Making semantic commitments can be delayed: Evidence from aspectual processing

Jitka Bartošová, Cassandra Chapman, Ivona Kučerová and Elisabet Service
McMaster University

Semantic re-interpretation has been shown to cause a processing cost. The present study provides evidence that the semantic interpretation of events interacts with real-time processing. The study focused on telicity, an aspectual property that does not solely depend on lexical items but instead on the semantic composition of constituents. Results from a working memory task showed that committing to a semantic interpretation incurs a processing cost and that some adverbials force the parser to commit to a particular aspectual interpretation. Specifically, *in-X-time* forces the parser to commit to a telic (completed) interpretation before the VP has been processed. In contrast, *for-X-time*, which is compatible with an atelic (completed or incomplete) interpretation, does not force the parser to make an early commitment. Results support the partial interpretation hypothesis following which the parser can delay making semantic commitments until it is necessary to do so, i.e., in atelic but not telic sentences.

1 Introduction

1.1 Goal of the study

The current study aims to provide a better understanding of the semantic processing of events. Until now, little work has investigated how the semantic interpretation of an event interacts with its processing in real-time. One exception is work by Frazier & Rayner (1990) who propose two competing hypotheses about semantic processing: the *immediate partial interpretation hypothesis* and the *immediate complete interpretation hypothesis*. These hypotheses make distinct predictions about semantic processing, especially with respect to the timing of semantic interpretation. Frazier & Rayner assume that processing difficulty arises from making a semantic commitment to a particular interpretation of the structure that is being processed. According to the *immediate complete interpretation hypothesis*, the parser makes a semantic commitment

immediately after each phrase is encountered. In contrast, following the *immediate partial interpretation hypothesis*, the parser can delay making a semantic commitment until later in the sentence unless doing so would lead to conflicting interpretations of the structure or to the structure not being interpreted at all. Therefore, the two hypotheses make different predictions with respect to *when* semantic commitments can be made.

To investigate which of the two hypotheses provides better empirical coverage, the present study explored *telicity*, an aspectual property of linguistic events. Telicity is argued to be a property of larger constituents within a sentence (Pustejovsky, 1991). More precisely, we investigated how and when temporal adverbials with distinct aspectual properties, such as *in-X-time* and *for-X-time*, affect the processing of English events.

1.2 Background on telic and atelic events

Linguistic events can be characterized based on whether or not the event has been completed. This property allows us to distinguish two types of events: telic and atelic. While telic events must be interpreted as completed, whether or not the event has been completed is irrelevant for the interpretation of atelic events. For example in the telic sentence in (1a), Peter must have reached the top of the mountain, but in the atelic sentence in (1b), Peter could have eaten some or all of the cookies.[1]

(1) a. *Telic*: Peter reached the top of the mountain.

b. *Atelic*: Peter ate cookies.

It is generally accepted that telicity is not only a property of the lexical verb itself but of its larger structure, i.e., the verb phrase (VP)[2] (Pustejovsky, 1991). Therefore, both the lexical semantics of the verb and the lexical semantics of other material in the structure (for instance, the noun phrase (NP) following the verb) matter for the semantic computation of telicity. As shown in (2), the

[1] The distinction between telic and atelic events has been discussed since Garey (1957) who used the term 'telic', derived from Greek *télos* (meaning 'goal'), for the first time while analyzing the aspectual properties of verbs and nominal arguments in French.

[2] Note that by "VP" we mean the minimal projection below vP, i.e., the verb and its internal argument, if any. We use the term "predicate" interchangeably with the term "VP".

verb *eat* is compatible with both telic and atelic interpretations, depending on the internal NP argument that the verb selects. For example, if the verb selects for a definite NP argument, the VP is interpreted as completed or telic, as in (2a), but if the verb selects for an indefinite NP argument, the VP is interpreted as atelic, as in (2b).

(2) a. *Telic*: Peter ate the cookies.

b. *Atelic*: Peter ate cookies.

However, it is important to note that the lexical semantics of individual verbs also plays a role and the division based on the type of NP argument that the verb selects, as shown in (2), is not always straightforward. The verb in (2) is neutral (or unspecified) with respect to telicity such that whether it is interpreted as telic or atelic is fully determined by its NP argument. There are also verbs whose lexical semantics is more specified. In such cases, the semantic properties of the verb's arguments do not affect its interpretation in the same way as we saw in (2). As shown in (3), the interpretation of the *lose* event is telic, irrespective of the type of NP argument that it selects (definite or indefinite NP). The opposite is true for the verb *love*, which tends to be interpreted as atelic, even if its internal argument is a definite NP, as shown in (4a).

(3) a. *Telic*: Peter lost the cookies.

b. *Telic*: Peter lost cookies.

(4) a. *Atelic*: Peter loved the cookies.

b. *Atelic*: Peter loved cookies.

Moreover, there are predicates that are ambiguous between telic and atelic interpretations, irrespective of the verb's NP argument (if any), as shown in (5). Whether the NP is definite (5a), indefinite (5b), or there is no overt NP in object position (5c) does not affect how the predicate will be interpreted, i.e., both readings are possible in all three cases. For example, (5a) can continue as: ... *and everything is now neatly folded in the wardrobe*, where the event would be interpreted as completed but it can also have the continuation: ...*that's why she couldn't go to the*

movies last night and she can't hang out with us today either as she is not done ironing yet, where the event is interpreted as incomplete.

(5) a. *Telic/Atelic*: Susana ironed the clothes.

b. *Telic/Atelic*: Peter ran a marathon.

c. *Telic/Atelic*: The beer fermented in the barrel.

Giorgi & Pianesi (2001) have shown that the adverbials *in-X-time* and *for-X-time* can be used as a test to distinguish between telic and atelic events.[3] Telic events can only be modified by *in-X-time* adverbials, as shown in (6a). In contrast, atelic events can only be modified by *for-X-time* adverbials, as shown in (6b). This test therefore provides a way to distinguish between telic and atelic interpretations of events independently of the lexical semantics of the verbs.

(6) a. *Telic*: Peter reached the top of the mountain in two hours / *for two hours.

b. *Atelic*: Peter ate cookies *in two hours / for two hours.

Importantly, both *in-X-time* and *for-X-time* adverbials are compatible with ambiguous predicates, as shown in (7). Notice that the event in (7a) is interpreted as completed since it is modified by *in-X-time*. In contrast, the event in (7b), which is modified by *for-X-time*, can either be interpreted as completed or incomplete, i.e., while (7a) cannot have a continuation *but she did not finish ironing the whole pile*, (7b) can. We therefore assume that if a predicate is unspecified for telicity, as in *iron* in (7), the *in-X-time/for-X-time* adverbials bring in the relevant information about telicity, and consequently restrict the interpretation of the event to be telic or atelic.

(7) a. The wife ironed the clothes **in thirty minutes**.

b. The wife ironed the clothes **for thirty minutes**.

To sum up, telicity is not a property of lexical items on their own but of the larger structure (at least VP). Certain lexical items bring in specific information about telicity. However, most of the

[3] The *in-X-time/for-X-time* adverbial test had been previously used to distinguish Vendler (1957)'s accomplishments and achievements from states and activities (for example in Dowty, 1986). However, Giorgi & Pianesi (2001) showed that the test distinguishes telic from atelic events.

time it is not the case that a single lexical item can unambiguously determine whether an event is going to be interpreted as telic or atelic. As noted above, not all verbs are lexically specified for telicity. In certain cases, if the verb selects for a definite NP argument, the event is interpreted as telic. However, this is not true for all VPs and certain VPs remain ambiguous between a telic and atelic interpretation, even if they combine with a definite or indefinite NP argument. We assume that in such cases, the adverbials *in-X-time* and *for-X-time* can restrict the interpretation of event to either telic or atelic. The fact that telicity is not a property of lexical items per se but depends on larger constituents provides us with a tool to investigate semantic processing at the sentential level. In addition, the investigation of telicity in the domain of real-time processing provides us with a way to distinguish between the *complete*, and *partial interpretation hypotheses* (Frazier & Rayner, 1990), which make distinct predictions about when a commitment to a semantic interpretation (in our case, telic or atelic) is made in real-time.

1.3 Processing of telicity

The current study builds on the general assumption that if we know which constituents bring in information about telicity, we can make predictions about how the processing of telic and atelic events unfolds in real-time. We follow Frazier & Rayner (1990) in assuming that making a semantic commitment to a particular interpretation is costly. Therefore, if we know that a constituent forces the parser to unambiguously interpret an event as either telic or atelic, we expect that committing to that semantic interpretation will incur a processing cost. Assuming that the difference between telic and atelic events belongs (at least partially) to semantics (Bach, 1986; Link, 1983; Verkuyl, 1993; Krifka, 1998; Borer, 2005; Ramchand, 2008, among others) and that the meanings of constituents are derived via the systematic semantic combination of lexical items (i.e., compositionally, see e.g., Heim & Kratzer, 1998), the present study investigated if the processing of telic and atelic events can improve our understanding of semantic processing in real-time. In particular, this study investigated the timing of commitments to particular semantic interpretations in the domain of real-time processing.

To be able to make predictions about how the semantic processing of telic and atelic events proceeds, we need to determine the primitives of telic and atelic interpretations. The relevant questions that we need to ask are: i) What does the telic or atelic interpretation correspond to? and ii) Which constituents bring in the telic or atelic interpretation? Only once we have answered these questions can we make predictions about the processing profiles of the two types of events.[4] Our main question concerns the point in real-time at which the parser commits to a semantic interpretation. Does the parser interpret the event as telic or atelic as soon as it encounters a constituent that brings in information about telicity? Can the parser delay making a semantic commitment to an interpretation under some circumstances? Before we lay out our experimental design aiming to address these questions, we will briefly summarize previous experimental work on telicity, from which we borrow a few notions relevant for the current experiment.

1.4 Previous experimental work on telicity

There have been two lines of experimental research in the domain of telicity. In both cases, researchers have focused on how the verb phrase (VP) is processed. The VP has been argued to be the relevant aspectual domain such that the locus of aspectual information is exclusively on the VP. This predicts that if there are any processing differences between telic and atelic events, these differences should be revealed once the VP has been processed.

One line of experimental research investigated how participants process sentences in which a VP with a particular aspectual profile (telic or atelic) is presented with an adverbial phrase (AdvP) that has a mismatching aspectual profile. This mismatch is predicted to lead to a processing cost resulting from semantic re-interpretation, called *aspectual coercion* (Pustejovsky, 1991).[5] Brennan & Pylkkänen (2008) examined aspectual coercion in sentences like (8a) and (8b). They

[4] We work with the assumption that telic and atelic events do not differ in terms of semantic complexity such that one of the interpretations would be more complex than the other.

[5] Note that the fact that a sentence containing an AdvP and a VP mismatching in telicity is not ungrammatical or implausible directly supports the view that telicity is not a property of lexical items per se, and not even just a VP in and of itself. If telicity were a property of lexical items or some fixed constituents, we would not expect lexical items or phrases specified as telic to co-occur with phrases specified as atelic within the same sentence. Information about telicity specified on lexical items/phrases therefore simply reflects a tendency which might be overwritten, as in the cases of aspectual coercion.

predicted a higher processing cost in sentences like (8a), which contains an AdvP and a VP that mismatch in telicity, i.e., an atelic, or *durative*, adverbial combined with a telic, or in their terms *punctual*, predicate, compared to (8b), in which the AdvP and VP have the same aspectual profile, i.e., when the adverbial is telic/punctual and the predicate is also telic/punctual. The authors indeed found that the VP was read more slowly in aspectual coercion conditions, (8a), compared to control conditions, (8b). Brennan & Pylkkänen (2008) argue that these results reflect a higher processing cost associated with mismatching aspectual profiles, compared to matching profiles. Aspectual coercion costs have since been reported also in an eye-tracking study conducted by Townsend (2013).

(8) a. *Coercion*: Throughout the day the student sneezed in the back of the classroom.
b. *Control*: After twenty minutes the student sneezed in the back of the classroom.

A second line of experimental work investigated the difference between telic and atelic events without using aspectual coercion. Stockall & Husband (2014) examined whether the aspectual properties of lexical verbs affect the processing of VPs. In a self-paced reading experiment, they compared the processing profiles of sentences containing verbs lexically specified as telic, such as *lose*, with sentences containing verbs unspecified for telicity, such as *read*. The relevant manipulation was the type of VP-internal argument (definite plural versus bare plural) that appeared with the verb. The type of internal argument either led to an overall telic interpretation, as in (9a)-(9c), or to an atelic interpretation, as in (9d). As was previously mentioned, verbs unspecified for telicity followed by definite NPs lead to a telic interpretation. However, the same verbs are interpreted as atelic if they are followed by indefinite NPs (bare plurals in this case). In contrast, verbs that are lexically specified as telic are interpreted as telic whether or not their argument is a definite or an indefinite NP.

(9) a. *Telic, definite plural*: The expert physicist lost the files on the formation of black holes.
b. *Telic, bare plural*: The expert physicist lost files on the formation of black holes.

c. *Unspecified, definite plural*: The expert physicist read the files on the formation of black holes.

d. *Unspecified, bare plural*: The expert physicist read files on the formation of black holes.

Stockall & Husband (2014) found longer reading times on the word following the VP, i.e., *on*, in the "atelic" condition, i.e., (9d), compared to this same word in the three other conditions.[6] Since this condition leads to an atelic interpretation but the others lead to a telic interpretation, the authors concluded that aspectual processing is affected by the aspectual properties of lexical verbs and that atelic interpretations are overall more difficult to process than telic interpretations.

In a second self-paced reading experiment, Stockall & Husband (2014) compared unspecified verbs, such as *inspect*, and verbs lexically specified as atelic, such as *roam*, as in (10a) and (10b). Regardless of what the overall interpretation of the sentence was (telic or atelic), the authors found no reading time differences between the sentences containing verbs lexically specified as atelic and unspecified verbs. Stockall & Husband (2014)'s results therefore suggest that English verbs can be divided into two groups: i) verbs inherently lexically specified as telic and ii) unspecified verbs. Unspecified verbs are more difficult to process. However, as the second conclusion is based on a null result, it must be treated with some caution.

(10) a. *Unspecified*: The local horticulturist inspected (the) gardens in the neighborhood.

b. *Atelic*: The local horticulturist roamed (the) gardens in the neighborhood.

To summarize, there are two main findings from previous experimental work on telicity that are relevant for the current study: i) adverbials and predicates mismatching in telicity lead to a processing cost called aspectual coercion, and ii) no difference in the processing profiles could be detected between sentences containing verbs lexically specified as atelic and unspecified verbs. However, sentences containing both of these types of verbs seemed to be more difficult to process

[6] By "atelic" codition we mean the condition containing an unspecified verb with a bare plural NP argument that leads to an overall atelic interpretation. We put "atelic" in quotations because it is our label we use to describe Stockall & Husband (2014)'s stimuli.

than sentences containing verbs lexically specified as telic. Both of these findings clearly suggest that aspectual information affects processing and that the VP is relevant for aspectual processing.

2 Current study

The current study builds on Stockall & Husband (2014)'s conclusion that telic and atelic events have distinct aspectual processing profiles. We also follow Brennan & Pylkkänen (2008) in assuming that AdvPs can affect the aspectual interpretation of verbs, and therefore, the semantic processing of events. However, we depart from previous experimental work in arguing that the VP is not the only aspectual domain responsible for the telic or atelic interpretation of an event. Instead, we argue that in some cases, AdvPs can restrict the semantic interpretation of events. We also depart from Stockall & Husband (2014) in that we question the conclusion that one type of event is overall more difficult to process than the other. Similarly to Frazier & Rayner (1990), we assume that making a semantic commitment to a particular interpretation is costly. We expect processing differences to arise in the syntactic position in which a semantic commitment can or must be made. In other words, we predict a local processing cost when the parser is able to make a semantic commitment. We used a working memory task to detect those processing costs.

While previous research has shown that the adverbials *in-X-time* and *for-X-time* affect the interpretation of events, it is not clear if the AdvPs themselves have an effect on real-time processing. In the current study, we asked: Do the *in-X-time* and *for-X-time* adverbials play a role in the processing of aspectual information? As mentioned above, previous experimental work on telicity has assumed that it is the VP that is the locus of aspectual information. However, the results from Brennan & Pylkkänen (2008) suggest that the AdvP also affects aspectual processing. These researchers focused on constructions in which the aspectual properties of the VP and AdvP mismatched in telicity. A related question is whether the AdvP on its own might affect the semantic processing of events. Such results would demonstrate that the VP is not the sole bearer of aspectual information and that other phrases, such as the AdvP, contribute to the aspectual interpretation of events. In the current study, we hypothesise that the adverbials

in-X-time and *for-X-time* play a role in the processing of aspectual information, even when coercion is not involved.

To investigate if *in-X-time/for-X-time* adverbials affect the semantic processing of events, our experimental design relied on two main components: (i) the existence of predicates that are ambiguous between the two interpretations (telic vs. atelic) when they appear on their own, as previously shown in (5), and (ii) the assumption that *in-X-time/for-X-time* adverbials restrict the aspectual interpretation of sentences containing ambiguous predicates to either telic or atelic.[7] Our stimuli therefore consisted of sentences containing ambiguous predicates combined with *in-X-time* and *for-X-time* adverbials, as previously shown in (7) and repeated here as (11).

(11) a. The wife ironed the clothes **in thirty minutes**.

b. The wife ironed the clothes **for thirty minutes**.

The current experimental design differs from that of Brennan & Pylkkänen (2008) in that we did not manipulate whether the VP and the AdvP mismatched in telicity. As the ambiguous predicates are unspecified with respect to telicity, they are fully compatible with both types of AdvPs. We therefore did not expect any instantiation of aspectual coercion, i.e., the telicity of one phrase being overwritten by the mismatching telicity of another phrase. In other words, we predicted that any potential processing differences between (11a) and (11b) would be solely due to the aspectual properties of the AdvPs.

To further isolate the potential effect of temporal adverbials on semantic processing, we also manipulated the syntactic position of the adverbials in the sentences. Our stimuli consisted of sentences in which the AdvP appeared either in canonical position, i.e., after the VP, as in (12a), or before the VP, as in (12b). The sentence-initial position of the AdvP (called *adverbial first*) allowed us to investigate whether the AdvP has an effect on the interpretation of the event before the VP has been encountered. Any processing effects found immediately following the AdvP would be solely due to that region itself and would not reflect effects that might interact with the

[7] We thank Bridget Copley for suggesting that we use a minimal pair design with scalar verbs to examine the processing of telic versus atelic events.

VP region.

(12) a. The wife ironed the clothes **in/for thirty minutes**.

b. **In/for thirty minutes** the wife ironed the clothes.

Even though it may seem as though both adverbials equally contribute to the aspectual interpretation of sentences with ambiguous predicates, we argue that only *in-X-time* forces the event to be interpreted as telic. We assume that the relevant property that allows us to distinguish telic from atelic events is whether or not the event was completed. As shown above, in sentences containing *in-X-time*, the event must be interpreted as completed. In contrast, if the sentence contains *for-X-time*, whether or not the event was completed is irrelevant. In other words, atelic events can either be interpreted as completed or as incomplete. In sentences containing ambiguous predicates, it is only the adverbial that brings in the relevant aspectual information. Building on the idea that event completion is the relevant property contributing to aspectual interpretation, we make the following assumptions about the adverbials: (i) only the *in-X-time* adverbial unambiguously brings in the telic interpretation and forces the parser to interpret the event as completed as soon as the AdvP is encountered, and (ii) the *for-X-time* adverbial does not bring in any particular aspectual interpretation to which the parser could commit as soon as the AdvP is encountered.

Using two distinct syntactic positions for the adverbials in combination with Frazier & Rayner (1990)'s two hypotheses about semantic processing, we are able to make precise predictions about when the parser can or must interpret an event as telic or atelic. Following Frazier & Rayner, we assume that committing to a particular semantic interpretation leads to an immediate processing cost. Crucially, the two semantic processing hypotheses make distinct predictions about when this processing cost will be incurred. According to the *immediate complete interpretation hypothesis*, the parser must interpret each phrase as soon as it is encountered. Following this hypothesis, we do not predict any processing differences to arise based on the type or syntactic position of the adverbial. If each phrase needs to be interpreted as soon as it is

encountered, the parser should commit to a particular interpretation as soon as it processes the phrase. In the current study, upon encountering one of the temporal adverbials, the parser should commit to one or the other aspectual interpretation. In contrast, following the *immediate partial interpretation hypothesis*, the parser may delay making semantic commitments to a particular interpretation in some circumstances. If this hypothesis is on the right track, we predict that both the type of adverbial and its syntactic position will affect processing. In particular, upon encountering *in-X-time*, the parser should immediately commit to a telic interpretation of the event. Since we manipulated the syntactic position of the adverbial (adverbial first or canonical position), we predict that such an effect should be stronger when the parser has been given a very early indication about the aspectual interpretation of the event, i.e., when the *in-X-time* adverbial appears at the beginning of the sentence. However, since *for-X-time* does not force the parser to commit to a particular interpretation (since the event can be interpreted as completed or not), we predict that the parser can delay making a semantic commitment until later in the sentence, irrespective of when it encounters the adverbial. As a result, we predict an interaction between the type (telic vs. atelic) and syntactic position of the adverbial in our sentences, if the partial interpretation hypothesis is on the right track. Before detailing these exact predictions, we will explain the methodology used in the current study, which enabled us to investigate potential local effects of the adverbial in the larger sentential context.

2.1 Working memory task

We used the complex span methodology, which was originally designed to measure individual working memory (WM) capacity (Daneman & Carpenter, 1980) but has been recently found to be sensitive to local effects of syntactic complexity (Chapman, Deschamps, et al., 2016; Chapman, Kuperman, & Service, 2016; Chapman & Service, 2016). In this task, participants read a series of sentences for comprehension and are asked to remember the last word of each sentence for later recall. The task is a dual task because it requires participants to i) read and comprehend sentences, and ii) remember memory words for later recall in the order they were presented to

them (serial). Since its original formulation, this task has been adapted by psycholinguists to investigate whether sentence processing interacts with WM capacity (see e.g., Just & Carpenter, 1992; Caplan & Waters, 1999; Fedorenko et al., 2006; Gordon et al., 2002, among others). The adapted version of the task, also called *reading, listening* or *sentence span*, asks participants to process sentences and remember unrelated memory words presented at the end of the sentence while simultaneously answering comprehension questions or making grammaticality judgements. Researchers have manipulated the syntactic complexity of the sentence in order to investigate how WM might be affected by more complex sentences. The critical dependent variable is the number of memory words that can be recalled in the correct order. In this task, the assumption is that lower memory word recall reflects a greater processing load. Generally, it has been found that recall is lower when the syntactic structure is more complex (i.e., object- versus subject-extracted relative clauses).

Chapman, Deschamps, et al. (2016) have recently proposed a novel variant of the complex span task in which memory words are placed *within* the processing sentences. The researchers argue that this method provides a more sensitive method to investigate whether memory encoding affects sentence processing because participants are asked to encode words as they process the sentences. As the critical manipulation, Chapman, Kuperman, & Service (2016); Chapman & Service (2016) investigated sentences containing subject- or object-extracted relative clauses, see (13a) and (13b) respectively, and found that participants were less likely to recall memory words placed within more complex sentences, e.g., object relative clauses, suggesting that this variant of complex span is sensitive enough to capture effects of syntactic complexity.

(13) a. *Subject*: The wrestler that impressed the cheerleader won the battle.

b. *Object*: The wrestler that the cheerleader impressed won the battle.

The complex span task results are in line with a body of work on subject- and object-extracted relative clauses. In previous work, researchers have consistently found that object-relatives, as in (13a), incur a greater processing cost compared to subject-relatives, as in (13b) (see for instance

Gibson, 1998, 2000; Grodner & Gibson, 2005; Hale, 2001; Levy, 2008; Lewis & Vasishth, 2005; Lewis et al., 2006; Staub, 2010; Traxler et al., 2002; Chapman, Kuperman, & Service, 2016; Chapman & Service, 2016, among others).

We propose that Chapman, Deschamps, et al. (2016)'s methodology allows us to investigate whether we can isolate the locus of aspectual information in the processing of sentences because memory words can be placed in different syntactic locations. Assuming that making a semantic commitment to an interpretation is locally costly, we expect to find poorer memory word recall after the commitment to a semantic interpretation has been made. For example, if the AdvP brings in the relevant aspectual information, we should be able to find differences in recall performance when the memory word is presented after the AdvP. More precisely, if the AdvP alone brings in the aspectual information, irrespective of the verb, such effects should be clearly seen when the AdvP is presented before the VP. Furthermore, if there are differences based on when the parser can make semantic commitments (*complete* versus *partial interpretation* hypotheses), this methodology allows us to investigate such effects. Since the location of the memory word can be manipulated, we can investigate local effects at different regions in the sentences. Note that even though it is possible for the two types of adverbials to attach at different syntactic positions, we do not expect any processing differences based on the syntactic attachment site of the adverbial because the sentences do not give rise to a syntactic attachment ambiguity, i.e., the differences in aspectual interpretation cannot be attributed to the attachment site of the adverbial. The differences in interpretation depend solely on the aspectual properties of the adverbials themselves, i.e., whether they are compatible with telic or atelic interpretations.

2.2 Methods

2.2.1 Participants.

Thirty three participants (2 male) were tested. The data from one participant was removed from the final analyses because they turned out to not be a native speaker of English. The remaining 32 participants were native speakers of English, between 17 and 28 years of age (M =

20.28 years, *SD* = 1.91) and were recruited through the participant pool in the Department of Linguistics and Languages at McMaster University. Participants received a course credit as compensation and provided informed consent prior to beginning the experiment. All participants were naïve as to the purposes of the present study.

2.2.2 Stimuli

Forty-five scalar predicates that are compatible with both *in-X-time* or *for-X-time* were used in the stimulus sentences. A total of ninety sentences, originally constructed by Baraniuk (2014), were modified and used in the present study. In Baraniuk (2014)'s study, stimuli consisted of minimal pairs in which the sentences differed only based on which adverbial was used, as in (14a) for the telic interpretation and (14b) for the atelic interpretation. Memory words were presented in two locations in the sentences: i) before the temporal adverbial phrase, i.e., WORD1, or ii) after the sentence, i.e., WORD2. In Baraniuk (2014)'s experiment, stimulus sentences always consisted of a full NP subject, a verb in the past tense, an object NP (in most cases),[8] and the critical temporal adverbial phrase.

(14) a. *Telic*: The wife ironed the clothes **WORD1** in two hours. **WORD2**

b. *Atelic*: The wife ironed the clothes **WORD1** for thirty minutes. **WORD2**

In the current study, the same minimal pairs were used as in Baraniuk (2014) but a second adverbial phrase appeared at the end of each sentence in order to avoid wrap-up effects, as in (15a)-(15d). In addition to the original conditions, we also included two conditions in which the critical adverbial phrase appeared before the VP, as in (15c)-(15d).[9] Memory words were presented in three locations in order to examine the effect of memory encoding on processing incrementally. Doing so allowed us to investigate local effects and specifically, examine whether

[8] Due to the fact that the number of scalar predicates in English is limited, 11 of the predicates used did not select for a VP object. The memory word was always placed after the VP and thus, either appeared after a verb that did not select for an object or after the verb and its argument.

[9] In the conditions in which the AdvP appeared first, a pronoun subject was used instead of a full NP, compare (15a)-(15b) to (15c)-(15d). This experiment was designed as a control for a different experiment in which this manipulation was crucial, however, this difference is irrelevant for the current experiment.

the aspectual properties of the adverbials affect encoding. Across all conditions, the memory word appeared after the whole VP, shown as WORD1 in (15a)-(15b) and as WORD2 in (15c)-(15d). Memory words could also appear after the AdvP, as WORD2 in (15a)-(15b) and WORD1 in (15c)-(15d). Finally, memory words could appear in a third position after the whole sentence had been presented, in this case, after the second AdvP in all conditions, shown as WORD3.

(15) a. *Telic, AdvP after VP*: The wife ironed the clothes **WORD1** in two hours **WORD2** at the house. **WORD3**

b. *Atelic, AdvP after VP*: The wife ironed the clothes **WORD1** for thirty minutes **WORD2** at the house. **WORD3**

c. *Telic, AdvP before VP*: In two hours **WORD1** she ironed the clothes **WORD2** at the house. **WORD3**

d. *Atelic, AdvP before VP*: For thirty minutes **WORD1** she ironed the clothes **WORD2** at the house. **WORD3**

In the complex span task, participants were presented with a total of 180 sentences: 90 test sentences as described above and 90 fillers from an unrelated experiment. Three lists of stimulus sentences were created such that memory words appeared in all three locations across all sentences; the locations were counterbalanced across lists. After the lists were created, the sentences were randomly divided into sets of five sentences. All sentences in a set always belonged to the same condition, i.e., (i) telic sentences with canonical word order, (15a), (ii) atelic sentences with canonical word order, (15b), (iii) telic sentences with the fronted temporal adverbial phrase, (15c), and (iv) atelic sentences with the fronted temporal adverbial phrase, (15d). The lists were created such that half of the sets were telic and the other half were atelic.

In total, participants saw nine telic and nine atelic sets of 5 sentences (i.e., ninety sentences). Out of these eighteen sets, half were in the canonical word order and the other half had the AdvP at the beginning of the sentence. Crucially, if participants saw the atelic sentence template in the canonical word order, they saw the corresponding telic sentence templates in the adverbial-first

word order. They never saw the same sentence template in both telicity conditions and in the same adverbial condition. Sets of sentences with the adverbial first word order were randomized such that in the minimally different sets containing the same sentence templates but with the canonical word order, the sentences were not presented in the same order as their matching counterparts.

Trial order was pseudo-randomized such that test trials were always followed by filler trials. No more than two sets in the same telicity condition, i.e., telic or atelic, appeared in a row (with a filler trial in between). Also, no more than three trials with the same word order condition, i.e., canonical or adverbial first, appeared in a row. These factors were all controlled to minimize the ability of the participants to guess the purpose of the study. In order to control for ordering effects, we also created a second list of stimuli in which trials were presented in the reverse order as compared to the first list (see Chapman, Deschamps, et al., 2016 for discussion about ordering effects in complex span). In total, six lists of stimulus sentences were created: three lists in which the locations of the memory words were counterbalanced across lists and two different stimulus orders.

Forty-five memory words were selected from the MRC Psycholinguistic Database (Coltheart, 1981). The memory words were four-letter nouns with familiarity (M = 557.51, SD = 29.53) and imageability (M = 576.93, SD = 27.98) ratings between 500 and 700. The memory words in the current experiment were divided such that each sentence with the same verb was assigned one memory word, i.e., all four sentences in (15a)-(15d) contained the same memory word across all the lists. This means that for each list, the same memory word appeared twice, i.e., once together with a canonical word order sentence and once with an adverbial-first word order sentence. Note that one of these sentences was telic and the other was atelic.[10]

To guarantee processing of the sentences, eighteen “control statements” were constructed for the test sentences, one following each set of 5 sentences. Participants were asked to determine if the statement, which was always related to one of the sentences from the set, was true or false (50% “true”). For instance, if the trial contained the sentence in (15a), the statement was: “The

[10] Note that for the ninety filler sentences, there were 90 memory words so participants saw a total of 135 memory words in the experiment.

wife ironed the clothes at the hotel." (answer: false). Participants answered the control statements correctly 62% of the time. We attribute this low accuracy to the difficulty of the task. Since participants were required to read sentences and encode memory words, this made it difficult to remember all the details from the sentences, especially if the sentence that was related to the control statement appeared early in the set of 5.

2.2.3 Procedure

Participants were instructed that they would be presented with sentences phrase-by-phrase at a rapid pace on a computer screen and that a word in red bold capitals would appear somewhere in each sentence. They were asked to memorize those red words in the order that they saw them for recall after each set of 5 sentences. At recall, they were asked to say "BLANK" if they could not remember a word in a particular positions. Participants were also instructed to read each sentence for comprehension as they might be asked about the sentence later in the experiment. A practice session at the beginning of the experiment consisted of three sets of 5 sentences, which were not analyzed.

The presentation for all sentences was experimenter-paced beginning with a fixation cross in the centre of the screen for 500 ms. For example, (16a) and (16b) indicate how the sentences were divided into regions in sentences containing *in-X-time* adverbial with the canonical or the adverbial first word order. Every region between two slashes was presented separately on a computer screen. The presentation time for each region was calculated based on the number of words in each phrase. Content words (personal pronouns included) were presented for 400 ms and function words were presented for 200 ms. The sentence words were presented in black lettering on white background in the middle of the screen. Memory words appeared for 1000 ms in bold red capitals in the centre of the screen in one of the three memory word locations within a sentence.

(16) a. *Canonical order - regions*: / The wife / ironed the clothes / in two hours / at the house. /

b. *Adverbial-first order - regions*: / In two hours / she / ironed the clothes / at the house. /

After each set of five sentences, a recall screen appeared asking participants to recall the red words aloud in the order they had appeared. The experimenter recorded their responses. After the recall task, participants were presented with a control statement related to one of the sentences. They were asked to press either Y for *yes* or N for *no* to verify the statement. After answering the control statement, the next trial appeared in the same fashion. The experiment lasted approximately 30 minutes.

2.3 Predictions

As the two types of adverbials carry different semantic interpretations, they make different predictions for processing. As previously discussed, *in-X-time* restricts the aspectual interpretation of the event such that it must be interpreted as completed. For example, in (15a), the wife must have completed the ironing event. Since this adverbial restricts the interpretation of the event, we predict that it forces the parser to make a semantic commitment to a telic interpretation. We assume that making a semantic commitment incurs a processing cost Frazier & Rayner (1990). We therefore predict that encountering the *in-X-time* adverbial should be costly. In the complex span task, where processing difficulty is measured by memory performance, encountering the *in-X-time* adverbial should lead to lower recall accuracy. In contrast, *for-X-time* does not restrict the aspectual interpretation of the event in the same way: whether or not the event was completed is irrelevant. For example, in (15b), whether the ironing event was completed is irrelevant. As the adverbial does not restrict the aspectual interpretation, we do not predict an immediate processing cost when the parser encounters this adverbial. In the complex span task, we should therefore observe better memory word performance following *for-X-time* compared to *in-X-time*.

Crucially, we also manipulated the syntactic position of the adverbial in our stimulus sentences such that it either appeared at the beginning of the sentence (adverbial first) or after the

VP (canonical). This manipulation enabled us to examine potential local effects of the adverbial before the VP is integrated within the sentence. If the adverbial appears at the beginning of the sentence, the parser receives a very early indication about how the event described by the VP (which is found further to the right) should be interpreted. In contrast, if the adverbial appears after the VP, the parser will not be able to adopt a particular aspectual interpretation until it has processed the whole sentence.

The two semantic processing hypotheses proposed by Frazier & Rayner (1990) make distinct predictions about whether the type and syntactic position of the adverbials should affect their processing cost. According to the *complete interpretation hypothesis*, we do not predict any differences based on the type of adverbial. Under this hypothesis, semantic commitments are made immediately after each phrase is encountered and consequently, we should not observe differences based on whether the adverbial restricts the aspectual interpretation. In other words, the parser should immediately commit to either a telic or atelic interpretation of the event upon encountering the adverbial, irrespective of where the adverbial appears in the syntactic structure. In contrast, according to the *partial interpretation hypothesis*, the parser can delay making a semantic commitment until later in the sentence, unless this delay would lead to (i) no interpretation at all, or (ii) conflicting interpretations. We hypothesise that whether a semantic commitment must be made immediately depends on whether or not the adverbial restricts the aspectual interpretation of the event. In the case of *in-X-time*, the event must be interpreted as completed and consequently, the parser can commit to this interpretation of the event as soon as it encounters the adverbial. However, in the case of *for-X-time*, the aspectual interpretation of the event is not restricted and consequently, the parser can delay committing to an atelic interpretation until later in the sentence. Nevertheless, the parser needs to make a semantic commitment to an aspectual interpretation at some point within the sentence. We expect a processing cost to be incurred when this commitment is made, i.e., towards the end of the sentence.

Thus, the *partial interpretation hypothesis* makes precise predictions based on the two types of adverbials. Following this hypothesis, we predict that *in-X-time* should incur a local processing

cost as soon as the adverbial is read because the adverbial enforces a telic (completed) interpretation. Since we are using the complex span task, we predict that recall should be lower immediately following the *in-X-time* adverbial, as soon as the parser commits to an aspectual interpretation. Since we also manipulated the syntactic position of the adverbial (before or after the VP), we are also able to investigate whether this potential effect differs based on when the adverbial is encountered. When the adverbial appears at the beginning of the sentence (before the VP), the parser has been given a very early indication that the event must be interpreted as completed. Consequently, we expect the effect of making a semantic commitment to a telic interpretation (processing cost, as measured by lower recall accuracy) to be stronger when the adverbial appears early in the sentence, compared to when it follows the VP. In contrast, we predict that *for-X-time* should incur a processing cost later in the sentence, only when a semantic commitment must be made. In this case, word recall should be lower for memory words presented after the whole sentence has been processed, i.e., the point at which the parser must make a decision about how the sentence will be interpreted. We do not predict any differences based on word order for the *for-X-time* adverbial because whether or not this adverbial appears early in the sentence does not change the aspectual interpretation. In other words, irrespective of the syntactic position of the adverbial, making a semantic commitment to a particular interpretation can be delayed until later in the sentence. The main predictions based on the complete and partial interpretation hypotheses are summarized in Table 1.

Table 1

Predictions based on the partial and complete interpretation hypotheses for in- and for-X-time adverbials

	Complete		Partial	
adverbial	*in-X-time*	*for-X-time*	*in-X-time*	*for-X-time*
delay?	no	no	**no**	**yes**

2.4 Results

Word recall accuracy was investigated using the complex span task to examine potential local effects of semantic processing costs. Participants' recall was scored based on strict serial order, i.e., they only received a point if they recalled the words in the order in which they were presented. Previous statistical modelling work has shown that participants and items contribute to random variance in psycholinguistic experiments. To control for this type of variance, we ran generalized linear mixed effects multiple regression models with participants and memory words as random effects (R. H. Baayen et al., 2008; H. R. Baayen, 2008; Pinheiro & Bates, 2000), as implemented in the lme4 package (version 1.1-10, Bates & Sarkar, 2007) for R (version 3.2.3, R Core Development Team, 2015). This type of statistical analysis allows several factors and predictors to be explored at once and takes into consideration any variance between participants and/or items. We initially fitted each model with a maximal random-effects structure (Barr et al., 2013) but then trimmed down each model to include only those random effects that significantly improved the model's performance, as determined by the likelihood ratio test. We also removed from the model all fixed effects that did not significantly contribute to the model's performance. In addition, outliers were removed: any data points that were +/- 2.5 SDs from the residual error of the model were removed. After removing outliers, the models were refitted. We report only the final fitted models after trimming.

As our dependent variable formed a binomial distribution (scored as 1 if the word was recalled correctly, 0 if it was not), we used the logical regression statistical model. Our main questions were whether the aspectual properties of the adverbial (*in-X-time* for telic, *for-X-time* for atelic) affected participants' recall performance and whether the syntactic position in which this adverb appeared (adverbial-first versus canonical word order) affected recall differently. We were also interested in whether the different memory word locations (After AdvP, After VP and After sentence) played a role, such that locations of increased processing load, where we hypothesised semantic commitments to be made, would result in poorer recall. In addition to these theoretically-specific questions, we also investigated whether the serial position of the recall

word within the set of 5 sentences (Number in set) affected recall performance (see e.g., Chapman, Deschamps, et al., 2016). Note that 0 in our model represents sentence number 3. We also investigated whether the block of sentences in which the memory words were presented affected recall (Block). Finally, as many previous studies have shown that the order of stimulus presentation can affect performance, we investigated the normalized order of the trials in the experiment (Trial number).

We fitted a linear mixed effects logistic regression model to the recall data with an interaction between memory word location (After AdvP, After VP, After sentence) and the syntactic position of the adverbial (see Table 2). While sentences with telic adverbials (*for-X-time*) had numerically higher recall scores compared to atelic adverbials, this effect was not statistically significant (β = 0.045, SE = 0.087, z = 0.518, p = 0.604, model not shown[11]). Note that a difference based on adverbial type is neither predicted by the complete nor the partial interpretation hypotheses. Such an effect would be expected only if the telic and atelic interpretations inherently differed in terms of semantic complexity (see Stockall & Husband, 2014). The complete interpretation hypothesis does not predict a difference based on the type of adverbial because, under this hypothesis, semantic commitments are made immediately after each phrase, regardless of their aspectual interpretation. Consequently, no differences between the two interpretations are expected. In contrast, the partial interpretation hypothesis does predict a processing difference based on the type of adverbial but the predicted effect depends on the syntactic position in which the parser can commit to a semantic commitment. Thus, an overall main effect of adverbial type is not expected.

The model did reveal a main effect of memory word location such that participants were less likely to recall words when they appeared after the AdvP compared to After the VP (β = 0.272, SE = 0.106, z = 2.557, p = 0.01, model not shown) and After the sentence (marginal: β = 0.175, SE = 0.106, z = 1.656, p = 0.098, model not shown). These results suggest that it was overall more difficult to recall words when they were presented After the AdvP compared to the two other memory word locations. We will examine the effect of memory word location in greater

[11] For reasons of space, we report in the tables the models that included interaction terms only. We report within the text the relevant statistical results for main effects.

detail in models that separate the results by adverbial type (reported below).

There was also a significant main effect of the word's serial position in the block of 5 sentences such that participants were more likely to recall words at the beginning of the set of 5 sentences compared to the end of the set (β = -0.379, SE = 0.069, z = -5.476, $p < 0.001$, model not shown). This reflects a standard serial order effect, as shown in Figure 1. In addition, there was a main effect of trial number in the experiment, such that participants were more likely to recall words correctly towards the end compared to the beginning of the experiment (β = 0.301, SE = 0.044, z = 6.843, $p < 0.001$, model not shown), probably reflecting learning over the course of the experiment.

The most interesting effects are interactions. There were significant or marginal interactions between the location of the memory word in the sentence and syntactic position of the adverbial (canonical vs. adverbial first), as shown in Figure 2 and as reported in Table 2. The interactions indicate that while participants were overall less likely to recall words when they appeared after the AdvP, this effect was stronger when the adverbial appeared first (compared to After sentence: $p < 0.01$, as shown in Table 2, and compared to After VP: β = 0.363, SE = 0.213, z = 1.703, p = 0.09, model not shown).

However, these results still do not allow us to distinguish between the partial and complete interpretation hypotheses because this statistical model collapses the data for the two types of adverbials. The crucial prediction from the partial interpretation hypothesis is that we should find differences between the two adverbial types based both on their position in the sentence and the location of the memory word, i.e., depending on when the parser can commit to an aspectual interpretation. To investigate these effects further, we looked at recall accuracy based on the aspectual properties of the adverbials in separate models. We predicted that if the partial interpretation hypothesis is on the right track, then the parser should be forced to make a semantic commitment to a telic interpretation as soon as *in-X-time* is processed. Therefore, we expected processing difficulty, as measured by lower recall performance, to arise immediately after *in-X-time* is processed. We predicted that this effect should be stronger when the adverbial is

encountered at the beginning of the sentence since the parser has been given an early indication that the sentence must be interpreted as telic. In contrast, since *for-X-time* does not force the parser to commit a particular aspectual interpretation, we expect that it will delay doing so until the end of the sentence. In this case, we should observe processing difficulty at the end of the sentence, irrespective of where in the sentence the adverbial is located. The parser only needs to commit to a particular interpretation of the sentence once it has processed the whole structure. Alternatively, according to the complete interpretation hypothesis, the parser should commit to a particular aspectual interpretation as soon as it reads the adverbial. Thus, there should be no delay in making semantic commitments based on the type of adverbial. Consequently, *in-X-time* and *for-X-time* should be processed similarly because both adverbials require that the parser immediately commit to an aspectual interpretation (telic or atelic, respectively).

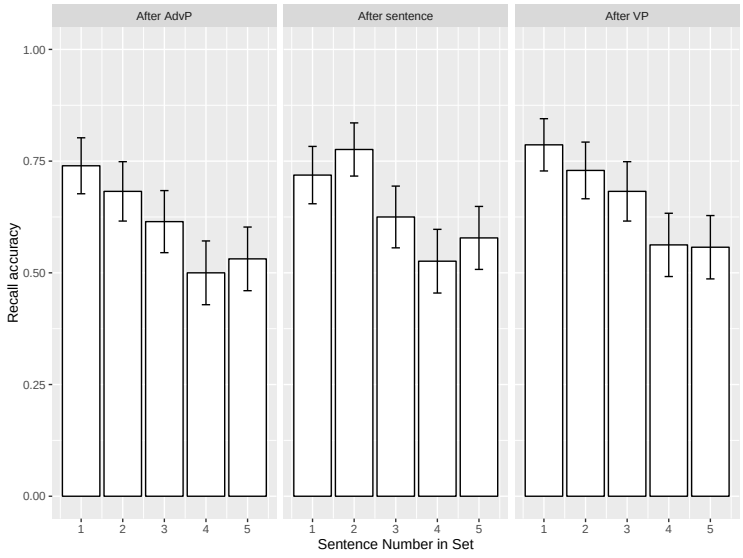

Fig. 1. Recall accuracy by sentence number in set and memory word location. Error bars represent 95 % confidence intervals.

2.4.1 In-X-time

We fitted a linear mixed effects logistic regression model to the recall data including an interaction between the location of the memory word and the syntactic position of the adverbial for sentences containing *in-X-time* (e.g., *The wife ironed the clothes in two hours at the house./In*

Table 2

Final mixed-effects model for serial recall accuracy (N = 2880 before trimming, 2874 after trimming), reported as the regression coefficient estimates, standard errors and z-values and p-values. Reference levels = After sentence, Atelic, Canonical word order

Variable	Coefficient estimate	Std. error	z-value	p-value
Intercept	0.750	0.237	3.158	0.002
Trial Number	0.304	0.044	6.878	0.000
After AdvP	0.123	0.15	0.823	0.411
After VP	0.213	0.151	1.415	0.157
Adverbial first	0.171	0.182	0.937	0.349
Telic	0.045	0.087	0.514	0.608
Number in set	-0.537	0.098	-5.467	0.000
After AdvP * Adverbial first	-0.596	0.212	-2.811	0.005
After VP * Adverbial first	-0.233	0.214	-1.090	0.276

Random effects: by-word intercept (SD = 0.256), by-participant intercept (SD = 1.069), by-block intercept (SD = 0.218), by-participant random slope for sentence number in set (SD = 0.448), and the correlation between by-participant intercept and slope (r = -0.94).

two hours she ironed the clothes at the house.), which restricts the interpretation to telic, as shown in Table 3. There was a significant main effect of the location of the memory word such that recall was less likely After the AdvP compared to After the sentence (β = 0.57, SE = 0.154, z = 3.693, $p < 0.001$, model not shown) or After the VP (β = 0.301, SE = 0.155, z = 1.938, p = 0.053, model not shown). Recall was also marginally more likely after the sentence compared to After the VP (β = -0.269, SE = 0.157, z = -1.716, p = 0.09, model not shown). However, we did not observe a significant main effect of the syntactic position of the adverbial in the sentence (β = -0.238, SE = 0.195, z = -1.221, p = 0.222, model not shown). Crucially, there were also significant interactions between the location of the memory word and the syntactic position of the adverbial, as shown in

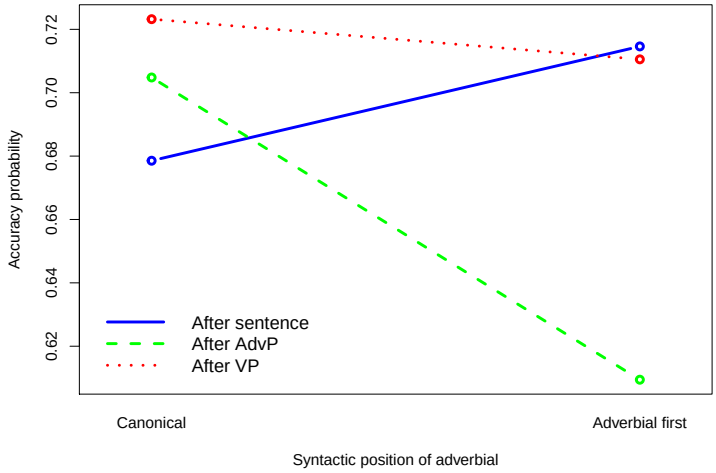

Fig. 2. Interaction plot showing an effect of memory word location by the syntactic position of the adverbial.

Figure 4. The significant interactions demonstrate that while recall was overall more likely After the sentence compared to After the AdvP, this effect was stronger when the adverbial appeared first compared to canonical position ($p < 0.01$, Table 3). The interaction between memory word location and syntactic position of the adverbial did not reach significance when comparing the after AdvP and after VP memory word locations, even though recall was overall more likely After the VP compared to After the AdvP. This suggests that the effect of syntactic position of the adverbial on memory word location only shows up at a much later region in the sentence (After AdvP vs. After sentence). We did however observe a significant interaction between the After VP and After sentence word locations (β = -0.62, SE = 0.312, z = -1.987, p = 0.047, model not shown), which again suggests that recall was easier at the end of the sentence when the adverbial appeared first. Since the effect appears at the end of the sentence in comparison to the other two word locations, it seems that processing becomes much easier once the parser can commit to a semantic interpretation signalled at the beginning of the sentence, i.e., when the adverbial appears first. When the adverbial appears in canonical position (after the VP), the parser has not been given an early indication that the sentence must be interpreted as telic and therefore it cannot commit to an aspectual interpretation until it reaches the adverbial later in the sentence.

Table 3

Final mixed-effects model for serial recall accuracy in telic sentences (N = 1440 before trimming, 1439 after trimming), reported as the regression coefficient estimates, standard errors and z-values and p-values. Reference levels = After AdvP, Canonical word order

Variable	Coefficient estimate	Std. error	z-value	p-value
Intercept	0.805	0.256	3.148	0.002
Trial Number	0.304	0.098	3.109	0.002
After sentence	0.089	0.214	0.416	0.677
After VP	0.143	0.216	0.661	0.509
Adverbial first	-0.619	0.250	-2.476	0.013
Number in set	-0.409	0.069	-5.897	0.000
After sentence* Adverbial first	0.884	0.308	2.870	0.004
After VP * Adverbial first	0.265	0.300	0.881	0.378

Random effects: by-word intercept (SD = 0.227), by-participant intercept (SD = 1.008), and by-block intercept (SD = 0.282).

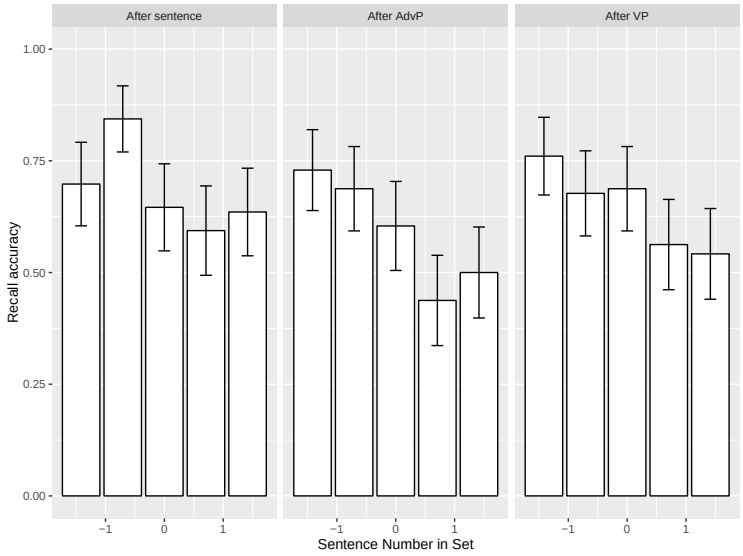

Fig. 3. Recall accuracy by sentence number in set and memory word location for sentences containing the telic adverbial (*in-X-time*). Error bars represent 95 % confidence intervals.

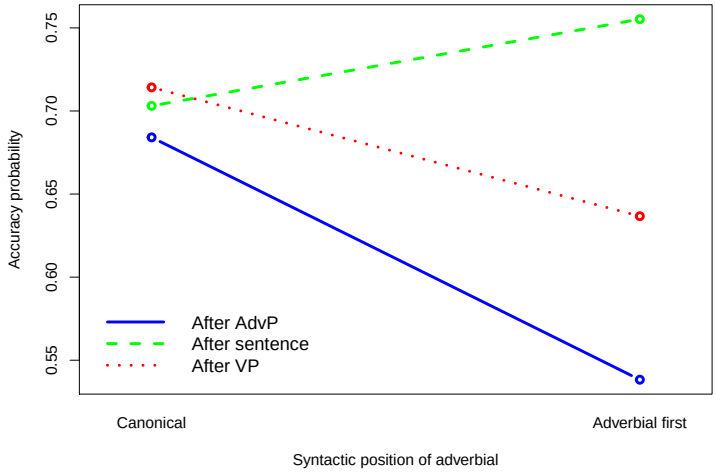

Fig. 4. Interaction plot showing an effect of the syntactic position of the adverbial on recall accuracy by memory word location for telic sentences.

2.4.2 For-X-time

We also fitted a linear mixed effects logistic regression model to the recall data including an interaction between the location of the memory word and the syntactic position of the adverbial for sentences containing *for-X-time* (e.g., *The wife ironed the clothes for thirty minutes./For thirty minutes she ironed the clothes at the house.*), which does not force a particular aspectual interpretation of the sentence, as shown in Table 4. There was a significant main effect of the location of the memory word such that recall was less likely After the sentence compared to After the VP ($p < 0.01$), as shown in Figure 5. Recall was also numerically less likely After the sentence compared to After the AdvP but this effect did not reach significance at the 5% threshold. We also did not observe a significant main effect of the syntactic position of the adverbial nor did we observe significant interactions between memory word location and the syntactic position of the adverbial, suggesting that the main effects in the atelic cases did not depend on the position of the adverbial within the sentence. We will discuss the theoretical implications of these results in the General Discussion section.

Table 4

Final mixed-effects model for serial recall accuracy in atelic sentences (N = 1440 before trimming, 1439 after trimming), reported as the regression coefficient estimates, standard errors and z-values and p-values. Reference levels = After AdvP, Canonical word order

Variable	Coefficient estimate	Std. error	z-value	p-value
Intercept	0.577	0.208	2.771	0.006
Trial Number	0.294	0.079	3.727	0.000
After AdvP	0.168	0.149	1.127	0.26
After VP	0.395	0.151	2.619	0.009
Adverbial first	0.165	0.124	0.133	0.894
Number in set	-0.544	0.110	-4.932	0.000

Random effects: by-word intercept (SD = 0.343), by-participant intercept (SD = 0.896), and by-participant random slope for sentence number in set (SD = 0.443), and the correlation between by-participant intercept and slope (r = -0.79).

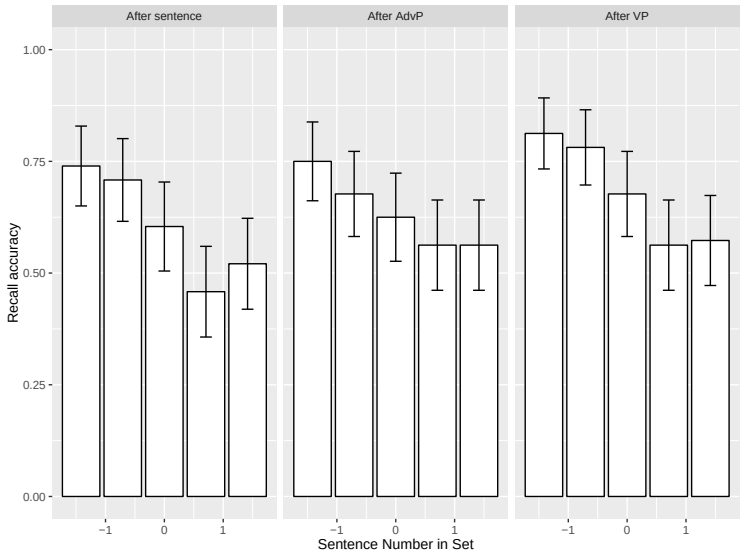

Fig. 5. Recall accuracy by sentence number in set and memory word location for sentences containing the atelic adverbial (*for-X-time*). Error bars represent 95 % confidence intervals.

3 General Discussion and Conclusions

To summarize our main findings, when we looked at the results collapsed across the two types of adverbials, we did not find significant differences between them, i.e., there was no main effect of adverbial type. However, we did observe a significant interaction between the syntactic position of the adverbial and the location in which the memory word was presented. These results suggested that memory word location effects were stronger in the adverbial first word order. To examine the potential effects of the adverbials directly, we analyzed the recall data based on the type of adverbial separately. Our results showed that there were differences in word recall performance based not only on the type of adverbial but also on its syntactic position in the sentence. Generally, these results suggest that the AdvP contributes to aspectual processing and therefore, that the VP is not the sole bearer of aspectual information. In addition, with respect to semantic commitment, our results provide direct evidence in favour of the *partial interpretation hypothesis* (Frazier & Rayner, 1990).

When the adverbial was *for-X-time* (atelic), we found an overall main effect of word location such that recall was more difficult at the end of the sentence compared to a sentence-internal position (After AdvP, After VP), suggesting that the parser was able to delay making a semantic commitment when the adverbial did not restrict the meaning of the event to a particular aspectual interpretation. This effect did not depend significantly on the word order of the sentence, i.e., whether the adverbial preceded or followed the VP. However, when the sentences contained *in-X-time* (telic), we found that recall was significantly worse when the memory word appeared following the AdvP, compared to the two other possible word locations. Crucially, we also observed an interaction between memory word location and the syntactic position of the adverbial such that the difference between After sentence and After AdvP was stronger when the adverbial appeared at the beginning of the sentence. These results do not support the *complete interpretation hypothesis* under which we would not expect any differences based on the aspectual interpretation of the adverbial since in both cases the parser needs to commit to

semantic interpretation as soon as it encounters the AdvP.

Our results provide strong evidence in favour of the *partial interpretation hypothesis* which assumes that the parser may delay making semantic commitments until later in the sentence in certain cases. As *in-X-time* restricts the interpretation to telic, a semantic commitment must be made as soon as the adverbial is processed. Such a commitment incurs a processing cost, as reflected by lower recall scores for memory words presented after the After AdvP. The significant interaction between the syntactic position of the adverbial and memory word location, however, suggests that the cost of making a semantic commitment to a telic interpretation as soon as the adverbial is processed differs based on when the adverbial is encountered in the sentence. When the adverbial appears early in the sentence (before the VP), the parser has a very early indication about the aspectual interpretation of the sentence and can therefore commit to a telic interpretation as soon as it reads the adverbial. In contrast, when the adverbial appears after the VP, it cannot commit to a particular interpretation until the adverbial is encountered, giving rise to later processing costs. As *for-X-time* does not restrict the aspectual interpretation of the event (it can be interpreted as completed or incomplete), the parser can delay making a semantic commitment until later in the sentence, irrespective of where the adverbial appears in the sentence. In atelic sentences, we found that word recall was worse after the sentence compared to the two other word locations. We interpret this effect as reflect a processing cost of committing to a semantic interpretation towards the end of the sentence. Since this effect did not differ based on the syntactic position of the adverbial, it seems that the parser can always delay committing to an atelic interpretation, even when the adverbial has been presented early in the sentence.

Similarly to previous experimental studies investigating telicity (Brennan & Pylkkänen, 2008; Stockall & Husband, 2014), we found processing differences based on the aspectual properties of phrases. Our results differ from previous studies in that we examined regions of the sentence beyond the VP in order to determine if other phrases contribute to aspectual interpretation. Indeed, we found that some temporal adverbials restrict the aspectual interpretation of the predicate. In future work, these adverbials should also be considered when examining the

processing of aspectual information. Overall, we did not find that either the telic or atelic interpretation was intrinsically more difficult (or more complex) than the other. Instead, our results suggest that both interpretations contribute to processing difficulty but differ based on timing, specifically, depending on when the parser commits to a semantic interpretation.

At a general level, these results suggest that Chapman, Deschamps, et al. (2016)'s variant of the complex span methodology is sensitive enough to investigate questions about semantic processing. By placing memory words within the processing sentences, we were able to tap into local effects that may not have been observable had we only measured recall ability at the end of the sentences.

Author Note

This research was partially supported by the Joseph-Armand Bombardier Doctoral Canada Graduate Scholarship 767-2014-1827 awarded by the Social Sciences and Humanities Research Council (SSHRC) to C. Chapman, and by a Natural Sciences and Engineering Research Council of Canada (NSERC) grant 311804-07 awarded to E. Service. We thank Bridget Copley and Malaree Baraniuk for discussion.

References

Baayen, H. R. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.

Bach, E. (1986). The algebra of events. *Linguistics and Philosophy*, *9*(1), 5–16.

Baraniuk, M. (2014). *Cognitively complex semantic processing in the domain of aspect.* (Unpublished Master's thesis, McMaster University)

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bates, D., & Sarkar, D. (2007). *The lme4 library.* Retrieved from `http://lib.stat.cmu.edu/R/CRAN/`

Borer, H. (2005). *In name only* (Vol. 1). Oxford University Press on Demand.

Brennan, J., & Pylkkänen, L. (2008). Processing events: Behavioral and neuromagnetic correlates of aspectual coercion. *Brain and Language*, *106*(2), 132–143.

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, *22*(1), 77–94.

Chapman, C., Deschamps, T., Kuperman, V., & Service, E. (2016). *Does simple sentence processing tax working memory? Evidence from complex span*. Manuscript submitted for publication.

Chapman, C., Kuperman, V., & Service, E. (2016). *Interaction between syntactic complexity and working memory: Evidence from complex span*. Manuscript submitted for publication.

Chapman, C., & Service, E. (2016, May). *Complex span evidence for shared resources between sentence processing and working memory*. Paper presented at International Meeting of the Psychonomic Society, Granada, Spain.

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology*, *33*(4), 497–505.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466.

Dowty, D. R. (1986). The effects of aspectual class on the temporal structure of discourse: semantics or pragmatics? *Linguistics and Philosophy*, *9*(1), 37–61.

Fedorenko, E., Gibson, E., & Rohde, D. (2006). The nature of working memory capacity in sentence comprehension: Evidence against domain-specific working memory resources. *Journal of Memory and Language*, *54*(4), 541–553.

Frazier, L., & Rayner, K. (1990). Taking on semantic commitments: Processing multiple meanings vs. multiple senses. *Journal of Memory and Language*, *29*(2), 181–200.

Garey, H. B. (1957). Verbal aspect in French. *Language*, *33*(2), 91–110.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*(1), 1–76.

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. A. O'Neil (Eds.), *Image, Language, Brain* (pp. 95–126). Cambridge, MA: MIT Press.

Giorgi, A., & Pianesi, F. (2001). Tense, attitudes and subjects. In R. Hastings, B. Jackon, & Z. Zvolenszky (Eds.), *Proceedings of Semantics and Linguistic Theory (SALT) XI* (pp. 212–230).

Gordon, P. C., Hendrick, R., & Levine, W. H. (2002). Memory-load interference in syntactic processing. *Psychological Science*, *13*(5), 425–430.

Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive Science*, *29*(2), 261–290.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the second meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies* (pp. 1–8).

Heim, I., & Kratzer, A. (1998). *Semantics in generative grammar*. Oxford: Blackwell.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*, 122–149.

Krifka, M. (1998). The origins of telicity. In S. Rothstein (Ed.), *Events and Grammar* (Vol. 70, p. 197-235). Dordrecht: Kluwer.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*(3), 375–419.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences*, *10*(10), 447–454.

Link, G. (1983). The logical analysis of plurals and mass terms: A lattice-theoretical approach. In R. Bäuerle, C. Schwarze, & A. von Stechow (Eds.), *Meaning, use, and interpretation of language* (p. 302-323). Berlin: de Gruyter.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed effects models in S and S-PLUS*. New York: Springer.

Pustejovsky, J. (1991). The generative lexicon. *Computational Linguistics*, *17*(4), 409–441.

Ramchand, G. C. (2008). *Verb meaning and the lexicon: A first phase syntax* (Vol. 116). Cambridge: Cambridge University Press.

Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition*, *116*(1), 71–86.

Stockall, L., & Husband, E. M. (2014). Processing (the) events: Lexical and structural ingredients of inner aspect. *Connectedness: Papers by and for Sarah Van Wagenen. UCLA Working Papers in Linguistics*, *18*.

Townsend, D. J. (2013). Aspectual coercion in eye movements. *Journal of Psycholinguistic Research*, *42*(3), 281–306.

Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language*, *47*(1), 69–90.

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

Verkuyl, H. (1993). *A theory of aspectuality (Cambridge Studies in Linguistics 64)*. Cambridge: Cambridge University Press.