The cross-linguistic processing of aspect – an eyetracking study on the time course of aspectual interpretation in Russian and German

Oliver Bott* and Anja Gattnar

Project B1, SFB 833, University of Tübingen, Nauklerstrasse 35, 72074 Tübingen, Germany; Project Composition in Context (CiC), University of Tübingen, Nauklerstrasse 35, 72074 Tübingen, Germany; Project C2, SFB 833, University of Tübingen, Nauklerstrasse 35, 72074 Tübingen, Germany

(Received 17 September 2014; accepted 22 February 2015)

This paper reports a cross-linguistic study on the time course of aspectual interpretation in an aspect language (Russian) and a non-aspect language (German). In Russian, mereological semantics led us to expect incremental mismatch detection independently of the presence or absence of the verbal arguments. In German, however, mismatch effects should be delayed until the processor has encountered the complete predication. These predictions were tested in two eyetracking during reading experiments. We investigated the processing of achievement verbs modified by aspectually mismatching adverbials in Russian (Exp. 1) and German (Exp. 2) and manipulated the word order in such a way that the mismatch occurred before or after the predication was complete. The data show that Russian readers immediately noticed the mismatch independently of whether the verb preceded or followed its arguments, whereas German readers showed mismatch effects only after a complete predication. We take this as evidence for cross-linguistically different increment sizes in event interpretation.

Keywords: incrementality; grammatical aspect; lexical aspect; cross-linguistic differences; German; Russian

Languages differ with respect to the grammatical means they have to encode semantic distinctions. A particularly prominent example is the category of grammatical aspect (e.g., Comrie, 1976). Speakers of an aspect language such as Russian have to choose a verbal form which is imperfective or perfective, that is, there is a formal opposition between perfective (pv) and imperfective (ipv) verbs. This is different for speakers of a non-aspect language like German, which does not have grammatical means to express completedness or ongoingness of an event (see von Stutterheim, Andermann, Carroll, Flecken, & Schmiedtová, 2012). The present paper investigates in which way the grammatical system of a language interacts with the time course of semantic interpretation during online comprehension. We will outline a semantic model for incremental interpretation that allows us to capture cross-linguistic differences of how temporality is encoded in language. Secondly, we will present the results of a cross-linguistic eyetracking study showing that the grammatical system of a language shapes the time-course of semantic interpretation. We will thereby assume that both, speakers of an aspect language as well as speakers of a non-aspect language, conceptually make the same distinctions with the only difference that the former arrive at the interpretation under the guidance of the grammatical system while the latter must use contextual information.

Consider the contrast between the German example (1) and its Russian counterpart in (2).

(1) Peter schrieb einen Brief…
Peter write-past-sg a letter…
“Peter was writing/wrote a letter…”

a. da wurde er durch einen Anruf unterbrochen.
when was he by a phone-call interrupt-past-sg.
“when he was interrupted by a phone call.”

b. und brachte ihn zur Post.
and bring-past-sg it to-the post-office.
“and took it to the post office.”

The German accomplishment (Vendler, 1957) schreiben eines Brief (write a letter) is ambiguous between an ongoing and a complete event interpretation and can be continued with either (1-a) or (1-b) which disambiguate towards an imperfective (ipv) or a perfective (pv) interpretation, respectively. This is impossible in Russian where perfectivity has to be encoded right from the start:

(2) a. Petr písal / *napísal, pis’mo, kogda on Peter write-ipv-past-sg/*write-pv-past-sg letter, when he byl prervan telefonnym zvonkom.
was interrupted by-phone call.
“Peter was writing/*wrote a letter when he was interrupted by a call.”
If an event is incomplete or in progress it can only be reported imperfectively, hence (2-a) is ungrammatical with *napisal, whereas complete events require a perfective verb. Therefore, (2-b) with imperfective *pisal is ungrammatical. Another way to put this cross-linguistic difference is that in an aspect language like Russian the viewpoint on an eventuality is immediately fixed by the grammatical marking of aspect, whereas in a non-aspect language like German the linguistic context plays a crucial role in determining the event representation. Obviously, this difference should have strong implications on how aspect is processed across languages.

Before diving right into our expectations for processing, we have to distinguish certain in principle independent but interacting semantic components in order to properly understand how temporality is expressed in language (cf. Klein, 2009 for an overview). Tense is a grammatical category of the verb which locates the time of a situation in relation to the time of the utterance. It can take the values past, present or future. Grammatical aspect denotes the viewpoint from which the eventuality is described, for instance, as a complete whole taking an outside perspective on the eventuality or from within the eventuality, e.g., as ongoing (Comrie, 1976). Finally, lexical aspect or Aktionsart serves to distinguish between basic event types including states, processes, accomplishments and achievements (Vendler, 1957). Although these three components are thought to be semantically universal, they vary a lot between languages with respect to the degree to which they are grammaticalized (e.g., Smith, 1997). Whereas in Russian both aspects are grammaticalized, English shows opposition between the grammaticalized progressive and the simple forms which are open to different aspectual interpretations, and German has no grammatical means to mark completed- or ongoingness. Thus, German and Russian are on the opposite ends regarding the marking of grammatical aspect.

Verkuyl (1972) was the first to investigate influences of the verbal arguments on the lexical aspect of the complete predication. He noticed that in a non-aspect language like English lexical aspect not only depends on the verb, but, crucially, that the arguments enter into aspectual composition (see also Bach, 1986; Krifka, 1998 and the dissertation by Champollion, 2010). This is illustrated for English in (3-a) to (3-c). Whether a sentence expresses an activity or an accomplishment is commonly diagnosed by the compatibility of a sentence with time span adverbials and durative adverbials, respectively (cf. Dowty, 1979, p. 69 for a list of linguistic tests). Thus, whereas (3-b) is telic and expresses an accomplishment, (3-a) and (3-c) are atelic activities. This suggests that the aspectual interpretation of the complete VP largely depends on the semantic properties of the nominal arguments.

$$\text{If an event is incomplete or in progress it can only be reported imperfectively, hence (2-a) is ungrammatical with *napisal, whereas complete events require a perfective verb. Therefore, (2-b) with imperfective *pisal is ungrammatical. Another way to put this cross-linguistic difference is that in an aspect language like Russian the viewpoint on an eventuality is immediately fixed by the grammatical marking of aspect, whereas in a non-aspect language like German the linguistic context plays a crucial role in determining the event representation. Obviously, this difference should have strong implications on how aspect is processed across languages.}$$

$$\text{Before diving right into our expectations for processing, we have to distinguish certain in principle independent but interacting semantic components in order to properly understand how temporality is expressed in language (cf. Klein, 2009 for an overview). Tense is a grammatical category of the verb which locates the time of a situation in relation to the time of the utterance. It can take the values past, present or future. Grammatical aspect denotes the viewpoint from which the eventuality is described, for instance, as a complete whole taking an outside perspective on the eventuality or from within the eventuality, e.g., as ongoing (Comrie, 1976). Finally, lexical aspect or Aktionsart serves to distinguish between basic event types including states, processes, accomplishments and achievements (Vendler, 1957). Although these three components are thought to be semantically universal, they vary a lot between languages with respect to the degree to which they are grammaticalized (e.g., Smith, 1997). Whereas in Russian both aspects are grammaticalized, English shows opposition between the grammaticalized progressive and the simple forms which are open to different aspectual interpretations, and German has no grammatical means to mark completed- or ongoingness. Thus, German and Russian are on the opposite ends regarding the marking of grammatical aspect.}$$

$$\text{Verkuyl (1972) was the first to investigate influences of the verbal arguments on the lexical aspect of the complete predication. He noticed that in a non-aspect language like English lexical aspect not only depends on the verb, but, crucially, that the arguments enter into aspectual composition (see also Bach, 1986; Krifka, 1998 and the dissertation by Champollion, 2010). This is illustrated for English in (3-a) to (3-c). Whether a sentence expresses an activity or an accomplishment is commonly diagnosed by the compatibility of a sentence with time span adverbials and durative adverbials, respectively (cf. Dowty, 1979, p. 69 for a list of linguistic tests). Thus, whereas (3-b) is telic and expresses an accomplishment, (3-a) and (3-c) are atelic activities. This suggests that the aspectual interpretation of the complete VP largely depends on the semantic properties of the nominal arguments.}$$
examples such as (6) semanticists have questioned for a long time whether any verb should be viewed as being inherently telic (see, e.g., Champollion, 2010; Krifka, 1989 and the discussion of achievements modified by for in van Geenhoven, 2004). According to the cumulativity hypothesis (e.g., Krifka, 1992), any verb, including achievement verbs start out as atelic and end up in a telic hypothesis (e.g., Krifka, 1992; van Geenhoven, 2004). According to the cumulativity hypothesis (e.g., Krifka, 1992), any verb, including achievement verbs start out as atelic and end up in a telic hypothesis (e.g., Krifka, 1992; van Geenhoven, 2004). According to the cumulativity hypothesis (e.g., Krifka, 1992), any verb, including achievement verbs start out as atelic and end up in a telic hypothesis (e.g., Krifka, 1992; van Geenhoven, 2004).

(6) a. Mary discovered fleas on her dog for six weeks.  
   (van Geenhoven, 2004, p. 136)

   b. John found his son’s tricycle in the driveway for six weeks.  
   (Dowty, 1979, p. 82)

   c. Visitors arrived all night.  
   (Steedman, 1997, p. 902)

Although the present paper is mainly concerned with aspectual composition of the complete VP and not so much with the aspectual properties of the individual verb, we will assume that the frequentative readings illustrated in (6) do involve aspectual coercion from a single event interpretation to an iterative interpretation (for an alternative proposal see Deo & Pinango, 2011 and the response in Champollion, 2013). Psycholinguistic evidence for iterative coercion in cases like these – also setting them apart from the commonly studied cases of iterative semelfactives (for an overview see, e.g., Bott, 2010; Brennan & Pylkkänen, 2008; Pylkkänen, Brennan, & Bemis, 2011) – shows that iteration of telic event types leads to a particularly strong increase in processing load (Bott, 2008).

The paper is structured as follows. First we will provide a characterization of perfective and imperfective aspect in Russian. We will then briefly introduce the relevant notions from formal semantics and discuss existing experimental work on aspectual interpretation. We will close the section with processing implications for the present study. Section 2 presents the results of an eyetracking experiment during reading which examined mismatch detection in sentences involving adverbial modification of complete and incomplete predications in Russian (Exp. 1). Section 3 presents a parallel experiment in German (Exp. 2). Finally, sections 4 and 5 present two completion studies that tested what kind of sentence continuations German and Russian comprehenders predict for the sentence beginnings tested in the eyetracking experiments (Exp. 3 & 4). The final section concludes the paper.

1.1. Imperfective and perfective verbal aspect in Russian

For those unfamiliar with Russian we will provide some background on Russian aspect. Verbal aspect in Russian is a grammatical category that encodes the opposition of imperfective (ipv) versus perfective (pv) aspect. This opposition is reflected by morphologically linked pairs of verbs with different functions. There are three ways to express aspect in Russian morphologically (cf. Breu, 2009, p. 211). For our purpose only the case of the so called secondary imperfectivization is relevant. (7-a) illustrates the suffix -yv(a)- used to express the secondary imperfective. Stem variation in (7-b) is another option. 34 out of the 36 verbs that were used in the present study are derived by secondary perfectivization. Note that the aspect marker only occurs at the end of the verb.

(7) a. otkryt’pv / otkryvat’ipv (open)
   b. zarnetit’pv / zamečat’ipv (notice)

Comrie (1976) and Smith (1997) distinguished situations as bounded or unbounded, telic or atelic. A simple characterization of the function of pv aspect is as follows: pv aspect is used to refer to single, completed (temporal closed) actions (8-b). It allows us to view an event as a whole, whereas ipv aspect refers to processes (8-a), states (8-c) and habitual states or iterative events (8-d). Ipv aspect thus focuses on the internal stages of an eventuality:

(8) a. On pisalipv pis’mo.  
   He write-ipv-past-sg letter.  
   “He was writing a/the letter.”

   b. On napisalipv pis’mo.  
   He write-pv-past-sg letter.  
   “He wrote a/the letter.”

   c. Ona znalaiipv otvet.  
   She know-ipv-past-sg answer.  
   “She knew the answer.”

   d. Na užin ja pokupaliipv sebe obyčno kolbasu.  
   For dinner I buy-ipv-past-sg me usually sausage.  
   “For dinner I bought me usually a sausage.”

Russian aspect plays a very prominent role in slavistics. We will mention only the most prominent semantic characterizations of pv and ipv aspect here. All of them explain the aspectual difference by an opposition to have a certain property or not. The situation is presented:

- as completed versus not completed,
- in its totality versus in its stages and
- with the presence versus the absence of an (internal) boundary (cf. Klein, 1995, p. 673ff.).

In each case, the perfective aspect possesses the particular property, whereas the imperfective aspect is characterised...
by a lack of the respective properties. There has been a long tradition of discussion on the interaction between grammatical aspect and lexical aspect. One issue is the interaction between grammatical aspect and the Vendlerian lexical classes. In this discussion, pv aspect has been taken to express accomplishments (predel’nye glagoly: Paducheva, 2009, p. 9) and achievements (momental’nye glagoly: Paducheva, 2009, p. 9), while the ipv aspect has been assumed to denote states and activities (Breth, 1985; Paducheva, 2009).

(9) a. Ivan el\_pv
   \textit{“Ivan ate.”}

b. Ivan s’el\_pv jabloko.
   \textit{“Ivan ate the apple.”}

c. Ivan el\_pv jabloki.
   \textit{“Ivan ate/was eating apples.”}

Paducheva (1996), in contrast to Breth, proposed that imperfective verbs include accomplishments, too. Imperfective accomplishments “describe an ongoing process that aims towards reaching its inherent limit, resulting in a change in the direct object”. In contrast, the perfective accomplishments denote that the inherent limit is already reached “resulting in a change in the direct object” (Braginsky & Rothstein, 2008, p. 10). Perfective and imperfective accomplishments form so called “bounded pairs”:

(10) a. Ivan s’el\_pv jabloko za dva časa.
   \textit{“Ivan ate the apple in two hours”}

b. Ivan el\_pv jabloko dva časa.
   \textit{“Ivan was eating an apple for two hours”}

As illustrated above in (4) and (5), aspectual languages exhibit a close correlation between indeterminate and determinate nouns and the use of ipv and pv aspect. As in other slavic languages pv aspect correlates with the definiteness of the object in Russian, too (Anstatt, 2003, p. 51).

(11) a. Mal’čik el\_pv sup.
   \textit{The boy was eating soup.}

b. Mal’čik s’el\_pv sup.
   \textit{The boy ate the soup.}

Sentence (11-a) expresses an atelic activity or a progressive, whereas (11-b) refers to a telic event. The examples illustrate the well-known correlation of the pv aspect with the definiteness of the direct object (Birkenmaier, 1979; Dahl, 1985).

1.2. Event semantics and incremental interpretation

This section presents an informal description of the semantics of tense and aspect in aspect and non-aspect languages (for a compositional analysis of tense and aspect across languages see Paslawska & von Stechow, 2003).

Mereological semantics serves as our starting point. This type of semantics explains the intricate interdependency between lexical aspectual classes and the semantic contribution of the arguments. These factors are brought together by a close linkage between the semantics of events, times and objects, in particular whether a verbal or nominal predicate has cumulative or quantized reference. According to this view, both events and objects include plurals and mass entities. All entities are assumed to stand in the parthood relation, usually taken as primitive (Link, 1983).

Figure 1 illustrates a part-whole structure. The ontological domain can thereby consist of an algebraic structure over objects, events or time intervals. For instance, if \(a, b\) and \(c\) denote amounts of water, Figure 1 exemplifies the semantics of the predicate \textit{water}. Whenever two entities are in the denotation of \textit{water}, their sum is, too. The same holds for events. If \(a, b\) and \(c\) are jogging events by John, then any of their sums is in the denotation of \textit{John is jogging}, too. Similarly, time intervals can form a mereology.

The mereological framework allows us to capture structural analogies and interactions between the denotations of verbal and nominal predicates. Crucially, it allows us to define second order properties which are crucial for the semantics of aspect and temporal modifiers such as \textit{for}-adverbs. The properties relevant for us are whether a predicate is quantized or homogenous, i.e. cumulative and divisible (Kripka, 1998). Here are the definitions:

(12) a. A predicate is quantized iff, whenever it holds of something, it doesn’t hold of any of its parts.

b. A predicate is cumulative iff the predicate applies to at least two distinct entities and, whenever it holds of two elements, it also holds of their sum.

c. A predicate is divisive iff, whenever it holds of something, it also holds of its (proper) parts.

d. A predicate is homogeneous iff it is both cumulative and divisive.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{An illustration of a mereology, a join semilattice with the bottom element removed. Edges of the graph correspond to proper parthood (<). \(\oplus\) corresponds to sum formation. Parthood and sum formation are interdefinable concepts.}
\end{figure}
Bare mass and bare plural nominal predicates pattern with
state and process predicates, i.e. they are homogeneous.
By contrast, count nominal predicates pattern with accom-
plishments (e.g., build a house) and achievements (e.g.,
reach the top) in so far as they are quantized. These
second order properties can also be used to capture
the semantic contribution of durative adverbials of the kind
for three hours. They take an event predicate as their
argument and constrain their runtime to the value speci-
fied in the measure phrase, for instance, three hours.
Furthermore, they presuppose that the predicate is divis-
ive. The presupposition is responsible for the fact that for-
adverbials are generally bad with quantized event pre-
dicates, which, per definitionem, are incompatible with
divisive event structures.

The property of being quantized turns out to be crucial
for perfectivity, as well. Following Križka (1992) we will
assume that whatever perfectivity may exactly be, it is a
fact that perfective aspect can only be assigned to verbal
predicates that are quantized, or rather, verbal predicates
that are telic. This entailment immediately explains why
aspect disambiguates the interpretation of nominals which
are ambiguous between a bounded and an unbounded
interpretation, cf. (4), (5) and (11).

Finally, for non-aspect languages it has been observed
that a predication only receives a quantized/telic event
interpretation if all verbal arguments are quantized
(Križka, 1992). Thus, lexical aspect is in fact not a
property of verbs, but of complete verb-argument struc-
tures. As a consequence, whether a VP is quantized or
cumulative and hence compatible with a for-adverbial can
only be decided after the predication is complete. This
describes the facts in a non-aspect language like German.
However, the situation is entirely different in an aspect
language like Russian. Here, the perfective aspect marking
of the verb is sufficient to provide a cue that the
eventuality expressed by the sentence is quantized. As a
consequence, the verb marked for aspect should be
enough to yield aspectual effects in connection with
temporal modifiers even in the absence of the arguments.
Obviously, this should affect the time course of aspectual
interpretation. Whereas aspectual modification in aspect
languages can be processed in a “verb-based manner”,
non-aspect languages are expected to be processed in a
more “argument-based manner”. After reviewing the for
our purposes relevant existing experimental work on the
processing of lexical and grammatical aspect, we will
make this prediction more precise.

1.3. Previous experimental studies on aspectual
interpretation and predictions for the present study

The semantic considerations in the preceding section
motivate why we think that the time course of aspectual
interpretation should be different in an aspect language
like Russian from that in a non-aspect language like
German. However, there are no studies that have system-
atically compared the time course of aspectual interpreta-
tion cross-linguistically.

Grammatical aspect has recently received increased
interest in psycholinguistics (for a comprehensive review
see Madden & Ferretti, 2009). An interesting line of
studies has investigated the accessibility of event partici-
pants in English progressive and simple sentences (e.g.
Ferretti, Kutas, & McRae, 2007). The experimental
findings provide evidence that the progressive presents
an event from the inside, making participants, instruments,
and places fully accessible, whereas the simple forms
present events as complete units with the event partici-
pants becoming less accessible. Similarly, it has been
shown that verbal aspect bears an influence on the
accessibility of event participants during discourse com-
prehension (Carreiras, Carriedo, Alonso, & Fernández,
1997). Using picture selection/verification Madden and
Zwaan (2003) found that participants were faster and more
likely to choose a picture showing a complete event rather
than a picture depicting an ongoing event after they had
read a sentence with an accomplishment verb in the
simple past. This indicates that English speakers interpret
accomplishments in the progressive as ongoing, but events
reported in the simple past as completed. This finding was
supported by a study by Anderson, Matlock, Fausey, and
Spivey (2008). They compared accomplishments describ-
ing a path such as Tom jogged/was jogging to the woods
and then stretched when he got there. Participants listened
to these sentences while they had to use the computer
mouse to drag and drop a human character in a visual
scene. When they heard a sentence in the past progressive,
many drops took place at the beginning and the centre of
the path, whereas in the simple past most drops were at
the end of the path. Flecken and Gerwien (2013)
investigated the interaction of perceived event duration
and viewpoint aspect in Dutch by comparing Dutch event
descriptions in the progressive vs. simple form. Their
findings show that the progressive form extends duration
estimations for short events, whereas it shortens the
perceived duration of inherently medium and long events.
This interaction provides further psycholinguistic evid-
ence for the different functions of progressive aspect and
simple forms, i.e., the progressive seems to give an
“inside” view of the event and focusses on a specific
internal time span of the event, whereas simple forms
provide an “outside perspective”. Taken together, the
studies have revealed that aspectual distinctions lead to
clear cognitive effects how a situation is mentally
represented – much in accordance with the semantic
work presented above.
Turning to cross-linguistic differences, von Stutterheim, Carroll, and Klein (2009) reported the findings of a production study comparing event descriptions of English, German, and Dutch speakers that were elicited while they were watching and reporting a silent video clip. Their study revealed clear differences between productions from the three languages. English speakers used the progressive to start event descriptions well before the endpoint was visible (e.g. *A car is going down a lane... to a farmhouse*). Germans showed a different behaviour relating their descriptions to the endpoint of a motion event. It thus seems as if production is constrained by the language in which it is realised. Recently, von Stutterheim et al. (2012) extended this line of research to a sample of seven languages and showed that the aspectual properties of a language influence how speakers conceptualise events. It is not only that speakers of different languages talked differently about motion events (e.g. mention of endpoints), their language also had a clear influence on their looking behaviour while watching the movies and influenced their memory capacity for those parts of the scene corresponding to endpoints.

Regarding the online comprehension of lexical and grammatical aspect, there are two lines of study which served as starting point for the present paper. The studies of Bott (2010) and Bott (2013) investigated the increment size of aspectual interpretation in German. Besides cases of aspectual coercion (achievements modified by *in x time*), he also investigated the reading times of cases of aspectual mismatch (achievement + *for x time*) and compared them with aspectual controls (achievements + *x time ago*). Consistently across a number of experiments, Bott (2013) observed aspectual coercion and mismatch effects not at the verb region but only after the verb had received its arguments, i.e. interpreters had encountered a minimal complete sentence. Another piece of evidence comes from a self-paced reading experiment (Bott, 2010, Exp.2) which compared reading times of activity biased verbs modified by *in*-adverbials in (13-a) vs. the control condition in (13-b).

(13)  
\[\begin{align*}
\text{a. Peter joggte} & \textit{ in zwanzig Minuten ...} \\
& \textit{Peter jogged in twenty minutes} \\
\text{b. Peter joggte} & \textit{zwanzig Minuten lang ...} \\
& \textit{Peter jogged for twenty minutes...}
\end{align*}\]

The study revealed a local mismatch effect at the adverbial in sentence (13-a) relative to (13-b), even though the sentence beginning with an *in*-adverbial can be perfectly well continued with a phrase introducing a path, such as *fünf Kilometer weit (a distance of five kilometers)*. And in fact, half of the conditions employed such bounded continuations.² Taken together, the just mentioned reading-time studies motivate the importance of the notion of a minimal complete sentence as processing domain. Linking this to our semantic considerations above, a minimal domain corresponds to the minimal incremental unit which allows for a presupposition check whether the predication is quantized.

Bott (2013) interpreted the findings of the reading time experiments as evidence for his late aspectual interpretation hypothesis (LAI).

*Late Aspectual Interpretation (LAI) Hypothesis: Lexical aspect is not computed before the verb has all its arguments.* (Bott, 2013, p. 199)

Without further qualifications, however, this hypothesis seems to run counter to the commonly held belief that language interpretation proceeds incrementally, that is, language comprehenders incorporate each word into the linguistic context as they encounter it (e.g., Marslen-Wilson, 1973, 1975). What is more, recent research has provided rather strong evidence that language comprehension not only proceeds incrementally, but that it even works in a forward-looking manner, that is, the processor anticipates what is most likely to come next even before seeing the input (see, e.g., DeLong, Urbach, & Kutas, 2005; Kamide, 2008). Thus, the LAI hypothesis in itself is not very plausible.

We would therefore like to refine the hypothesis taking into account the above outlined differences in the cross-linguistic grammaticalization of aspect and propose that the comprehension system is geared towards incremental interpretation, but is constrained by the degree to which a semantic distinction is encoded in the grammar. This is captured in Bott and Hamm (2014)’s Cross-linguistic Aspectual Variation Hypothesis (CAV).

*Cross-linguistic Aspectual Variation (CAV) Hypothesis: If a language has the grammatical means to express an aspectual distinction, the processor does immediately commit to an aspectual interpretation. Else, the aspectual interpretation remains underspecified.* (Bott & Hamm, 2014)

Bott and Hamm (2014) conducted a cross-linguistic self-paced reading study in which they compared processing difficulty of aspectual coercion in German and English. They investigated whether the amount of processing difficulty resulting from coercion of an accomplishment (e.g. *build a monument*) into an activity varies between languages depending on whether the language has the grammatical means to mark an eventuality as ongoing (as does the progressive in English) or not (the German system). The study revealed clear coercion costs in
English but not in German lending initial support to the CAV-Hypothesis.

The hypothesis directly follows from our semantic considerations. We have seen that the intuition that lexical aspect is a property of lexical verbs (as, for instance, proposed by Vendler, 1957) runs into problems once we take a closer look at the interactions of verbs and their arguments. Thus, without the guidance of grammatical aspect the bottom-up construction of an event type based on the compositional interpretation of the complete predication is as incremental as we can get. We have further proposed that the situation is different in an aspect language like Russian where grammatical aspect provides a top-down constraint for the upcoming sentence.

The predictions of our proposal can thus be summarised as follows. We expect immediate mismatch detection and hence processing difficulty of aspectual mismatch between adverbial modifiers and verbs in an aspect language like Russian, irrespective of the presence or absence of the verbal arguments. In a non-aspect language like German, however, we expect delayed mismatch effects until the point when interpreters have received a complete verb-argument structure.

2. Processing aspectual information in Russian – Exp. 1

The first experiment tested whether Russian readers show immediate mismatch effects when they encounter a perfective achievement modified by a durative adverbial. We employed a 2 × 2 within design and manipulated the kind of adverbial (Russian for-adverbials vs. control, i.e. ago-adverbials) and the word order of the test sentences (subject verb object adverbial vs. adverbial verb object subject). Besides their different aspect systems, Russian and German are particularly well suited for the aims of our study because word order is relatively free in both languages. In Russian, a word order with an adverb preceding the verb has even been shown to be the unmarked order for adverb placement (Kallestinova, 2007, Exp. 3). Otherwise, SVO order is preferred over other, more marked word orders which need to be licenced information structurally.

2.1. Methods

2.1.1. Materials

We constructed 36 parallel items in Russian and German (see supplemental data and research materials for the complete list of experimental items). A sample item, for convenience transliterated into the Latin alphabet, is provided in (14).³ In the actual experiment all materials were in cyrillic. Asterisks indicate segmentation into regions of interest (ROIs).

All items contained clear instances of achievement verbs such as win, find, reach, notice, comprehend, and so forth. Critical ROIs for the analysis were the verb region, i.e. the third ROI in (14-a)/(14-b) and the second ROI in (14-c)/(14-d), and the adverbial region in (14-a) and (14-b). Each item received a question querying whether the sentence was sensible.

A latin square was used to distribute the experimental sentences over six lists (= six data points per participant and condition). The same 120 fillers (40 non-sensical) were added to each list. Each experimental item and 64 of the distractors received a question querying whether the sentence was sensible. 56 fillers were accompanied by ordinary comprehension questions in order to prevent participants from developing strategies. Non-sensical fillers never received ordinary comprehension questions.

2.1.2. Participants

36 native Russian speakers (27 female; mean age: 27.25 y., sd.: 4.56 y., min.: 20 y., max. 41 y.) took part for a payment of €10. All were students at the University of Tübingen and were tested in Tübingen in the same eyetracking lab as the German participants. On average, they had spent 7.1 y. in Germany. Each of the participants estimated their command of German somewhere between good and excellent. All participants had normal or corrected to normal vision. The data of nine additional participants had to be excluded from the analysis because of calibration problems.
2.1.3. Procedure

A desktop-mounted Eyelink 1000 eyetracker monitored the gaze location of participants’ dominant eye. Participants viewed the stimuli binocularly on a 21 inch monitor 70 cm from their eyes. A head rest minimised head movements. The experiment was implemented using the SR Research Experiment Builder software and eyetracking data were preprocessed and exported with the SR Research Data Viewer.

The trial began with the presentation of a screen which served as calibration check with a little black dot in the position where the centre of the first word would appear. If no fixation was registered within five seconds, recalibration was enforced. Otherwise a sentence appeared in the centre of a navy blue screen in yellow font. Two characters corresponded approximately to one degree of visual angle. After reading participants had to move their eyes to an asterisk at the bottom of the screen which triggered the presentation of the question screen. For the reading stage, there was a time limit of 20 seconds after which recalibration was automatically enforced. There was no time limit for answering the questions. The experiment consisted of a practice of 10 sentences, followed by the experimental trials in two blocks with individually randomised sentences. Before each experimental block the tracker was calibrated using a 3 × 3 grid guaranteeing that all fixations were less than 0.5 degrees apart from the calibration stimuli. An average experimental session lasted approximately half an hour.

2.1.4. Data analysis

Prior to all analyses we preprocessed the data. We checked for major track loss (no failed calibration checks/timeouts in the experimental trials) and excluded all fixations that immediately preceded or followed a blink. The remaining fixations shorter than 80 ms and within one character space of the previous or next fixation were merged with this fixation. We then excluded the still remaining fixations shorter than 80 ms and fixations longer than 1200 ms. This affected 3.8% of all fixations. The reading times in the two word orders were analysed by computing paired t-tests with the independent variable adverbial. To analyse the proportions of regressions and the proportions of “yes, sensible” judgements we computed separate logit mixed effects model (GLME) analyses for each word order including adverbial as fixed effect and the random effects (intercepts and slopes) of participants and items (Barr, Levy, Scheepers, & Tily, 2013; Jäger, 2008). Judgement times were corrected for outliers by removing all data points up to 200 ms and above 2.5 standard deviations of a participant’s mean RT affecting 5.9% of the data.

We carried out analyses of six eyetracking measures. First-pass time is the total time spent in a ROI before the reader moves on or looks back in the text. Last fixation duration is the duration of the last fixation on a ROI regardless of whether it is the only fixation or the last of multiple fixations on it. Last fixation durations are not among the standardly reported measures and, therefore, some motivation may be in order (but see Rayner, Balota, & Pollatsek, 1986). There were two reasons for this. Firstly, as described in the introduction aspect marking of Russian achievements is regularly marked via the secondary imperfective, i.e. a suffix (or lack thereof in the perfective; this was the case in 34 out of the 36 verbs included in our study). As a consequence, to decide upon grammatical aspect Russian readers have to interpret the final part of the verb ROI. The second reason had to do with the German experiment (Exp. 2). Here, we wanted to investigate the possibility that mismatch in the Adv-VSO order affected the preview from the last fixation of the adverbial region upon the verb. Regression-path duration represents the sum of all fixations starting with the first fixation in a region and ending with the first forward saccade past that region. If a ROI was skipped during first-pass, it was counted as a missing value for all three just mentioned measures, respectively. Finally, total time is the sum of all fixations made on a region. We also measured two types of proportions of regressions: first-pass regression ratios, i.e. the proportions of how often readers launched a regression from a region during first pass reading. Again, skipped regions were treated as missing values. The proportion of regressions in a region is a measure of how often it was entered from the right. In the following, we will loosely refer to first fixation duration, first-pass time, regression-path duration and first-pass regression ratios as “early” eyetracking measures and to total times and proportions of regressions in as “late” measures without really suggesting that this distinction reflects well-defined processing stages (see, e.g., Rayner, 1998 for a comprehensive review).

2.2. Results and discussion

The behavioural data showed strong mismatch effects. The for x time mismatch conditions were rejected 91.2% of the time in the SVO-Adv order and 94.4% of the time in the Adv-VOS order. By contrast, controls with x time ago were generally accepted: 78.2% of the time in the SVO-Adv condition and 74.1% of the time in the Adv-VOS order. The GLME analysis revealed only a reliable fixed effect of adverbial (estimate = 4.92, |z| = 10.16, p < .01) but neither order (|z| = .92) nor the interaction were significant (|z| = .18). Thus, the two word orders were judged equally acceptable. Judgement RTs showed a similar pattern. The control conditions had mean judgement RTs of 1267 ms and 1381 ms in the SOV-Adv and the Adv-VOS orders, respectively. The mismatch conditions were on average 240 ms faster with RTs of 1074 ms and 1094 ms in the SVO-Adv and Adv-VOS conditions. ANOVAs revealed only a significant main effect of adverbial (F1(1, 35) = 17.46, MSE = 118509.42, p < .01;
Figure 2. Aspectual mismatch effects in Russian sentences with SVO-Adv order. Means and standard errors (se) of the six eyetracking measures in Exp. 1. S–1, first part of subject (ROI 1); S–2, second part of subject (ROI 2); V, verb (ROI 3); O, object (ROI 4); Adv, adverbial (ROI 5). Note: se computed on the basis of the by participants analyses.

\[ F_2(1, 35) = 19.77, \text{MSE} = 102845.27, p < 0.01 \]. That the mismatch conditions were judged faster than the control conditions already suggests that readers of Russian had noticed the aspectual mismatch well before the end of the sentence which in turn allowed them to anticipate the acceptability judgement.

Figures 2 and 3 provide an overview region by region of the six eyetracking measures in the SVO-Adv and the Adv-VOS orders, respectively. As expected, the SVO-Adv conditions showed the expected mismatch effect starting right at the adverbial region. At the adverbial mismatch led to longer first-pass times relative to control (\( t_1(35) = \)
2.66, \( p < .05 \); \( z_2(35) = 2.03, p < .05 \), longer regression-path durations (\( t_1(35) = 6.21, p < .01 \); \( z_2(35) = 5.08, p < .01 \)) and a significant increase of first-pass regression ratios (GLME: estimate = 1.44, \( |z| = 5.96, p < .01 \)).\(^4\) The analysis of “late” eyetracking measures, i.e. total times and proportions of regressions in, suggests that aspectual mismatch was reflected by regressions back to the verb but not beyond it to reinspect the first two ROIs, too. The first ROI showed even the opposite pattern: The total times in the control condition were (by participants) significantly longer than in the mismatch condition (\( t_1(35) = 2.40, p < .05 \); \( z_2(35) = 1.58, p = .12 \)). Also, the control
condition received more regressions back into the first ROI than the mismatch condition (GLME: estimate = -0.48, \(|z| = 2.04, p < .05\). At the second ROI mismatch and control did not differ either in total times (|t1/2| < 1) or in their respective proportions of regressions in (GLME: \(|z| = .02\). Regressions due to mismatch affected only the verb and the subsequent object ROI: verb ROI (total times: \(t_1(35) = 2.64, p < .05\); \(t_2(35) = 2.61, p < .05\); \% regression in: estimate = .93, \(|z| = 3.71, p < .01\)), object ROI (total times: \(t_1(35) = 1.88, p = .07\); \(t_2(35) = 2.43, p < .05\); \% regression in: estimate = .81, \(|z| = 3.28, p < .01\)). Thus, readers strongly focused on the verb-(object)-adverbial complex when processing aspectual mismatch.

The Adv-VOS conditions showed immediate mismatch effects right at the verb region. Last fixation durations were on average longer in the mismatch condition (294.9 ms) than in the control condition (263.0 ms). In t-tests, this difference was reflected by a significant effect of adverbial (\(t_1(35) = 2.47, p < .05\); \(t_2(35) = 2.92, p < .01\)). Apart from the last fixation effect t-tests of regression-path durations revealed a marginally significant mismatch effect (\(t_1(35) = 1.75, p = .07\); \(t_2(35) = 1.77, p = .09\). The same pattern was present in the analysis of first-pass regression ratios. In the Adv-VOS order, mismatch led to 16.4% regressions out of the verb region during first-pass reading relative to only 23.3% in the Adv-VOS control condition. In most trials, the verb received multiple fixations. The analysis of the first fixation durations of this region revealed no difference between mismatch and control (mismatch: 292.7 ms; control: 296.9 ms; \(t(35) = -.37\)). Taken together with the last fixation effect, this suggests that readers detected aspectual mismatch before leaving the region, but not immediately when entering it.

The following direct object region revealed the same pattern of effects. Mismatch in the Adv-VOS order had a mean regression-path duration of 758.3 ms relative to 582.0 ms in the control condition (SVO-Adv for comparison: 546.0 ms (mismatch) vs. 513.2 ms (control)). This slowdown was accompanied by a 34.3% first-pass regression ratio in the Adv-VOS mismatch condition relative to only 23.3% in the Adv-VOS control condition. In the statistical analyses these differences were reflected by significant mismatch effects in regression-path durations (\(t_1(35) = 3.57, p < .01\); \(t_2(35) = 4.69, p < .01\)) and first-pass regression ratios (GLME: estimate = .59, \(|z| = 2.50, p < .05\)). However, in the analyses of first fixation durations and first-pass times the interaction did not turn out to be reliable (all \(t_{1/2} < 1\)). Taken together this suggests that the mismatch effects at the object region were due to regressions launched from this region during first-pass reading rather than differences in the ease of integrating the direct object.

The analysis of total times and proportions of regressions in complemented the reported analyses of “early” eyetracking measures of the verb and object ROIs. In 63.4% of the trials in the Adv-VOS mismatch condition participants regressed back into the adverbial region relative to 44.4% in the control condition resulting in a significant mismatch effect (GLME: estimate = .94, \(|z| = 3.75, p < .01\)). Furthermore, the analyses of total times showed an interesting pattern. Whereas the adverbial-verb complex was read more slowly in the mismatch condition than in the control condition (adverbial: \(t_1(35) = 3.75, p < .01\); \(t_2(35) = 3.17, p < .01\); verb: \(t_1(35) = 4.76, p < .01\); \(t_2(35) = 4.66, p < .01\)), on the subsequent regions participants spent less time in the mismatch sentences than in the aspectually well-formed controls (object: \(t_{1/2} < 1\); subject (first part, ROI 4): \(t_1(35) = -2.73, p < .01\); \(t_2(35) = -2.43, p < .05\); subject (second part, ROI 5): \(t_1(35) = -3.32, p < .01\); \(t_2(35) = -2.56, p < .05\)). This suggests that after having detected the mismatch at the verb region they were no longer processing the sentence deeply.

To summarise, the findings of the present experiment provide clear evidence that Russian readers experience aspectual mismatch immediately when they encounter a durative adverbial and a perfective achievement verb. Immediate mismatch effects across word orders show that mismatch detection is completely unaffected by the presence or absence of verbal arguments. Thus, the reading time data are fully consistent with our predictions.

3. Processing aspectual information in German – Exp. 1
To investigate aspectual mismatch detection in a non-aspect language we conducted a parallel eyetracking experiment in German. Like Russian, German has relatively free word order and adverbial modifiers can appear before or after the verb and its arguments. If aspectual processing is delayed until the verb-argument structure is complete, we get the following specific predictions. During first-pass reading aspectual mismatch should not cause any delay or regressions out of regions that are encountered before the transitive verb has received both arguments. Only then readers should slow down and/or launch regressions to earlier parts of the sentence. Thus, in the SVO-Adv conditions we expected mismatch effects to immediately show up at the adverbial, whereas in the
Adv-VSO conditions we expected delayed effects of “early” reading time measures (i.e. first-pass times, last fixation durations, first-pass regression ratios and regression path durations) showing up at the object region.

3.1. Methods

In the following, we will only describe those aspects of the experiment in which the German version differed from the Russian version. In all other respects, methods were identical to those of the previous experiment.

3.1.1. Materials

A German sample item in the four conditions is provided in (15). Instead of the Adv-VOS conditions tested in the Russian eyetracking experiment we included Adv-VSO sentences in the German version. This was done because these two word orders correspond to the intuitively unmarked cases with topicalized adverbials in Russian and German.

A latin square was used to distribute the experimental sentences over lists such that each list contained each item under only one condition.5 The same 122 fillers (40 nonsensical) were added to each list. Each experimental item and 62 of the distractors were followed by a question querying whether the sentence was sensible. 60 fillers were followed by ordinary comprehension questions.

To make sure that the predicted delay of aspectual mismatch effects in German Adv-VSO sentences cannot be attributed to processing a marked word order that requires special licencing conditions, we ran a grammaticality judgement pretest and gathered judgements for the two word orders using the thermometer judgement method (Featherston, 2008). We only tested the control conditions, i.e. transitive achievements modified by an ago-adverbial. To find out if both 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 (Featherston, 2008). Figure 4 depicts the mean judgements from 20 German native speakers. The Adv-VSO word order was rated even better than the SVO-Adv condition ($t_1(19) = 2.39, p < .05$; $t_2(35) = 2.65, p < .05$). However, both word orders were in the range of fully grammatical sentences (= fillers of category B).

3.1.2. Participants

36 native German speakers from the University of Tübingen took part in the experiment (29 female; mean age: 25.4 y., sd.: 4.4 y., min.: 19 y., max.: 35 y.). Six additional participants had to be excluded from the analysis due to calibration problems ($N = 4$) or error rates above 40% in the practice ($N = 2$). Participants received a payment of €8. All had normal or corrected to normal vision. Six participants were randomly assigned to each list.

Figure 4. Mean grammaticality judgements (+ 95% confidence intervals computed on the basis of the by participants analysis) for the control condition in the two word orders. Also shown are mean judgements of five categories of normed filler sentences ranging from perceived natural (cat. A) to strongly marked (cat. E).
3.1.3. Data analysis

Again, we excluded all fixations that immediately preceded or followed a blink and excluded all trials with major track losses (0.57% of all trials). The remaining fixations shorter than 80 ms and within one character space of the previous or next fixation were merged with this fixation. The remaining fixations shorter than 80 ms and longer than 1200 ms were excluded. This procedure affected 5.6% of the fixations. Behavioural and eyetracking data were then analysed as in Exp. 1. The outlier correction of judgement RTs affected 2.5% of the data.

3.2. Results and discussion

Participants perceived a strong aspectual mismatch of a similar size as in the previous experiment. The for x time mismatch conditions were rejected 83.7% of the time in the SVO-Adv order and 80.1% of the time in the Adv-VOS order. Controls with x time ago were generally accepted: 80.6% of the time in the SVO-Adv condition and 89.5% of the time in the Adv-VOS order. The GLME analysis revealed a reliable fixed effect of adverbial (estimate = –4.02, |z| = 11.29, p < .01) and of order (estimate = –.82, |z| = 2.64, p < .01) but no significant interaction (|z| = 1.64). The main effect of order was due to a slightly generally higher acceptance of Adv-SVO sentences in accordance with the results of the pretest. In contrast to the previous experiment, judgement RTs did not differ reliably between mismatch and control (effects involving adverbial: all F_{1/2} < 1.5). The former conditions had a mean judgement RT of 1668 ms while the latter had a mean RT of 1703 ms. This lack of difference already indicates that German readers were still considering sensicality of the mismatch sentences when entering the judgement stage.

Figures 5 and 6 present an overview over all six eyetracking measures in the SVO-Adv and in the Adv-VSO word orders, respectively. We will again start with the statistical analyses of the SVO-Adv conditions by only comparing the SVO-Adv mismatch and control conditions. Afterwards we will consider mismatch detection in the Adv-VSO order via pairwise comparisons of the mismatch and the control condition.

The analysis of “early” eyetracking measures in the SVO-Adv word order revealed substantial differences between mismatch and control starting immediately at the adverbial ROI. Here, mismatch led to significantly longer first-pass times than control (mismatch: M = 704 ms, control: M = 581 ms; t(135) = 4.56, p < .01; t(235) = 3.56, p < .01), longer regression-path durations (mismatch: M = 704 ms, control: M = 581 ms; t(135) = 5.49, p < .01; t(235) = 4.85, p < .01) and larger first-pass regression ratios (mismatch: M = 19.0%, control: M = 13.5%; GLME: estimate = .94, |z| = 2.33, p < .05). The analysis of “late” eyetracking measures, i.e. total times and the proportions of regressions in showed that the disruption caused by aspectual mismatch did not lead to rereading earlier parts of the sentence. Total times didn’t differ significantly between mismatch and control for any of the first three ROIs preceding the adverbial (all t_{1} < 1.8; all t_{2} < 1.4) but only for the adverbial region (mismatch: M = 1094 ms, control: M = 831 ms; t_{1}(35) = 5.66, p < .01; t_{2}(35) = 4.17, p < .01).

The analysis of proportions of regressions in revealed no mismatch effects in the first two ROIs, the subject (mismatch: M = 40.7%, control: M = 40.4%; GLME: |z| = .28) and the verb regions (mismatch: M = 35.9%, control: M = 32.8%; GLME: |z| = 1.02). Regressions due to mismatch affected only the third region, i.e. the direct object (mismatch: M = 35.9%, control: M = 25.0%; GLME: estimate = 1.07, |z| = 3.37, p < .01). This is quite different from the Russian version of the experiment where readers reinspected the verbs after having encountered a mismatching adverbial.

In line with our predictions, mismatch detection in the Adv-VSO order did not lead to any effects before the object region (ROI 4). Last-fixation durations on the first ROI were even longer in the control condition than in the mismatch condition (t_{1}(35) = –3.92, p < .01; t_{2}(35) = –2.90, p < .01), thus there was no indication of a mismatch effect during preview from the adverbial ROI. At the following three ROIs including the adverbial last fixation durations did not differ reliably between mismatch and control (all t_{1/2} < 1.25). The same pattern was observed for first-pass times (all t_{1/2} < 1) and regression-path durations (all t_{1/2} < 1.85) from the verb region (ROI 2) up to the adverbial (ROI 4). When integrating the verb, aspectual mismatch or coercion thus went unnoticed in the Adv-VSO conditions. At the object, however, there was a first indication of mismatch detection. The statistical analysis of first-pass regression ratios revealed larger first-pass regression ratios out of the adverbial region in the mismatch condition than in the control condition (mismatch: M = 21.1%; control: M = 13.0%; GLME: estimate = .64, |z| = 2.30, p < .05). That mismatch led to rereading of the sentences was further supported by the statistical analyses of “late” eyetracking measures. The analysis of total times revealed mismatch effects for the first ROI – the adverbial – (t_{1}(35) = 4.55, p < .01; t_{2}(35) = 4.22, p < .01), the verb region (t_{1}(35) = 3.47, p < .01; t_{2}(35) = 3.01, p < .01), the subject region (t_{1}(35) = 2.13, p < .05; t_{2}(35) = 2.41, p < .05) and a marginal effect at the object region (t_{1}(35) = 2.05, p < .05; t_{2}(35) = 1.82, p = .08). The analysis of proportions of regressions in showed that mismatch led to a significant increase of regressive eye-movements to the very beginning of the sentence (GLME: estimate = .68, |z| = 2.76, p < .01). Mismatch triggered 59.7% regressions back into the adverbial region, whereas in the control condition readers regressed back to adverbial in only 46.5% of the trials. Logit mixed effects model analyses revealed
that the mismatch effect did neither extend to the verb region (ROI 2: mismatch: $M = 39.3\%$; control: $M = 34.4\%$; $|z| = 1.06$) nor the subject region (ROI 3: mismatch: $M = 21.8\%$; control: $M = 22.0\%$; $|z| = .12$).

Taken together it thus seems as if German readers detected an aspectual mismatch only after they had processed a complete minimal sentence which in turn led them to reread the sentence once again. Thus, in a non-
aspect language like German Aktionsart can only be determined after the verb-argument structure is complete. Aspectual processing in a non-aspect language seems in fact to require processing units with an increment size larger than single words. Our data provide evidence that the increment size of aspectual processing may be a minimal complete sentence since this was the earliest point where we observed aspectual mismatch effects. After they had integrated the direct object in the Adv-VSO order, readers started to launch regressions back to the beginning.
of the sentence. However, the conclusion that aspectual processing in a non-aspect language is delayed until a minimal complete sentence has been encountered may be a little premature. This is because there is an alternative, incremental explanation of the German eyetracking data.

Investigating adverbial modification in yet incomplete verb-argument structures raises the question whether readers might be able to predict an argument compatible with the adverbial? Consider (16-a) with the two continuations in (16-b) and (16-c).

(16) a. Ganze zwei Stunden erreichte …
   Whole two hours reach-past-sg …
   b. *… der Bergsteiger den Gipfel.
   *… the mountaineer the mountain-top.
   c. … der Bergsteiger keinen Gipfel.
   … the mountaineer no mountain-top.

As (16-c) shows, (16-a) can be continued in a meaningful way, although the most typical continuation of a yet incomplete achievement in (16-b) yields a semantically ill-formed sentence. When the processor encounters the sentence beginning in (16-a), it may predict material that is yet to come (see, e.g., Kamide, 2008 for predictive processing). The predictive capabilities of the parser are thus absolutely crucial and lead to different expectations about when processing difficulty emerges in sentences like (16-a) with the semantically anomalous continuation (16-b).

Let us for the moment assume that aspectual processing is predictive in nature and that the sentence processor is able to consider the complete range of possible arguments yet to come. It will then interpret the incomplete sentence in (16-a) with the expectation of a continuation like (16-c).

As a result, including the adverbial, the sentence fragment is predicted to be well formed. Only when a continuation like (16-b) is encountered, is the expectation disconfirmed and processing difficulty emerges. Thus, we would expect delayed processing precisely because of incremental interpretation with anticipation of the right kind of argument (including negation, bare plurals etc.) making the sentence well-formed.

In the present paper we have implicitly entertained a non-incremental theoretical alternative, namely that the processor does not incrementally develop expectations about a likely continuation that aspectually fits the sentence up to this point. Instead, we interpreted the results of the present experiment as evidence for delayed processing in non-aspect languages.

4. Anticipation of semantically appropriate arguments in German – Exp. 3

To decide between these theoretical alternatives, we conducted a continuation experiment and measured the interpretation of incomplete sentences like (16-a). A prediction based account would let us expect that readers have no difficulty to identify a sensible continuation. If, however, they are not able to provide a continuation, this would provide clear counter-evidence against such an account. The latter result would, on the other hand, be fully compatible with the delayed aspectual processing account proposed in this paper. This is because comprehenders that try to come up with a sensible continuation may be mislead by strongly associated arguments like “the top” such that more unusual and infrequent continuations of the kind illustrated in (16-c) do not become accessible to them.

4.1. Methods

The present experiment employed an offline continuation task with no time pressure. This ensured that participants had the opportunity to find the most sensible continuation. If they are not able to come up with arguments that change lexical aspect to fit the requirements of an otherwise mismatching adverbial in an offline task like this, it is even less likely that they are able to do so during real-time comprehension. The present experiment has already been reported in Bott (2010, Exp.7).

4.1.1. Materials

The first 30 items from Exp. 2 were tested in the aspectual mismatch condition: an achievement combined with a for-adverbial. We therefore truncated the experimental sentences and constructed the three types of prompts in (17-a)–(17-c).

   The mountaineer( s) reach-past-sg/pl two hours long …
   zwei Stunden lang … For two hours reach-past-sg/pl …
   The top reach-past-sg/pl for two hours …
   b. Den GipfelAcc. erreichte_sg/erreichten_pl.
   The mountaineer( s) reach-past-sg/pl two hours long …
   for two hours reach-past-sg/pl …
   c. Zwei Stunden lang erreichte_pl/ erreichten_pl …
   For two hours reach-past-sg/pl …
   for two hours reach-past-sg/pl …

In (17-c) the bare verb is combined with the adverbial (Adv-V). In this condition readers have maximal freedom in choosing the appropriate arguments to satisfy the input requirements of the adverbial. Example (17-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 (SV-Adv). In (17-b) the case-disambiguated object den Gipfel is realised preverbally in topicalized position, but the sentence lacks a subject (OV-Adv). We included all three conditions to see how constraining our subjects and objects were and whether there is a subject/object asymmetry in aspectual composition, a question that has been
sometimes addressed in the theoretical literature (e.g., Verkuyl, 2005).
If aspectual processing is predictive as outlined above, the number information of the verb might already provide a cue to what is yet to come. The typical examples proving an aspectual semantic influence of the arguments involve cases with bare plurals (e.g. *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 in a factorial $3 \times 2$ (word order $\times$ number) within design.

Additionally, 40 distractors were included in the experiment. 30 of them allowed for a sensible continuation while 10 clearly did not. The latter contained tense violations like morgen kam… (tomorrow came…) and aspectual violations of different sorts 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 using a latin square design.

4.1.2. 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 provided with the opportunity to reject the sentence as lacking a sensible continuation.

60 German native speakers (23 female; mean age: 29.4 y., min.: 19 y., max.: 56 y.) took part in the experiment. None of them had participated in the previous experiment. Among them, six prizes of 50€ were distributed by lot. Participants were randomly assigned to lists (10 participants per list). The experiment was administered over the internet using WebExp2 (Mayo, Corley, & Keller, 2006).

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.

4.2. Results and discussion
Figure 7 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 generally 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 a logit mixed effects model analysis this difference revealed a significant effect of sentence condition (contrast OV-Adv vs. Adv-V: estimate $= 1.07$, $|z| = 2.71$, $p < .01$; contrast SV-Adv vs. Adv-V: $|z| < 1$). The main effect was due to the fact that the OV-Adv sentences received significantly less continuations than the SV-Adv and the Adv-V sentences. This difference was mediated by number, as shown by significant interactions between OV-Adv vs. Adv-V and number (estimate $= -1.10$, $|z| = 3.31$, $p < .01$) as well as between SV-Adv vs. Adv-V and number (estimate $= -0.71$, $|z| = 2.32$, $p < .05$). Separate analyses for the three sentence conditions revealed that these interactions were due to the fact that participants produced significantly less continuations in the singular than in the plural Adv-V condition (estimate $= 1.23$, $|z| = 4.13$, $p < .01$). This was different in the SV-Adv and the OV-Adv conditions, where no reliable effect of number was observed (both $|z| < 1.3$).

The present 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 adverbal. The findings indicate that in most of the trials this was not the case. The initial part of sentences containing an achievement which is modified by a for-adverbial were by and large judged as nonsensical. This suggests that readers just predict lexical material on an associative basis without deep aspectual analysis. As it seems, comprehenders are generally unable to make use of the full set of combinatorial possibilities but rely on relatively superficial lexical associations.

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 had difficulty to predict the right kinds of arguments in an offline task like this it is even less likely that readers engage in predictive aspectual
processing during ordinary comprehension. The results of the present experiment thus confirm our interpretation of the eyetracking data from the previous experiment.

Nevertheless, the findings show that 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 important than the external argument in line with what has been claimed in the theoretical literature.

Interestingly, the number information of the verb did not have a uniform influence on the ability to predict material that is yet to come. In the OV-Adv and, unsurprisingly, in the SV-Adv conditions, participants were as likely to provide a sensible continuation when the verb had plural morphology as when it was singular. In the Adv-V conditions, however, participants used number information as a guide to a sensible continuation. On the basis of the few continuations provided, we do not have a ready explanation for this pattern of results and have to leave this an issue for future research.

To sum up, German comprehenders generally seem to be unable to come up with a sensible continuation for sentences with an achievement verb and a for-adverbial. Nevertheless, even though finding a continuation only occurred rather rarely, it was not impossible to provide sensible continuations for the experimental items. Under the right circumstances, comprehenders can even use grammatical cues, such as the number inflection of the verb as a guide to the right kind of continuation.

5. Offline comparison with Russian – Exp. 4

Obviously, this should be different in an aspect language like Russian, where no perfectly marked verb – no matter of its lexical semantics – should be compatible with a for-adverbial. Thus, there should be no sensible continuations, whatsoever. To complete the picture, we ran a parallel continuation study in Russian to test whether this prediction is borne out.

5.1. Methods

Methods were the same as in the previous experiment.

5.1.1. Materials

The 36 items from Exp. 1 were tested in the aspectual mismatch condition (18-a) and in the control condition (18-b). Again, we only presented the initial part of the Adv-V sentences and asked whether there is a sensible continuation.9

(18) a. Celyx polčasa dostigpv …
    For half an hour reach-pv-past-sg …

b. polčasa nazad dostigpv,…
    Half an hour ago reach-pv-past-sg…

The sentence (18-a) with a perfective verb modified by a for-adverbial involves aspectual mismatch. Hence, readers should not be able to come up with a sensible continuation. Obviously, this should be different in the control condition (18-b).

In addition, 54 fillers were included. 36 clearly allowed a continuation while 18 did not. The latter again included tense violations like zavtra znal …(tomorrow knew …).

5.1.2. Participants

27 native Russian speaker (22 female, mean age 31.14 y., min 21 y., max 56 y.) took part in the experiment. None of them had participated in Experiment 1. The experiment was a questionnaire administered over the internet. It was implemented using OnExp (http://onexp.textstrukturen.uni-goettingen.de).

5.2. Results and discussion

The experimental items in the mismatch condition were generally rejected. Only in 8.3% of the cases participants found a possible completion. Moreover, the analysis of the provided completions revealed that without a single exception all of the cases in which participants came up with a sensible continuation involved a shift in meaning of the for-adverbials from an adverbial modifier into an argument of the verbal phrase. Let us give an example of this interpretation:

(19) [Celyx sorok minut]acc polučilpv, dlja rešenija zadanija po matematike.
    The whole forty minutes get-pv-past-sg for the solution of the task in mathematics.

“He got forty minutes to solve the math exercise.”

In the control condition (18-b) participants provided sensible continuations 95.7% of the time. The even more pronounced difference between rejection rates in the mismatch and the control conditions in the present relative to the previous experiment indicates that in Russian, other than in German, the combination of a perfective achievement with a for-adverbial is in fact ungrammatical.

6. General discussion

We considered semantic differences between aspect and non-aspect languages and derived cross-linguistic
predictions about the time course of aspectual mismatch detection. These predictions were tested in two closely parallel eyetracking during reading experiments in Russian and German. The observed online processing of aspectual mismatch in Russian and German supports our claim that the grammatical marking of aspect has a clear impact on the relative time course of semantic processing.

Following Krifka (1992) and others we hypothesised that perfective aspect disambiguates the interpretation of the nominal arguments in aspect languages, whereas in non-aspect languages the opposite direction of influence is found. We therefore predicted that aspectual mismatch detection between (perfective) achievements and for-adverbials requires incremental units of different sizes in aspect as opposed to non-aspect languages. Whereas the former allow mismatch detection on the sole basis of the perfective verb and the adverbial, the latter were predicted to depend on a complete verb-argument structure because both, the verb and its arguments, bear an influence on lexical aspect. This is exactly what we observed in online processing. Russian readers showed aspectual mismatch effects immediately at the verb or adverbial region independently of whether the verb preceded or followed its nominal arguments. German readers behaved rather differently. Aspectual mismatch effects were delayed until the predication was complete, i.e. the obligatorily transitive achievement verbs had received their subject and direct object arguments, respectively. We considered two alternative explanations for these delayed effects. The first was a prediction based account according to which the combination of a German for-adverbial with a German achievement verb should trigger the expectation of an upcoming non-quantized argument. When encountering quantized arguments later on, however, this prediction is disconfirmed and aspectual mismatch occurs. Hence mismatch detection appears to be late, although aspectual processing operates in a forward-looking manner. The results of Exp. 3 disconfirms this kind of explanation. German readers were generally unable to come up with a sensible continuation for the experimental items. This leaves us with the second theoretical alternative according to which the increment size of aspectual processing varies across languages. In a non-aspect language like German the relevant increment size for lexical aspect is bigger than the lexical verb and consists of a minimal complete sentence.

What may be the reason for this? We think that the answer is an optimal adaptation of the processing to the grammatical system of the language at hand. One of the main driving forces for incremental interpretation is the limitation of working memory capacity calling for immediate integration of the incoming information into the existing representation (Marslen-Wilson, 1973; Frazier, 1987). On the other hand, it is received wisdom that reanalysis is costly. We think that having an increment size larger than the lexical verb is the optimal solution for the trade-off between these two opposing processing constraints. This may seem counterintuitive in the case of achievement verbs since they constitute almost unambiguous cases of telic events as shown in Exp. 3. However, if we consider other verbs, too, we find that they are used in sentences broadly distributing across lexical aspectual classes. As an example, we would like to contrast two examples from a corpus study reported in Bott (2010, ch.4). The German verb bauen (build) may be thought of as a prototypical example of an accomplishment verb given the prominence of build a house as an example of an accomplishment in the semantic literature. However, if we look at its actual aspectual distribution in a random sample of 200 occurrences from the German Cosmas Corpus (http://www.ids-mannheim.de/cosmas2/) we find that in only 57% of the cases it was used in accomplishment sentences (32% activities, 8% achievements, 3% statives). Moreover, we might think that backen (bake) should show a very similar distribution since both verbs are verbs of creation (in VerbNet (Kipper Schuler, 2006), for instance, they share the same class (build 26.1-1)). The actual distribution, however, turns out to be very different from that of build: bake occurred only 20% in accomplishment sentences, 61% in activity sentences and 19% in statives. This indicates that we may not have reliable intuitions about the most frequent lexical aspect of a given verb. Obviously, more systematic corpus work needs to be done looking at large sets of verbs, but, nevertheless, this already might give an impression how often aspectual reanalysis would be called for if lexical aspect was already encoded in the verb and not the result of aspectual composition. The facts are completely different in Russian. The reason is that any perfective verb, no matter of its lexical semantics, is incompatible with a for-adverbial.

Do the findings of our study imply that there is no event interpretation before having processed the complete VP in a non-aspect language? To be clear, this question must be answered in the negative. We know from quite a number of online experiments that thematic roles of event participants are assigned incrementally, no matter whether the language under investigation is an aspect language or not (e.g., Knoeferle, Crocker, Scheepers, & Pickering, 2005). Similarly, grammatical aspect has been shown to yield immediate effects, for instance, the priming of typical locations which only occurs in progressive sentences but not in sentences with the same verb in the simple past (e.g., Ferretti et al., 2007). What our results suggest, however, is that lexical aspect or Aktionsart is not determined at the lexical level, but depends on a complete verb-argument structure. Furthermore, our study shows that grammatical differences between languages lead to clear differences how these languages are processed during online comprehension.
Acknowledgements
The authors would like to thank Polina Berezovskaya, Fritz Hamm, Robin Hörmig, Janina Radó, Fabian Schlotterbeck, Wolfgang Sternewald and the participants of the Oberseminar WS 2014/15 of Sigrid Beck and Sonja Tiemann for comments on earlier versions of this paper. Thanks also go to Manuel Carreiras and an anonymous reviewer of Language, Cognition and Neuroscience. We would also like to thank Anna Schmid, Anna Pryslopska and Johanna Heininger for data collection and Salome Dimmer for commenting on an earlier draft of this paper. The usual disclaimer applies – all remaining errors are the responsibility of the authors.

Disclosure statement
No potential conflict of interest was reported by the authors.

Funding
This research was funded by the German Science Foundation within projects B1 and C2 of the Collaborative Research Centre SFB 833 “The Construction of Meaning” and project Composition in Context (CiC, XPrag.de).

Supplemental data and research materials
The underlying research materials (i.e. the full set of Russian and German materials) for this article can be accessed at http://hdl.handle.net/11022/0000-0000-4D88-1@ds1

Notes
1. Perspective is, of course, a purely metaphorical description which has to be made more precise (cf. Klein, 1994, 2009). We will nevertheless stick to this metaphor since it can be made precise.
2. Besides clear activity verbs such jog (which all can be used intransitively) the experiment also tested accomplishment verbs and the local mismatch effect took the form of an interaction between verb type and adverbial. For details, the interested reader is referred to Bott (2010, Exp.2).
3. The actual experiment also included two coercion conditions, namely imperfective sentences with ago-adverbials. Since these are not relevant for the purposes of the present paper, we ignore these conditions and treat trials in these conditions as fillers.
4. Last fixation durations did not differ significantly \(|\tau_{1,2}| < .5\).
5. As in the Russian version, the German experiment included two coercion conditions. Since these are not relevant for our present purposes, we will treat them as fillers. There were six lists with six data points per list and experimental condition.
6. The analysis of last fixation durations revealed no significant differences.
7. Proportions of regressions in the adverbial region did not differ between mismatch and control (mismatch: \(M = 16.2\%\), control: \(M = 9.7\%\); GLME: \(|z| = 1.08\).
8. An anonymous reviewer pointed out that the claim that no aspectual processing occurred before having processed the complete predication may be too strong. We fully agree. Our data only provide evidence that in German lexical aspect/Aktionsart is assigned to the complete predication. Crucially, they do not imply that absolutely no event interpretation occurred before that point. We know from the existing literature, for instance, that nominal arguments are immediately related to the event argument via incremental thematic role assignment (e.g., Knoeferle et al., 2005). Our claim thus only concerns lexical aspect/Aktionsart. Furthermore, on the basis of the present results we do not know whether our results also generalise to non-aspect languages other than German. However, we have to leave this a question for future research.
9. The experiment included a third condition which is not relevant for our present purposes and will thus be ignored.

References