like *the top* is expected. The predicted object is semantically incongruous with the *for*-adverbial. Thus, difficulty is expected immediately at the adverbial even before the object is encountered.

Before coming to the online experiments, we have to decide between these two alternatives. For this purpose, we measured the interpretation of incomplete sentences like (9-a) by asking comprehenders for a continuation.

8.2.1 Method

The present experiment was a production experiment with no time pressure. This ensured that participants had the opportunity to find the most sensible continuation. If in an offline task like this they are not able to come up with arguments that change lexical aspect to fit the requirements of an otherwise mismatching adverbial, it is even less likely that they are able to do so during real-time comprehension.

8.2.1.1 Materials

The same thirty items that were used in the online experiments were tested in the aspectual mismatch condition: an achievement combined with a *for*-adverbial. The ends of the experimental sentences were eliminated. This yielded the conditions in (10-a)–(10-c).

- (10) a. Der/Die Bergsteiger_{nom.} erreichte/n_{sing./pl.} zwei Stunden lang...
 The mountaineer(s) reached two hours long...
 b. Den Gipfel_{acc.} erreichte/n_{sing./pl.} zwei Stunden lang...
 The top reached for two hours...
 c. Zwei Stunden lang erreichte/n_{sing./pl.}...
 For two hours reached..

Example (10-a) contains the (in the singular case disambiguated) subject *der/die Bergsteiger*, an unambiguously transitive achievement verb and a *for*-adverbial, but the object is still missing. In (10-b) the case-disambiguated object *den Gipfel* is realized preverbally in topicalized position, but the sentence lacks a subject. In (10-c) the bare verb is tested with the adverbial. In this condition readers have maximal freedom in choosing the appropriate arguments to satisfy the input requirements of the adverbial.

If aspectual processing is highly predictive as outlined above, the number information of the verb might provide an important cue to what is yet to come. The typical examples proving an aspectual semantic influence of the arguments involve cases with bare plurals (eg. *visitors/*a visitor arrived all night*). When encountering a plural verb it might be that aspectual processing automatically predicts a bare plural subject. To test this, the number of the verb (*singular vs. plural*) was manipulated yielding a total of six conditions employing a 3×2 (*word order* \times *number*) factorial design.

Additionally, 40 distractors were included in the experiment. Thirty of them allowed for a sensible continuation while ten clearly did not. The latter contained tense violations like *morgen kam ... (tomorrow came...)* and aspectual violations of a different sort such as *Hans war gerade dabei intelligent zu sein, als ... (Hans was being intelligent, when ...)*. The experimental items and the filler sentences were arranged in six lists in a latin square design.

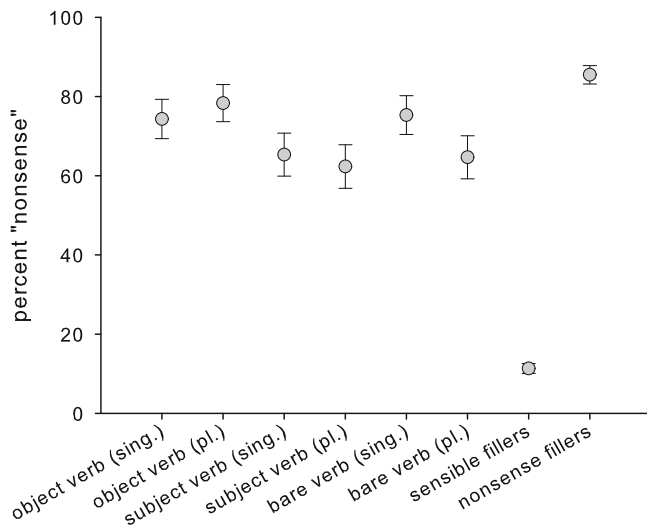


Fig. 8.2 Percent “nonsense” answers in Experiment 1 (+ 95% CIs)

8.2.1.2 Procedure and Participants

The experiment employed a combined acceptability rating/sentence completion task. Participants were asked to come up with a meaningful completion of the sentence. If they were not able to do so they were prompted to reject the sentence as nonsensical.

Sixty German native speakers (23 female; mean age 29.4 years) took part in the experiment. Among them, six prizes of 50€ were distributed by lot. Participants were randomly assigned to lists (ten participants per list). An experimental session took approximately 30 min.

For purposes of quantitative analysis, the percent of “nonsense” ratings were computed. In addition to “nonsense” button presses, all continuations that yielded sentences which were clearly not sensible or incomplete were also counted as “nonsense”. This affected 13.5% of the trials with experimental items. A qualitative analysis of the provided continuations can be found in [Bott \(2010\)](#).

8.2.2 Results

Figure 8.2 depicts the percent of “nonsense” answers for the experimental items and the distractors. The performance on the fillers shows that participants had understood the task and provided a completion if this was possible.

The experimental items were overwhelmingly rejected as nonsensical with a mean of 70.1% nonsense answers. There were, however, differences among the conditions.

First of all, participants provided more sensible completions when they had to choose an object (63.8% “nonsense”) than when the subject was missing (76.3% “nonsense”). In repeated measures ANOVAs this difference revealed a significant main effect of *verb argument structure* ($F_1(2, 118) = 10.70; p < .01; F_2(2, 58) = 4.74; p < .05$). ANOVAs which just compared the *missing object* and the *missing subject* conditions yielded a significant main effect of *word order* ($F_1(1, 59) = 22.74; p < .01; F_2(1, 29) = 7.12; p < .05$), but neither a reliable effect of *number* nor an interaction between *word order* and *number* (all $F_s < 2$).

Secondly, the interaction between *word order* and *number* was significant ($F_1(2, 118) = 5.46; p < .05; F_2(2, 58) = 4.57; p < .05$). The interaction was due to the *bare verb* conditions receiving more completions when the verb was plural than when it was singular ($t_1(59) = 4.28; p < .01; t_2(29) = 2.95; p < .01$), but the *missing object* and *missing subject* condition not showing a number effect. The main effect of *number* was not reliable ($F_1(1, 59) = 3.64; p = .06; F_2(1, 29) = 2.02; p = .17$).

8.2.3 Discussion

This experiment investigated whether readers can predict forthcoming arguments that shift the lexical aspect of a yet incomplete verb-argument structure in accordance with the input requirements of an aspectually mismatching adverbial. The findings clearly indicate that this is not the case. The initial part of sentences containing an achievement which is modified by a *for*-adverbial were overwhelmingly judged as nonsensical. This shows that readers just predict lexical material on an associative basis without deep aspectual analysis. As it seems, comprehenders aren't able to make use of the full set of combinatorial possibilities but rely on superficial lexical associations.

Nevertheless, the predictive capabilities depend on the parts of the verb-argument structure that have been encountered. Participants were able to come up with a sensible continuation more easily when the object than when the subject was missing. Although both, the internal and the external argument, matter with respect to lexical aspect, the internal argument seems to be more accessible than the external argument.

Interestingly, the number information of the verb did not have a big influence on the ability to predict material that is yet to come. In the *missing subject* conditions, participants were as likely to provide a sensible continuation when the verb had plural morphology as when it was singular. Thus, even with supportive morphological information there was no evidence of predicting the right kind of argument. Beyond the purposes of the present experiment this is an interesting finding since it demonstrates clear limitations of predictive processing.

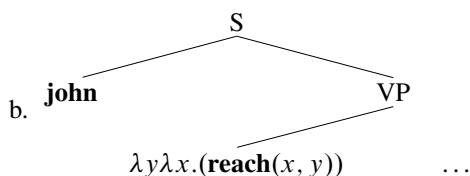
There was no time pressure to provide a completion. During ordinary reading, however, the processor is forced to decide much faster on the interpretation of the incoming material. Thus, if readers were not able to predict the right kinds of arguments in an offline task like the one employed here it is even less likely that the processor will engage in highly predictive parsing during ordinary comprehension. Assuming incremental aspectual parsing along the lines of the IAIH, readers can therefore be expected to stumble across mismatching aspectual information as soon as they encounter it.

8.3 Experiment 2: Complete Sentences Versus Extraposed Objects

Can an aspectually mismatching or coercing adverbial be immediately combined with a verb before the complete VP has been processed? The present experiment investigated this hypothesis by measuring reading times at adverbials that either matched the lexical aspect of achievement verbs, called for additive coercion or were aspectually mismatching. Processing was studied in SVOA and in SVAO sentences. In the latter the adverbial appeared at a point where the direct object of the unambiguously transitive verbs was still missing.

Under standard assumptions about the way compositional interpretation of the sentence works the subject cannot be combined with a transitive verb before the direct object is present (Heim and Kratzer, 1998).³ Consider the first words of a simple sentence in (11-a) with the simplistic semantic representation in (11-b).

(11) a. John reaches ...



Functional application of the subject node and the verb node is not possible before the VP node is semantically determined. But this depends on the object. As a result, composition has to wait until the object is present. This illustrates that common semantic practice cannot be easily brought together with incremental interpretation. Finding an immediate effect in the SVAO sentences would thus be particularly interesting when it comes to developing a cognitively realistic semantics.

³However, Heim and Kratzer (1998) actually argue for *top-down* interpretation which is somewhat different from the *bottom-up* approach chosen here. What is crucial here is that *top-down* interpretation also requires a complete sentence to compute a meaning for it.

8.3.1 Method

8.3.1.1 Materials

The experiment used a 3×2 factorial design with the factors *adverbial* (three levels: *control* vs. *additive coercion* vs. *mismatch*) and *word order* (two levels: *SVOA* vs. *SVAO*). The conditions are illustrated in the sample item in (12). Vertical lines indicate segmentation. (12-a) has SVOA word order, whereas (12-b) is SVAO. Instead of an *ago*-adverbial, the aspectual mismatch conditions had a *for*-adverbial (always of the type *x Zeit lang*, eg. *zehn Minuten lang*) and the additive coercion conditions had an *in*-adverbial (always of the type *in x Zeit*, eg. *in zehn Minuten*).

- (12) a. Der Förster_{subj.} | entdeckte | die Falle_{obj.} | vor (in/ganze) zehn
 The ranger | spotted | the trap | ago (in/for) ten
 Minuten | im | Wald.
 minutes | in-the | forest.
- b. Der Förster_{subj.} | entdeckte | vor (in/ganze) zehn Minuten | im |
 The ranger | spotted | ago (in/for) ten minutes | in-the |
 Wald | die Falle_{obj.} | für | Bären.
 forest | the trap | for | bears.

In the SVOA order, the adverbial was presented in one segment. It was always followed by a PP which was split up into two regions. These served as spillover regions and were included to see whether mismatch and coercion effects showed up before the end of the sentence. For statistical analysis, reading times were aggregated over the last two segments.

In the SVAO word order, the 30 items were constructed with two spillover regions. The adverbial was followed by a prepositional phrase which was divided into two regions, the preposition and the rest of the PP which was followed by the direct object. An effect of *aspect* at the direct object region is thus very unlikely to be a spillover effect from the adverbial region. Following the object, another PP was included as second spillover region. Like the first PP, it was divided into two segments. It was always attached to the object to make the noun phrase heavier and thus more natural in extraposed position. Statistical analyses used reading times that were aggregated over the two PP segments.

Additionally, 75 filler sentences were included in the experiment. They encompassed all kinds of aspectual classes and 25 of them were semantically ill-formed resulting in a overall ratio of 2:1 of well-formed to ill-formed sentences. The experimental items and the distractors were distributed over six lists in a latin square design. For each participant this yielded five data points per condition.

8.3.1.2 Procedure

The experiment was a self-paced reading study using moving window presentation. Each sentence was followed by a question. In the experimental items and half of the

fillers, questions queried whether the sentence made sense. To prevent participants from anticipating this kind of question, the other half of the filler sentences were followed by ordinary comprehension questions. Questions had to be answered with a time limit of 3 s.

The experiment started with written instructions. Then followed a practice session with ten trials. The practice items contained no aspectual violations. After the practice session the experiment followed in one block with an individually randomized order of sentences. An experimental session took about 20 min.

8.3.1.3 Participants

Thirty students from Tübingen University (all native German speakers, 24 female, mean age = 22.9 years) participated in the experiment. Each subject was paid 5€ for participation. The participants were randomly assigned to lists (five subjects per list).

8.3.1.4 Data Analysis

Reading times longer than 2,500 ms were trimmed to correct for outliers. This affected less than 0.5% of the data. Performance on the comprehension questions revealed that participants read attentively. Each of them answered more than 75% of the questions correctly.

8.3.2 Results

8.3.2.1 “Makes Sense” Judgements

The mean judgments are depicted in Fig. 8.3. In the SVOA conditions, participants judged the control condition as sensible in 93.2%, additive coercion in 48.5% and mismatch in 16.0%. The pattern was similar in the SVAO conditions. Aspectual control was judged sensible in 90.7%, additive coercion in 63.3% and aspectual mismatch in 29.9%. In ANOVAs, this difference led to a significant main effect of *adverbial* ($F_1(2, 58) = 169.70; p < .01; F_2(2, 58) = 98.64; p < .01$), a significant main effect of *word order* ($F_1(1, 29) = 8.91; p < .01; F_2(1, 29) = 12.58; p < .01$) and a significant interaction between the *adverbial* and *word order* ($F_1(2, 58) = 4.91; p < .05; F_2(2, 58) = 3.48; p < .01$).

Although the patterns are similar, mismatch detection was better in the SVOA mismatch condition than in the SVAO mismatch condition. Also, additive coercion was judged acceptable less often in SVOA sentences than in SVAO sentences. However, a direct comparison between the judgment results of the two word order variants is difficult, because the experimental items in the two conditions differed in

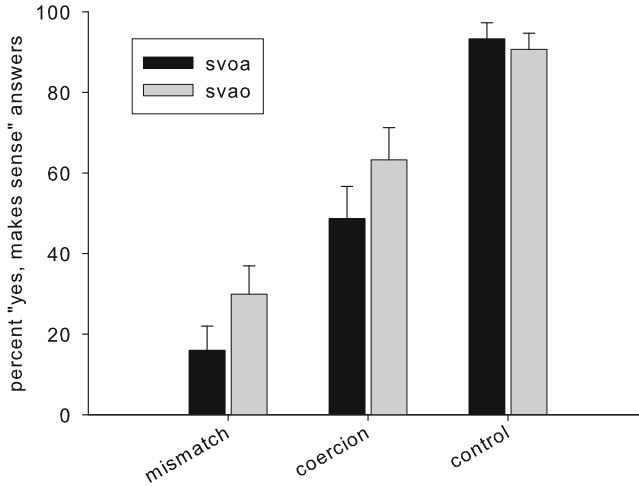


Fig. 8.3 Makes sense judgments of SVOA sentences in *black* and of SVAO sentences in *grey* (+95% CIs) in Experiment 2

length and furthermore the items in the SVAO conditions involved an additional PP which the items in the SVOA conditions did not.

Mean judgment times ranged between 1,400 and 1,580 ms, but there were no systematic differences between the conditions. Accordingly, ANOVAs analyzing judgment times did reveal neither significant main effects of *adverbial* or *order* (all $F_{1/2} < 1$) nor a significant interaction between them ($F_1(2, 58) = 2.11$; $p = .11$; $F_2(2, 58) = 1.76$; $p = .18$).

8.3.2.2 Reading Times: SVOA Word Order

Figure 8.4 shows mean reading times of sentences involving coercion and mismatch compared to control for the whole sentence.

Up to the adverbial phrase the three aspectual conditions were identical and did not differ in reading time (all $F_{1/2} < 1$).

When readers encountered the adverbial phrase they slowed down in case of a *for*-adverbial (mean RT 60.0 ms/char) and in case of an *in*-adverbial (mean RT 60.9 ms/char) compared to aspectual control (mean RT 54.1 ms/char). In ANOVAs, this difference was reflected in a significant main effect of *adverbial* ($F_1(2, 58) = 4.05$; $p < .05$; $F_2(2, 58) = 3.49$; $p < .05$). Paired t-tests using a Bonferroni adjusted alpha revealed that mismatch was read more slowly than control ($t_1(29) = 2.26$; $p < .025$; $t_2(29) = 2.32$; $p < .025$) and that coercion was read more slowly than control ($t_1(29) = 2.34$; $p < .025$; $t_2(29) = 2.44$; $p < .025$).

In the additive coercion condition the slow-down extended to the subsequent PP region (mean RT 105.5 ms/char), while mismatch (mean RT 93.4 ms/char) and

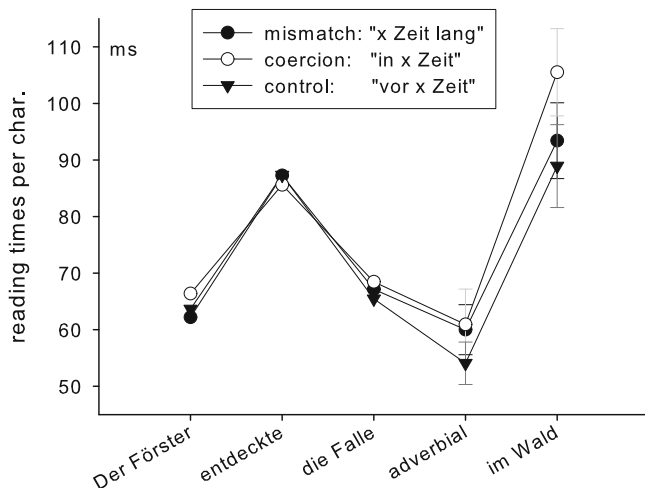


Fig. 8.4 Reading times per character in conditions with SVAO word order (+95% CIs) in Experiment 2

control (mean RT 88.9 ms/char) were roughly the same. In ANOVAs, this difference resulted in a significant main effect of *adverbial* ($F_1(2, 58) = 7.56$; $F_2(2, 58) = 4, 83$; $p < .05$). Paired t-tests revealed that reading times in the coercion condition were slower than in the mismatch condition ($t_1(29) = 3.04$; $p < .025$; $t_2(29) = 2.42$; $p < .025$). There was, however, no significant difference between mismatch and control ($t_1(29) = 1.21$; $p = .24$; $t_2(29) = .78$; $p = .44$).

Furthermore, reading times of coerced sentences were analyzed contingent on judgments. Thus, those trials in which participants judged a sentence with an *in*-adverbial semantically acceptable were analyzed separately from those in which they were considered semantically ill-formed. The former were trials in which subjects actually computed a coerced meaning (henceforth *coercion trials*) while the latter are trials where they failed to accomplish coercion (henceforth *failed reanalysis trials*). Table 8.1 presents the results.

On the regions preceding and including the adverbial, coercion trials and failed reanalysis trials had reading times of comparable length. At the sentence final PP, however, failed reanalysis trials were slower than coercion trials which didn't differ from control. The former difference was significant as revealed by an independent samples t-test ($t(146) = 2.59$; $p < .025$). In contrast, the numerical difference between coercion trials and control was not reliable ($t(209) = 1.26$; $p = .21$).

8.3.2.3 Reading Times: SVAO Word Order

Figure 8.5 shows mean reading times of SVAO word order across the three conditions. At the adverbial region control had a mean RT of 50.70 ms/char, mismatch had 51.16 ms/char and coercion had 50.54 ms/char. At the following

Table 8.1 Mean reading times in SVOA word order conditionalized on judgments in Experiment 2

	“Yes” coercion	“No” coercion	“Yes” control
Identical regions			
<i>Subject</i>	62.60 (4.53)	61.63 (3.10)	59.69 (2.47)
<i>Verb</i>	83.51 (3.88)	80.48 (3.37)	81.16 (2.74)
<i>Object</i>	66.06 (3.76)	63.08 (3.12)	60.87 (2.45)
Adverbial and PP			
<i>In-adverbial</i>	59.20 (4.37)	61.88 (4.68)	52.14 (1.89)
<i>PP</i>	94.76 (5.05)	114.57 (5.71)	86.85 (3.70)

Note: reading times per character in milliseconds (plus mean standard errors)

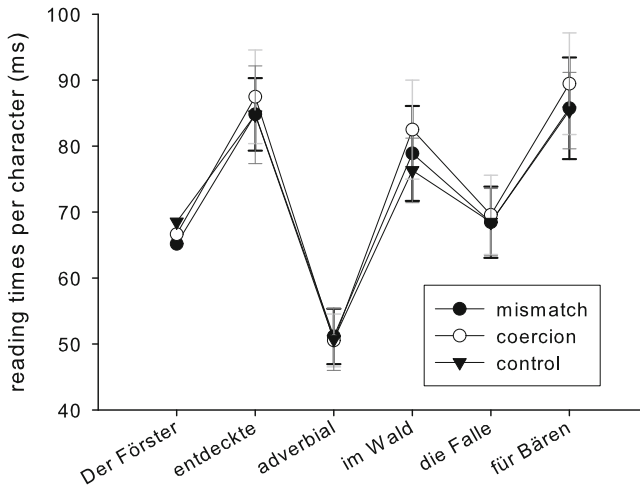


Fig. 8.5 Reading times per character in conditions with SVAO word order (+95% CIs) in Experiment 2

spillover region separating the adverbial from the object, control was numerically read fastest with a mean RT of 76.31 ms/char, mismatch had 78.90 ms/char and coercion had 82.49 ms/char. At the object region, control had a mean RT of 68.47 ms/char, mismatch had 68.46 ms/char and coercion had 69.58 ms/char. The sentence final segment had mean RTs of 85.38 ms/char in control and 85.73 and 89.46 ms/char in mismatch and coercion, respectively. Statistical analyses of the reading times revealed neither a significant difference at the adverbial region ($F_{1,2} < 0.5$) nor at any of the following segments (all $F_{1,s} < 1.5$; all $F_{2,s} < 1$). Since there was an overall numerical trend going slightly in the expected direction, we further analysed RTs of the end of the sentence by adding up the reading times of the last four segments. ANOVAs analyzing these cumulated RTs did not reveal any reliable differences between the three adverbials either ($F_{1,2} < .05$).

8.3.3 Discussion

This experiment investigated the processing of sentences involving aspectual mismatch and additive coercion. This type of coercion has so far not been investigated in the psycholinguistic literature. The findings provide evidence for additive coercion leading to considerable processing difficulty. First, judgments suggested that coercion was carried out only in approximately 50% of all trials. Second, coercion led to longer reading times than control. This effect cannot be attributed to semantic markedness of the coerced sentences because reading pace also slowed down in coercion trials which were judged “yes, sensible”. Finding processing difficulty across different types of coercion lends further support to the claim that aspectual coercion is a cognitively difficult operation generalizing over the few aspectual coercion types that have been investigated so far.

Besides the coercion effect we obtained a mismatch effect at the adverbial region in the SVOA word order. It is important to note that both coercion and mismatch were present at the region before the final segment. This indicates that the most extreme version of the LAIH – aspectual processing delayed until the very end of the sentence – cannot be true. Instead, we have to allude to the notions of a complete verb-argument structure and/or predication to properly lay out the range of possible hypotheses.

Crucial for the questions addressed in this paper, however, is that coercion and mismatch effects were only elicited by adverbials modifying a complete verb-argument structure. What is particularly striking about the results is the lack of a mismatch effect, even though the judgment data reveal that subjects were well aware of the aspectual mismatch. This shows that the information of the subject plus the verb is not enough to determine lexical aspect.

This finding is interesting because at first sight it conflicts with the incrementality assumption usually made in the processing literature at least for syntax (e.g. [Frazier \(1987\)](#), [Crocker \(1996\)](#), [Hagoort \(2003\)](#)). Although readers could in principle immediately interpret the initial part of the achievements, lexical aspect was not immediately determined and interpretation seemed to be delayed. However, at the end of the sentence, when providing a sensibility judgment, participants clearly had accomplished an aspectual interpretation. Judgments were relatively fast and were equally easy in the coercion condition, the mismatch condition and aspectual control. The findings thus provide evidence against the incremental aspectual interpretation hypothesis (IAIH).

How can this be, given the abundant evidence for incrementality from a wide range of psycholinguistic phenomena? In this experiment parsing would have been more efficient if the processor had immediately decided on an aspectual class, because the aspectual information of the lexical items could then directly be integrated into a situation model. The subject and the verb were maximally informative with respect to which lexical aspect had to be chosen. But note that this was due to the fact that the verbs used in this study were carefully selected. They provided clear instances of transitive achievement verbs with bounded subjects.

In the “real world”, however, matters are often not that clear. In the majority of cases, looking at the verbal information alone may not tell the comprehender anything about the relevant situation type. Often, aspectual distinctions are far from clear cut and we are dealing more with a continuum than with a discrete system. Immediately deciding on the aspectual class would thus lead to a vast amount of rather costly aspectual reinterpretation. As a result, the processor might work most efficiently by waiting until the verb has received the minimally required arguments. This may be even more so, since incremental syntactic interpretation already provides a structured representation that can be kept in working memory keeping memory load comparably low.

If aspectual processing was delayed, why did no effects show up further downstream the sentence when readers eventually encountered the extraposed object? A possible explanation for this lack of effect may be that the materials contained adjuncts – the first spillover region – that intervened between the adverbial and the direct object. Although they were kept constant across conditions, the intervening material may have slowed down processing of the following material in general.⁴ In turn, potential aspectual effects may have been obscured. In fact, there is psycholinguistic evidence for difficulty caused by intervening adjuncts (see eg. Staub et al. (2006)). In any case, we have to be very careful of drawing hasty conclusions because the suggested interpretation of results crucially relies on analysing null effects in the SVAO order.

To deal with this problem we decided to leave out the intervening adjuncts in the self-paced reading experiment testing OVAS constructions which will be reported in the next section. We will see clear indication of delayed effects there. Furthermore, the eyetracking experiment (Experiment 4) will provide additional evidence to substantiate the tentative claims made here.

8.4 Experiment 3: Modification of Complete VPs

Does the verb with its internal argument form a natural unit with respect to lexical aspect? Intuitive judgments reveal that the VP already encodes a minimal situation. For instance, we can talk about situations like *to build a house* while we leave it open who is actually building it. The examples in (13) illustrate that actually no local subject is needed to determine the aspectual class.

⁴An anonymous reviewer pointed to an alternative explanation which is exactly the opposite, ie. a speed-up due to irrelevant intervening material. Again, this may have obscured a potential effect at the object region. We fully agree that this is possible, too. No matter which explanation is correct, we think it is crucial to show that aspectual processing is delayed until the verb has received its minimally required arguments. Anticipating what is yet to come, this is exactly the kind of effect we observed in both of the following experiments.

- (13) a. Es wurde begonnen den Schlüssel zu suchen.
 It was begun the key to search.
 Somebody began to search the key.
- b. *Es wurde begonnen den Schlüssel zu finden.
 It was begun the key to find.
 *Somebody began to find the key.

Begin states that there was a start event of some durative process. In (13-a) *search the key* is of the required type, but the achievement *find the key* in (13-b) is not. Crucially, in the constructions in (13) the expletive *it* only serves as a dummy subject which lacks any semantic content.

Given these linguistic facts, it is quite plausible to assume that the processor determines lexical aspect at the level of the verb and its internal argument(s). This is stated in the *Complete Verb Phrase Hypothesis* in (14).

- (14) Complete Verb Phrase Hypothesis (CVPH)
 A complete VP is specified for lexical aspect.

The CVPH stands in sharp opposition to other linguistic facts. Above, we used the sentence *visitors arrived all night* to demonstrate that the right choice of subject bears an important influence on the aspectual class of the whole sentence. At first sight, these linguistic facts are providing conflicting evidence. On the one hand, the VP seems to be sufficient to determine lexical aspect, but on the other hand, complete verb-argument structures have to be considered. It is thus interesting to investigate whether adverbial modification of a complete VP will reveal mismatch or coercion effects well before the subject is present. The present experiment tested the CVPH by looking at the processing of OVAS sentences, i.e. adverbial modification in constructions with extraposed subjects.

8.4.1 Method

The present self-paced reading experiment tested the CVPH using slightly modified materials of the previous experiments with OVAS word order. (15) is a sample item, vertical lines indicate segmentation.

- (15) a. Den Haarriss_{obj.} | am Wasserrohr | bemerkte | vor dreißig
 The hairline-crack | at-the water-pipe | noticed | ago thirty
 Minuten | ...
 minutes | ...
 Thirty minutes ago, [...] noticed the hairline crack at the water-pipe.
- b. Den Haarriss_{obj.} | am Wasserrohr | bemerkte | in dreißig
 The hairline-crack | at-the water-pipe | noticed | in thirty
 Minuten | ...
 minutes | ...
 In thirty minutes, [...] noticed the hairline crack at the water-pipe.

- c. Den Haarriss_{obj.} | am Wasserrohr | bemerkte | dreißig Minuten lang
 The hairline-crack | at-the water-pipe | noticed | thirty minutes long
 | ...
 | ...
 For thirty minutes, [...] noticed the hairline crack at the water-pipe.
- d. ... ein aufmerksamer Klempner
 ... an attentive plumber

Example (15-a) is aspectual control, (15-b) involves additive coercion and (15-c) contains an aspectual mismatch. The case disambiguated object always appeared in the sentence initial position. The object was always definite and maximally specific to license it in that position. Furthermore, to make the object-initial word order expected, all sentences, items as well as fillers, had an object before subject word order.

The number of the verb may provide some information about the forthcoming subject. A bare plural subject, for instance, is ungrammatical following a singular verb. For this reason, besides *adverbial*, *number* was manipulated in a 3×2 factorial design resulting in a total of six conditions. Each item in each aspectual condition was constructed in two versions, with a singular subject (eg. *an attentive plumber*) and with a plural subject (eg. *a few attentive plumbers*).

The 75 fillers from the previous experiment were transformed into object-initial sentences. Items and fillers were assigned to six lists in a latin square design. The experimental procedure was identical to the previous experiment: after reading a sentence participants had to provide a sensuality judgment.

42 native German speakers (31 female; mean age 23.0 years) from Tübingen University took part in the study for a payment of 5€. Participants were randomly assigned to lists (five subjects per list).

8.4.2 Results

8.4.2.1 Makes Sense Judgments and Judgment Times

While *control* was accepted in 89.7% (*sing.*: 92.9% vs. *pl.*: 86.6%), *mismatch* was only accepted in 31.3% (*sing.*: 26.8% vs. *pl.*: 35.7%). *Coercion* was intermediate with 63.0% “yes” responses (*sing.*: 58.7% vs. *pl.*: 67.6%). The sentences involving aspectual coercion were judged as sensible in the majority of cases, as was confirmed by a t-test testing whether *coercion* significantly differed from 50% ($t_1(41) = 3.64$; $p < .01$; $t_2(29) = 3.51$; $p < .01$).

In ANOVAs, the difference between the aspectual conditions was reflected by a significant main effect of *adverbial* ($F_1(2, 82) = 124.50$; $p < .01$; $F_2(2, 58) = 91.48$; $p < .01$). *Number* had a comparably weaker influence on the judgments.

Table 8.2 Mean judgment times in Experiment 3

	Judgment time	N (out of 210)
Singular		
“No” to mismatch	1,254 (664)	154
“Yes” to coercion	1,549 (759)	121
“Yes” to control	1,341 (633)	195
Plural		
“No” to mismatch	1,294 (655)	134
“Yes” to coercion	1,495 (666)	142
“Yes” to control	1,331 (672)	181

Note: Judgment times (+std.) of expected answers in ms

While the main effect of *number* was not reliable ($F_1(1, 41) = 3.25$; $p = .08$; $F_2(1, 29) = 2.07$; $p = .17$), the interaction between *number* and *adverbial* was significant ($F_1(2, 82) = 5.32$; $p < .05$; $F_2(2, 58) = 4.18$; $p < .05$). The interaction is due to the fact that the differences between the aspectual conditions are somewhat bigger in the *singular* than in the *plural* conditions.

Table 8.2 shows the judgment times for “no” responses in the *mismatch* conditions and “yes” responses in the *coercion* and *control* conditions.

In both *number* conditions, judgments took longer for sentences involving aspectual coercion than for controls or sentences involving an aspectual mismatch. In ANOVAs which analyzed judgment times of expected answers (= “no” with respect to *mismatch* and “yes” with respect to *coercion* and *control*), this difference was reflected by a main effect of *adverbial* that was significant by participants ($F_1(2, 82) = 5.14$; $p < .05$; $F_2(2, 58) = 2.84$; $p = .08$). Neither the main effect of *number* ($F_{1,2} < 1$) nor the interaction between *adverbial* and *number* was reliable ($F_{1,2} < 1$). A paired t-test comparing judgment times for coercion versus control (pooled over *number* conditions) revealed a reliable difference between these two conditions ($t_1(41) = 2.40$; $p < .05$; $t_2(29) = 3.10$; $p < .01$).

8.4.2.2 Reading Times

The reading times for the three aspectual conditions are depicted in Fig. 8.6. They were longer in the aspectual mismatch and the coercion condition compared to control. Since ANOVAs revealed that the pattern was the same in the singular and the plural conditions, the data were aggregated over the corresponding *singular* and *plural* conditions.

A difference in reading times only showed up at the head noun of the subject phrase (*mismatch*: 88.3 ms/char vs. *coercion*: 88.1 ms/char vs. *control*: 77.3 ms/char).

At the adverbial region, the aspectual conditions did not differ. *Mismatch* had a mean RT of 50.42 ms/char, *coercion* had 52.92 ms/char and *control* had 50.26 ms/char. ANOVAs didn’t reveal a significant main effect of *aspect*

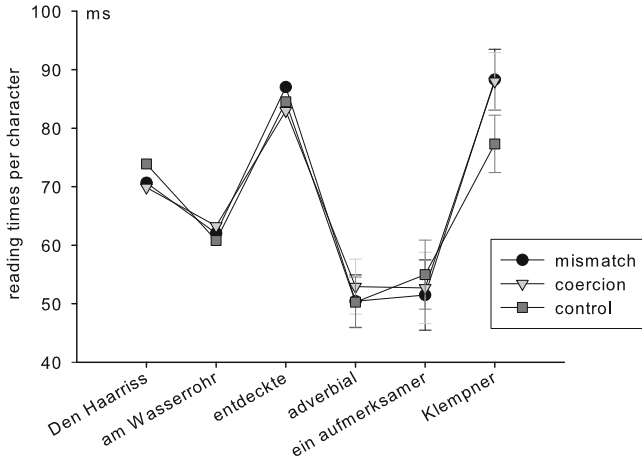


Fig. 8.6 Mean reading times (+95% CIs) in Experiment 3

($F_1(2, 82) = 2.64$; $p = .09$; $F_2(2, 58) = 1.01$; $p = .36$). At the following first part of the subject phrase, there were also no differences in reading time. Numerically, control was even slowest.

When readers encountered the head noun of the subject phrase, reading times were slower in the mismatch and the coercion conditions than in the *singular* and *plural* controls. In ANOVAs, this difference was reflected by a significant main effect of *aspect* ($F_1(2, 82) = 7.32$; $p < .01$; $F_2(2, 58) = 7.88$; $p < .01$).

8.4.3 Discussion

The present experiment provides additional evidence that processing sentences with aspectual mismatch and coercion are more difficult than aspectual controls. Interestingly, the difficulty only emerged after readers had encountered the extraposed subject phrase, that is only at the point when the verb had received its minimally required arguments. In contrast, at the critical adverbial and the subsequent region all three conditions were read equally fast. The results thus provide clear evidence against the Complete Verb Phrase Hypothesis (CVPH). The VP did not contain enough information to allow for aspectual mismatch and coercion effects when it was combined with a mismatching or coercing adverbial. Furthermore, since only delayed effects were found, it is not surprising that the *number* information wasn't used to predict what kind of subject is yet to come.

In contrast to the findings of the previous experiment, the results of the present study show delayed aspectual effects. This delayed effect can best be explained by a hierarchical organization of aspectual processing, where first the eventuality

of the verb-argument structure has to be computed and only in a second step is adverbial modification possible. The findings perfectly match the predictions of the Late Aspectual Interpretation Hypothesis (LAIH).

Can these late effects be due to lexical aspect being underspecified until readers cross a sentence boundary? On the basis of the findings reported in this paper this actually seems to be a viable option. Additional evidence from an experiment investigating the processing of sentences like (16-a) vs. (16-b) in Bott (2010, Experiment 2) makes this explanation, however, very unlikely.

- (16) a. Peter joggte in fünfzehn Minuten. . .
 Peter jogged in fifteen minutes. . .
 b. Peter joggte fünfzehn Minuten lang. . .
 Peter jogged for fifteen minutes. . .

In that experiment, reading times of the adverbial phrases in examples (16-a) vs. (16-b) indicated enhanced difficulty in (16-a) as compared to (16-b). This is interesting, since (16-a) can be continued in a sensible way, for instance by providing the right kind of path argument *drei Kilometer (three kilometers)*. It thus seems that aspectual processing is in fact delayed until a minimal verb-argument structure is complete.

Taken together, Experiments 2 and 3 thus demonstrate a fascinating interplay between the parsing of argument structure and of lexical aspect. The former seems to be prior to aspectual processing. This adds an interesting new parameter to the incrementality debate, namely the domain size with respect to a particular phenomenon. In the next section we will further elaborate on another facet of incremental interpretation, that is which stages of processing are affected by processing lexical aspect.

8.5 Experiment 4: SVOA Versus AVSO Sentences

Is it possible that self-paced reading data are too coarse to detect aspectual effects in yet incomplete verb-argument structures? To check whether this was the case, we conducted an experiment in which we measured eye movements while participants were reading SVOA versus AVSO sentences. The latter construction allows us to keep track of aspectual processing while the verb and its arguments come in one piece after the other.

Eye-movement data may yield additional information with regard to the SVOA construction, too. They provide a more fine-grained measure of the stages of processing that are targeted by aspectual mismatch and aspectual coercion, respectively (cf. Rayner (1998) for an overview). Do mismatch and coercion already affect the initial analysis or will mismatch and coercion effects only show up in regressive eye-movements?

8.5.1 Method

8.5.1.1 Materials

We constructed 36 unambiguously transitive achievement sentences in six conditions according to a 3 (*adverbial: mismatch vs. additive coercion vs. control*) \times 2 (*word order: SVOA vs. AVSO*) factorial design. A sample item is provided in (17-a) and (17-b), vertical lines indicate interest area boundaries. All items subordinated an *although*-clause which was segmented into four spillover regions. Line breaks always occurred after the first spillover region *obwohl* (*although*). The full set of materials is contained in the appendix.

- (17) a. Der Ringer | gewann | das Turnier | ganze drei Stunden (in 3 h
 The wrestler | won | the tournament | whole three hours (in 3 h
 / vor 3 h), | obwohl | es | viele | Konkurrenten gab.
 / ago 3 h), | although | it | many | competitors were.
 The wrestler won the tournament for three hours (in three hours/three
 hours ago), although there were many competitors.
- b. Ganze drei Stunden (In 3 h / Vor 3 h) | gewann | der Ringer | das
 Whole three hours (In 3 h / Ago 3 h) | won | the wrestler | the
 Turnier, | obwohl | es | viele | Konkurrenten | gab.
 tournament, | although | it | many | competitors | were.
 For three hours (In three hours/Three hours ago), the wrestler won the
 tournament, although there were many competitors.

A latin square was used to distribute the experimental sentences over six lists. One hundred and twenty-two fillers (40 non-sensical) were added to each list. Each experimental item and 62 of the distractors were followed by a question querying whether the sentence was sensible. Sixty fillers were followed by ordinary comprehension questions to prevent participants from anticipating the judgment while reading the sentence.

8.5.1.2 Participants

Participants were 24 students from Tübingen University (mean age 26.1, range from 19 to 33 years; 18 female) who received 8€ for their participation. None of them had participated in any of the previous experiments. Four participants were randomly assigned to each list. Five additional participants had to be excluded from the analysis due to calibration problems ($N = 3$) or error rates above 40% in the practice ($N = 2$).

8.5.1.3 Procedure

A desktop-mounted Eyelink 1,000 eyetracker monitored the gaze location of the participant's dominant eye. The eyetracker has a spatial resolution of 0.01° of visual angle and samples gaze location every millisecond. Participants viewed the stimuli binocularly on a 19 in. monitor 70 cm from their eyes. A head rest minimized head movements. The experiment was implemented using the Experiment Builder software and eyetracking data were exported with the Data Viewer software package.

Subjects were tested individually. The tracker was calibrated using a 3×3 grid guaranteeing that all fixations were less than 0.5° apart from the calibration stimuli. After calibration was completed, participants read the experimental instructions on the screen. This was followed by a practice session of ten items. In the experiment, each trial started with a calibration check. The tracker was recalibrated as necessary. Eye-movements were recorded during reading.

The trial began with the presentation of a screen which served as calibration control with a little black dot in the position where the center of the first word would appear. If no fixation was registered within 5 s, recalibration was enforced. Otherwise a sentence in yellow 15 point font size letters appeared in the center of a navy blue screen. Three characters corresponded approximately to 1° of visual angle. After reading the sentence participants had to move their eyes to an asterisk at the bottom of the screen. Fixating the area around the asterisk triggered the presentation of the question screen querying whether the sentence was sensible. There was no time limit for providing an answer.

8.5.1.4 Data Analysis

Prior to all analyses we preprocessed the data. Fixations that were shorter than 80 ms and within one character space of the previous or next fixation were assimilated to this fixation. The remaining fixations shorter than 80 ms or longer than 1,200 ms were excluded. This affected 5.6% of the data.

We analyzed fixation times with respect to five eyetracking measures.⁵ *First-pass time* is the total time spent in an interest area before the reader moves on or looks back in the text. *Regression path durations* are the sum of fixation durations from the time the reader enters a region, to the time when the reader enters the following region, that is it includes first-pass time plus the time spent on regressions. Finally, *total reading time* is the sum of all fixations on a particular region. If a region was skipped during first-pass or never fixated at all, we replaced the missing value in the

⁵In addition to the measures reported here we analyzed first fixation durations. Since there were no differences we refrain from reporting these.

first-pass times, the regression path duration or the total times by a value of zero.⁶ As for first-pass and total times, we analyzed reading times per character to compensate for systematic length differences between the three adverbial types (mean number of characters were 17.1 (coercion), 18.1 (control) and 19.6 (mismatch)). We also measured two types of proportions of regressions: *first pass regression ratios*,⁷ ie. the proportions of how often readers launched a regression from a region during first pass (forward) reading. The proportion of *regressions in* a region is a measure of how often it was entered from the right.

8.5.1.5 Predictions

If aspectual processing is delayed until the verb-argument structure is complete, we will get the following predictions. During first-pass reading aspectual mismatch and coercion should not cause any delay or regressions out of regions that are encountered before the transitive verb has received both arguments. Only then should readers slow down and/or launch regressions to earlier parts of the sentence. Thus, in the SVOA conditions we expected mismatch and coercion effects to immediately show up at the adverbial, whereas in the AVSO conditions we expected delayed effects of ‘early’ reading time measures (ie. first-pass times, first-pass regression ratios and regression path durations) showing up at the object region. Indeed, this is what we found.

8.5.2 Results

The conditions were judged as follows: in the SVOA word order, *mismatch* was falsely accepted 14.6%, *coercion* was accepted 84.7% and *control* was accepted 81.3% of the time. Acceptance rates were similar in the AVSO word order: *mismatch* was falsely accepted 18.1%, *coercion* was accepted 88.9% and *control* was accepted 91.7% of the time. ANOVAs analyzing ‘correct’ judgments revealed no reliable main effects of *word order* or *adverbial* (all $F_{1/2} < 2.6$), but an interaction that was marginal by participants ($F_1(1, 23) = 2.76$; $p = .08$; $F_2(1, 35) = 3.34$; $p < .05$). The lack of a main effect of *adverbial* indicates that unlike in the previous experiments there were no consistent differences between the adverbial types. Additive coercion can be as felicitous as control when the context provides the relevant information (eg. some obstacle to the culminating event mentioned in the *although* clause that indicates what the preparatory process may have been).

⁶If a region receives no fixations, its information was most probably already available due to parafoveal preview of the preceding segment or predictive parsing.

⁷We will refer to them loosely as *regressions out*.

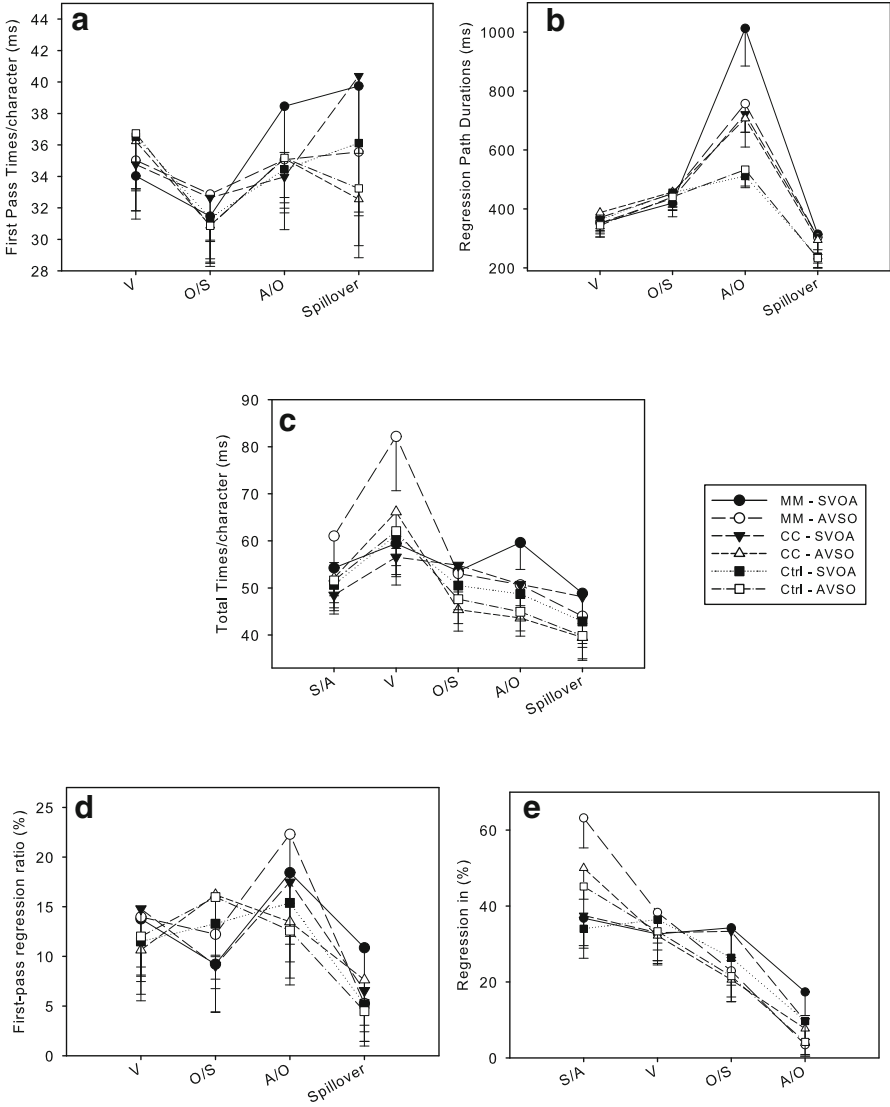


Fig. 8.7 Mean reading times and proportions of regressions (+lower limit of 95% CIs) in Experiment 4. Abbreviations: Panel a) first-pass times. Panel b) regression path durations. Panel c) total times. Panel d) first-pass regression ratios. Panel e) proportions of regression in. A adverbial, V verb, S subject, O object

Figure 8.7 displays the mean first-pass time, regression path duration, total time and proportions of regressions in all six conditions up to the first spillover region. The first region was left out of the graphs because of length differences between the different adverbials. In the following paragraphs, we will walk through the eye-tracking record region by region.

At the first region of interest (ROI) there were big lexical differences between conditions. To investigate whether a potential mismatch effect already affected the preview of the verb, we compared the first-pass times in the AVSO mismatch and the AVSO control condition in a pairwise comparison. The difference between mismatch (38.6 ms/char) and control (37.3 ms/char) was not significant ($t_{1/2} < .8$; $p_{1/2} > .4$). It is thus unlikely that aspectual mismatch was detected during preview of the verb from the adverbial ROI. Proportions of regressions into this region revealed a clear difference between adverbial types in the AVSO word order. Mismatch had 63.2% regressions into the adverbial region, whereas coercion and control had 50 and 45.1%, respectively. By contrast, in the SVOA word order proportions of regressions in were roughly the same (mismatch: 36.8%; coercion: 37.5%; control: 34.0%). In ANOVAs, these differences led to significant main effects of *order* ($F_1(1, 23) = 12.00$; $p < .01$; $F_2(1, 35) = 19.09$; $p < .01$) and *adverbial* ($F_1(2, 46) = 4.33$; $p < .05$; $F_2(2, 70) = 3.52$; $p < .05$), but no significant interaction ($F_1(1, 23) = 2.03$; $p = .14$; $F_2(2, 70) = 1.80$; $p = .17$). Pairwise comparisons revealed a significant mismatch effect in the AVSO order (mismatch vs. control: $t_1(23) = 2.92$; $p < .01$; $t_2(35) = 2.96$; $p < .01$), but not in the SVOA order ($t_{1/2} < 1$).

At the verb region, the AVSO conditions did not differ either in first-pass times or regression path duration. ANOVAs revealed no reliable main effect of *adverbial* nor a significant interaction between *adverbial* and *order* (all $F_{1/2} < 1.2$). Also, first-pass regression ratios did not differ between conditions (all $F_{1/2} < 1$). When integrating the verb, aspectual mismatch or coercion thus went unnoticed in the AVSO conditions. The proportions of regressions in didn't differ in the verb region (all $F_{1/2} < 1.2$), either. In total times, however, the kind of adverbial made a clear difference in the AVSO sentences. In the mismatch condition total times were longer (82.2 ms/char) than in the coercion (66.2 ms/char) or the control condition (62.1 ms/char). This difference was absent in the SVOA conditions. ANOVAs analyzing the total times in all six conditions revealed a significant main effect of *order* ($F_1(1, 23) = 7.97$; $p < .01$; $F_2(1, 35) = 10.07$; $p < .01$), a significant main effect of *adverbial* ($F_1(2, 46) = 4.19$; $p < .05$; $F_2(2, 70) = 4.58$; $p < .05$) and an interaction that was significant by participants and marginal by items ($F_1(2, 46) = 3.34$; $p < .05$; $F_2(2, 70) = 2.83$; $p = .08$). This effect in total times in combination with the lack of effects in the earlier reading time measures indicates that the mismatch effect in the AVSO mismatch condition came from readers noticing a problem with the verb while rereading the sentence.

The third ROI contained the direct object in the SVOA conditions and the subject in the AVSO conditions. ANOVAs analyzing first-pass times and regression path durations revealed no significant main effects or a reliable interaction between the two (all $F_{1/2} < 1.3$). Again, there was no mismatch or coercion effect in the AVSO order. This is further corroborated by first-pass regression ratios. Numerically, in both word orders control led to slightly even more regressions than mismatch or

coercion. In total time, conditions were more or less the same. ANOVAs revealed a by participants significant main effect of *order* ($F_1(1, 23) = 6.39$; $p < .05$; $F_2(1, 35) = 2.15$; $p = .15$) due to SVOA conditions having slightly higher total times than the AVSO conditions. Neither the main effect of *adverbial* nor the interaction was reliable (both $F_{1/2} < 2.4$). This suggests that the arguments were not as important as the verb when it came to regressive eye movements due to aspectual mismatch. Proportions of regressions into this region differed between the two word orders (main effect of *order*: $F_1(1, 23) = 15.06$; $p < .01$; $F_2(1, 35) = 12.01$; $p < .01$). SVOA sentences received on average 9.2% more regressions than did AVSO sentences. Neither the main effect of *adverbial* nor the interaction were significant.

The next ROI was the critical segment. In the SVOA word order, it was the region where readers encountered a mismatching or coercing adverbial. In the AVSO word order, the readers got the direct object saturating the second argument slot. First-pass times didn't differ significantly between conditions (all $F_{1/2} < 1.6$). Yet, pairwise comparisons between mismatch and control revealed that in the SVOA order mismatching adverbials were read slower than *ago*-adverbials in the control condition (38.5 ms/char vs. 34.4 ms/char). This difference was significant by participants and marginal by items ($t_1(23) = 2.10$; $p < .05$; $t_2(35) = 1.85$; $p = .07$). There was, however, no difference between mismatch and control in the AVSO order (35.1 ms/char vs. 35.2 ms/char: $t_{1/2} < .1$). First-pass regression ratios, however, indicated an early mismatch effect in the AVSO word order, too. In AVSO sentences, mismatch led to 22.9% regressions out of the object region as compared to 13.9% in coercion and 12.6% in control sentences. In the SVOA conditions, the proportions ranged between 16.0 and 18.6%. ANOVAs revealed a marginally significant main effect of *adverbial* ($F_1(2, 46) = 2.93$; $p = .07$; $F_2(2, 70) = 3.78$; $p < .05$) but no significant main effect of *order* or their interaction. In pairwise comparisons, the mismatch effect turned out significant in the AVSO order ($t_1(23) = 2.87$; $p < .01$; $t_2(35) = 2.67$; $p < .05$), but not in the SVOA order ($t_{1/2} < 1$). In regression path duration we found a clear mismatch effect in both word orders. The SVOA mismatch condition had a mean regression path duration of 1,012 ms, whereas coercion and control had 757 and 722 ms. In the AVSO order we observed the same pattern: mismatch was read slowest with a mean regression path duration of 707 ms, whereas coercion and control were much faster with 512 and 533 ms. In ANOVAs this was reflected by significant main effects of *order* ($F_1(1, 23) = 30.23$; $p < .01$; $F_2(1, 35) = 35.49$; $p < .01$) and *adverbial* ($F_1(2, 46) = 11.03$; $p < .01$; $F_2(2, 70) = 17.77$; $p < .01$), but no reliable interaction ($F_{1/2} < 1$). Thus, when readers encountered an aspectually mismatching adverbial in SVOA word order, they launched a regression. In the AVSO order they regressed, too, but only launched a mismatch-induced regression after predication was complete. The total times followed the same pattern. Mismatch lead to longer RT than coercion and control. Statistically, the main effects of

order ($F_1(1, 23) = 12.87$; $p < .01$; $F_2(1, 35) = 7.78$; $p < .01$) and *adverbial* ($F_1(2, 46) = 10.88$; $p < .01$; $F_2(2, 70) = 7.74$; $p < .01$) were reliable, but there was no reliable interaction between *order* and *adverbial* ($F_{1/2} < 1$). Regressions in also differed between conditions. In the SVOA order there was a mismatch effect of 7.5% more regressions into mismatching adverbials than into control conditions, but the AVSO order was roughly the same. In ANOVAs this led to a significant main effect of *order* ($F_1(1, 23) = 8.61$; $p < .01$; $F_2(1, 35) = 15.00$; $p < .01$), a non significant main effect of *adverbial* ($F_{1/2} < 2$) and a by participants significant interaction between *order* and *adverbial* ($F_1(2, 46) = 4.61$; $p < .05$; $F_2(2, 70) = 2.80$; $p = .07$).

At the following spillover region, there were no reliable differences in any of the eyetracking measures.

8.5.3 Discussion

The present experiment provides clear support for the LAIH. The makes sense judgments show that readers noticed the aspectual mismatch in the aspectual mismatch condition. The eye movements indicate that mismatch detection in the SVOA mismatch condition was very fast. Immediately when readers encountered the mismatching adverbial reading was slower than in the control condition. The time course was different in the AVSO mismatch condition. Before having read the arguments, the lexical aspect of the verb was not composed with the mismatching adverbial. Only after the complete predication a delayed mismatch effect emerged. Like in the SVOA order, mismatch detection affected early eyetracking measures, namely first-pass regression ratio.

May a potential early mismatch effect in the AVSO conditions have gone unnoticed because of too small sample size in the present study? This is a legitimate concern because we are basically interpreting null effects. Nevertheless, we think that this is unlikely to be true. After completing the study we tested additional 12 participants to gain more statistical power. Still, the pattern of results was exactly the same as reported here (cf. Bott (2011)).

What is puzzling about the results of the present experiment is that both coercion conditions perfectly lined up with the control conditions.⁸ This doesn't fit the results of the two self-paced reading studies reported earlier. An explanation for the divergent findings might be that the sentences in the present experiment were always continued with an *although* clause mentioning an obstacle that made the culmination hard to achieve. For instance, in the sample item (17-a)–(17-b) the culminating event *win the tournament* when combined with the coercing *in*-adverbial called

⁸We cannot go into the details here, but statistical analyses of the complete sentence also taking into account the subsequent spillover regions didn't yield a coercion effect whatsoever.

for a preparatory process not expressed in the main clause. The *although* clause implicitly stated what the preparation probably was, namely fighting a lot of fights. In this sense, the *although* clause may have resolved additive coercion in the present experiment. The self-paced reading experiments didn't have this kind of continuation, so it may have been left to the reader to come up with an appropriate preparatory process. This explanation receives independent support from an event related potentials (ERP) study on additive coercion (Bott, 2010) using the same kind of materials that were tested in the present experiment. The study showed that additive coercion differed qualitatively from aspectual mismatch. While the latter led to a P600 effect, the former only elicited a working memory LAN. Based on these findings we have argued that additive coercion involves a smooth update of the aspectual representation without revising it first. This kind of smooth update may have gone unnoticed in the present experiment. It has to be left to further research to investigate whether a coercion effect would show up in an eyetracking experiment, too, when the sentence doesn't contain any information about what the missing eventuality might have been.

8.6 General Discussion

The present paper investigated the processing domain of lexical aspect. We formulated three alternative hypotheses, *incremental aspectual interpretation* (IAIH), the *complete verb phrase hypothesis* (CVPH) versus the *late aspectual interpretation hypothesis* (LAIH). The first hypothesis is inspired by much psycholinguistic work on sentence processing which shows that the sentence representation is constructed on an (at least) word-by-word basis. By contrast, the LAIH takes into account semantic work on lexical aspect like Dowty (1979), Verkuyl (1993) and Krifka (1998) which demonstrates that the arguments have a great impact and that aspect can only be determined at the sentential level.

In three reading time studies we used adverbial modification of yet incomplete verb-argument structures to investigate whether aspectual mismatch and additive coercion slow down reading of the adverbial when arguments are still missing. The results of the experiments provide evidence for the LAIH: the adverbial only showed semantic effects after the verb had received all its arguments. These findings are particularly striking since the completion study showed that the same sentence fragments were judged to be semantically ill-formed with comprehenders not being able to continue them in a sensible way. The findings are thus clearly inconsistent with the IAIH and the CVPH. Lexical aspect seems to be determined at the sentence level at the earliest.

Does this mean that lexical aspect isn't processed incrementally? Reflecting upon the notion of *incrementality*, two senses have to be distinguished. First, *incrementality* sometimes means *immediacy* which reflects whether some kind of

information is taken into account immediately, that is during first interpretation. Second, *incremental interpretation* sometimes is used to refer to processing that proceeds *word-by-word*. In principle, these two aspects are independent from each other and have to be kept apart. Whereas lexical aspect depends on a bigger processing domain than the word or even the phrase, the time-course of mismatch and coercion effects speak in favor of immediate aspectual processing. In the eyetracking experiment, mismatch detection in the SVOA condition occurred immediately at the adverbial as indicated by enhanced first-pass times. This lends support to assumption that aspectual processing is incremental, in principle, but that the processor operates on increment units that are bigger than the word or even the phrase.

It is an open question, however, whether the present findings can be generalized to other aspectual classes or languages. To date, we can only speculate about these issues. We find it plausible to assume that the aspectual system of a language has a big influence on how the language is processed. For instance, in a language with grammatical means to distinguish telic from atelic processes we would expect to find immediate mismatch effects irrespective of potentially missing arguments. We are planning experiments testing these predictions by looking at crosslinguistic differences in domain size comparing German and Russian. Turning to other aspectual classes, we expect the findings to be fully generalizable. In any of these, the arguments and the construction play a crucial role, as (18-a) and (18-b) demonstrate for activities and semelfactives, respectively.

- (18) a. In zwei Stunden joggte... (Peter bis ins nächste Dorf.)
 In two hours jogged... (Peter to the next village)
- b. In fünf Minuten hustete... (Peter das Tuch über den ganzen Tisch.)
 In five minutes coughed... (Peter the cloth over the whole table.)

In conclusion, this paper addresses the question at what hierarchical level of verb-argument structure the processor constructs an atomic event unit. Let us continue the analogy from chemistry. In the same way that the properties of an atom do not depend only upon the nucleus but also upon the number of electrons, the atomic orbitals and their occupancy, lexical aspect seems to be determined only at a supralexical level. Just as in chemistry and physics this doesn't mean that an atomic unit has no internal structure, but rather that our means of investigation – the kinds of adverbials used here – are only sensitive to the holistic properties of the atomic event as a whole.

Acknowledgements The author would like to thank Sam Featherston, Fritz Hamm, Janina Radó and two anonymous reviewers for their valuable comments on earlier drafts of this paper. Any remaining errors are, of course, the author's responsibility. The research reported here was supported by the German Science Foundation (DFG) funding the Collaborative Research Centres 441 and 833.

Appendix: Sentence Materials Used in Experiment 4

- 1 Der Ringer gewann das Turnier vor drei Stunden, obwohl viele Konkurrenten gab.
- 2 Die Yachter reichelten den Hafen vor einer halben Stunde, obwohl kein starker Wind blies.
- 3 Die Truppeleroberte die Festung vor sieben Tagen, obwohl die Gegnerin der Überzahl waren.
- 4 Die Juryleer nannte die Sprecherin vor drei Stunden, obwohl die Mitglieder verschiedener Meinung waren.
- 5 Der Stürmerlerkannte die Chance vor einer Minute, obwohl die Abwehr sehr dicht stand.
- 6 Der Kommissar überführte die Bankräuber vor zwei Tagen, obwohl sie kaum Spuren hinterlassen hatten.
- 7 Der Archäologe entdeckte die Skulptur vor zwei Wochen, obwohl das Grabmal teilweise verschüttet war.
- 8 Der Mechaniker identifizierte den Schaden vor zwanzig Minuten, obwohl der Motor nahezu unzugänglich war.
- 9 Der Wilderer legte den Hirsch vor zwanzig Minuten, obwohl drei Kugeln daneben gingen.
- 10 Der Jungel tötete die Stechmücke vor fünf Minuten, obwohl sie sehr schnell war.
- 11 Der Abiturienter hielt das Zeugnis vor vierzig Minuten, obwohl er der Letzter im Alphabet war.
- 12 Die Ehefrau fand die Liebesbriefe vor zwei Stunden, obwohl ihr Mann sie gut versteckt hatte.
- 13 Der Übersetzer verstand den Artikel vor zwei Tagen, obwohl er sehr schwierig war.
- 14 Der Tüftler fand die Maschine vor fünf Tagen, obwohl die Konstruktion sehr kompliziert war.
- 15 Das Kind bekam das Paket vor zwei Tagen, obwohl die Adresse nicht stimmte.
- 16 Der Klempner bemerkte den Rohrbruch vor zehn Minuten, obwohl das Rohr unter Putz verlegt war.
- 17 Der Forscher verließ die Höhle vor fünfzehn Minuten, obwohl seine Taschenlampe nicht funktionierte.
- 18 Der Schüler bewältigte die Aufgabe vor einer Viertelstunde, obwohl er nicht schnell schreiben konnte.
- 19 Der Mathematiker durchschaute die Idee vor zwei Tagen, obwohl sie unklar formuliert war.
- 20 Das Militär startete die Operation vor einer Stunde, trotz kleinerer Unstimmigkeiten im Führungsstab.
- 21 Das Virus befell den Patienten vor vierundzwanzig Stunden, trotz sorgfältigster Maßnahmen zur Quarantäne.
- 22 Der Rennfahrer besiegte den Konkurrenten vor dreißig Minuten, trotz Motorschaden in der ersten Runde.
- 23 Der Matrosen spähte den Leuchtturm vor fünf Minuten, obwohl es überaus neblig war.
- 24 Der Arzt erblickte den Kollegen vor zehn Minuten, obwohl viele Leutnanten der Konferenz teilnahmen.
- 25 Die Schülerin fasste das Problem vor drei Minuten, obwohl die Aufgabe verwirrend formuliert war.
- 26 Der Hausmeister erwischte den Jungen vor drei Minuten, obwohl dieser ihn lauszutricksen versuchte.

- 27 Die Großmutter|öffnet|den Brief|vor zwei Minuten,|obwohl|dieser|extra|versiegelt war.
- 28 Die Student|en|gründeten|die Initiative|vor zwei Wochen,|obwohl|die Unileitung|gegensiegiert hatte.
- 29 Die Putzfrau|schloss|das Fenster|vor einer Minute,|obwohl|es|heftig|geklemt hatte.
- 30 Der Mann|vergaß|die Telefonnummer|vor fünf Minuten,|obwohl|er sie|zu|memorieren versuchte.
- 31 Der Professor|akzeptiert|den Beweis|vor zwei Stunden,|obwohl|er|anfangs|eine Prämissel nicht verstanden hatte.
- 32 Die Nachbarin|erfuhr|das Gerücht|vor vierzig Minuten,|trotz|der|Verschwiegenheit|ihrer Nachbarn.
- 33 Das Mädchen|lergte|telleine Karte|vor zwei Tagen,|obwohl|der|Schwarzmarkt|quasi leergefegt war.
- 34 Die Polizei|ertappt|den Einbrecher|vor dreißig Minuten,|obwohl|dieser|alle|Eventualitäten einberechnet hatte.
- 35 Der Minister|eröffnet|das Museum|vor fünfzig Minuten,|obwohl|mehrere|Ansprachen gehalten wurden.
- 36 Der Einbrecher|zerbrach|die Scheibe|vor dreißig Minuten,|obwohl|dieser|aus|Panzer|glas bestand.

Note: all sentences in the SVOA control condition. Vertical lines indicate regions of interest.

References

- Altmann, G.T., and Y. Kamide. 1999. Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition* 73:247–264.
- Bott, O. 2010. *The processing of events, linguistics today, vol. 162*. Amsterdam: John Benjamins.
- Bott, O. 2011. *Limits to the immediate specification of lexical aspect*. Poster presented at CUNY 2011, Palo Alto.
- Brennan, J., and L. Pykkänen. 2008. Processing events: behavioral and neuromagnetic correlates of aspectual coercion. *Brain and Language* 106:132–143.
- Crocker, M. 1996. *Computational psycholinguistics*. Dordrecht: Kluwer Academic Publishers.
- Dowty, D. 1979. *Word meaning and Montague grammar*. Dordrecht: Reidel.
- Featherston, S. 2008. Thermometer judgments as linguistic evidence. In *Was ist linguistische Evidenz*, eds. Riehl, C., and A. Rothe. Aachen: Shaker.
- Ferretti, T.R., Kutas, M., and K. McRae. 2007. Verb aspect and the activation of event knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 33(1):182–196.
- Frazier, L. 1987. Sentence processing: a tutorial review. In *Attention and performance XII. The psychology of reading*, ed. Coltheart, M., 559–586. Hillsdale: Erlbaum.
- Hagoort, P. 2003. How the brain solves the binding problem: a neurocognitive model of syntactic processing. *Neuroimage* 20:S18–S29.
- Hamm, F., and M. van Lambalgen. 2005. *The proper treatment of events*. Malden: Blackwell.
- Heim, I., and A. Kratzer. 1998. *Semantics in generative grammar*. Malden: Blackwell.
- Krifka, M. 1998. The origins of telicity. In *Events and grammar*, ed. Rothstein, S., 197–237. Dordrecht: Kluwer.
- Malaia, E., Wilbur, R.B., and C. Weber-Fox. this volume. Event end-point primes the undergoer argument: a look at neurobiological bases of event structure processing. In *The subatomic semantics of event predicates*, eds. Arsenijevic, B., Gehrke, B., and R. Marín. Amsterdam: Springer.
- Moens, M., and M. Steedman. 1988. Temporal ontology and temporal reference. *Computational Linguistics* 14(2):15–28.

- Pickering, M.J., McElree, B., Frisson, S., Chen, L., and M.J. Traxler. 2006. Underspecification and coercion. *Discourse Processes* 42(2):131–155.
- Piñango, M.M., Zurif, E., and R. Jackendoff. 1999. Real-time processing implications of enriched composition at the syntax-semantics interface. *Journal of Psycholinguistic Research* 28(4):395–414.
- Piñango, M.M., Winnick, A., Ullah, R., and E. Zurif. 2006. Time-course of semantic composition: the case of aspectual coercion. *Journal of Psycholinguistic Research* 35:233–244.
- Pylkkänen, L., and B. McElree. 2006. The syntax-semantic interface: on-line composition of sentence meaning. In *Handbook of Psycholinguistics*, vol. 2, eds. Traxler, M., and M.A. Gernsbacher, 537–577. New York: Elsevier.
- Rayner, K. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin* 124(3):372–422.
- Staub, A., Chuck Clifton, J., and L. Frazier. 2006. Heavy np shift is the parser's last resort: evidence from eye movements. *Journal of Memory and Language* 54(3):389–406.
- Todorova, M., Straub, K., Badecker, W., and R. Frank. 2000. Aspectual coercion and the online computation of sentential aspect. In *Proceedings of CogSci 2000*, eds. L. R. Gleitman and A.K. Joshi, 523–528. Philadelphia: Cognitive Science Society.
- Vendler, Z. 1957. Verbs and times. *The Philosophical Review* 66(2):143–160.
- Verkuyl, H. 1993. *A theory of aspectuality*. Cambridge: Cambridge University Press.