Making semantic commitments can be delayed:

Evidence from aspectual processing

Jitka Bartošová[1], Cassandra Chapman[2], Ivona Kučerová[1]and Elisabet Service[1]

[1]Department of Linguistics and Languages, McMaster University, Hamilton, Ontario, Canada

[2]Department of Linguistics, University of Toronto, Toronto, Ontario, Canada

Address for correspondence:

Cassandra Chapman

University of Toronto, Department of Linguistics

100 St George Street

Toronto, Ontario, Canada M5S 3G3

E-mail: cassandra.chapman@utoronto.ca

Telephone +1-(416)-978-4029

Funding: 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.

Conflict of Interest: The authors declare that they have no conflict of interest.

Running head: MAKING SEMANTIC COMMITMENTS

Making semantic commitments can be delayed: Evidence from aspectual processing

Semantic interpretation of aspectual verbs has been shown to cause a processing cost. The present study provides additional evidence that the semantic interpretation of events interacts with sentence processing. The study focused on telicity, an aspectual property that does not solely depend on lexical items but instead on the semantic composition of verb phrase (VP)-level events. 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* adverbials force the parser to commit to a telic (completed/terminated) interpretation before the VP has been processed. In contrast, *for-X-time* adverbials, which are compatible with an atelic (completed or incomplete) interpretation, do not force the parser to make an early commitment to a particular semantic interpretation. Instead, processing is always delayed until the VP has been completely parsed. Results support the partial interpretation hypothesis according to which the parser can delay making semantic commitments until it is necessary to do so, i.e., in atelic but not telic sentences.

Public significance statement: This work presents results from a working memory task demonstrating that the semantic interpretation of verbal events has an effect on sentence processing. Results demonstrate that committing to a semantic interpretation of an event incurs a processing cost (lower working memory capacity) but that the cognitive system responsible for sentence processing only commits to a certain interpretation when it is forced to do so. These results demonstrate that the semantic interpretation of events has an effect on human cognition (sentence processing) and working memory capacity more broadly.

1 Introduction

1.1 Goal of the study

The current study aims to provide a better understanding of the semantic processing of events, specifically in the domain of telicity, an aspectual property associated with the semantics of events. Although there is a significant amount of work that addresses the issue of event processing (for example, McElree, Traxler, Pickering, Seely, & Jackendoff, 2001; De Almeida, 2004; Pickering, McElree, & Traxler, 2006; Pylkkänen & McElree, 2006; Bott & Hamm, 2014; Stockall & Husband, 2014; Piñango & Deo, 2016; Philipp, Graf, Kretzschmar, & Primus, 2017), there are still unanswered questions about the exact time course of semantic processing in relation to syntactic processing. The question of the time course of semantic and syntactic processing has been asked, for instance, in work by Frazier and Rayner (1990), who propose two competing hypotheses about the timing of semantic processing: the *immediate partial interpretation hypothesis* and the *immediate complete interpretation hypothesis*. Frazier and Rayner (1990) 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[1] 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 create the need to maintain multiple interpretations for a word, phrase or a relation, or the structure would not be interpreted at all.[2] While the immediate complete interpretation hypothesis predicts strictly deterministic and linear semantic processing, its partial counterpart allows for the time course of semantic processing to deviate from syntactic processing.

To investigate which of the two hypotheses provides better empirical coverage, the present

[1] By "parser," we mean the cognitive system responsible for structure building in real-time processing.
[2] Frazier and Rayner (1990) were interested in the question of polysemy, i.e., the interpretation of lexical items (see, e.g., Frisson (2009) for an overview), but the core proposal has since been extended to the processing of semantic structures, including aspectual event properties (e.g., in Piñango & Deo, 2016).

study explored *telicity*, an aspectual property of linguistic events, by isolating distinct semantic components that contribute to telicity. 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 terminated (has been completed), yielding the distinction between telic (terminated) and atelic (unspecified with respect to whether the event reached its culmination point) events (Vendler, 1957/1967; Dowty, 1979).[3] In English, the telic interpretation is not determined by a single lexical item. Instead, telicity arises from the verb, its argument(s), temporal adjuncts, and, in some cases, from the pragmatic context (e.g., Verkuyl, 1972; 1993; Dowty, 1979; Pustejovsky, 1991; Krifka, 1989; 1992; 1998; Tenny, 1987; 1994; Giorgi & Pianesi, 2001). That is, telicity is a property of VPs.[4] As shown in (1),the 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 (1a), but if the verb selects for an indefinite NP argument, the completion of the event remains unspecified (atelic), as in (1b).[5]

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

b. *Atelic*: Peter ate cookies.

Some verbs are more specific in their lexical semantics. As shown in (2), the interpretation of the *lose* event is telic, irrespective of the type of NP argument that it selects (definite or indefinite

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

[4] This is a simplification. A number of proposals argue that telicity is located in a functional projection above VP or vP (e.g., Folli, 2001; Borer, 2005; Ramchand, 2008; Rothstein, 2008). We use the term VP as a cover term for the extended nominal projection, including the corresponding aspectual projections.

[5] The literature is split on the question of whether atelicity is a specific semantic interpretation (e.g., Rothstein 2004; Dowty, 1991; Kennedy & Levin, 2008; Krifka, 1989, 1992; Tenny, 1994; Verkuyl, 1993), or the lack of telicity, i.e., of event termination or culmination (e.g., Borer, 2005; Schein, 2002, Stockall & Husband, 2014). We side with the underspecified approach but the distinction does not play a crucial role in the present study.

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, (3).

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

b. *Telic*: Peter lost cookies.

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

b. *Atelic*: Peter loved cookies.

Crucially, some predicates are ambiguous between telic and atelic interpretations, irrespective of the verb's NP argument (if any), as shown in (4).

(4) a. *Telic/Atelic*: Peter ironed the clothes.

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

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

Moreover, the distinction between telic/atelic interpretations arises with certain temporal adjuncts. As shown in (5), the adverbials *in-X-time* and *for-X-time* can be used as a test to distinguish between telic and atelic events independent of the lexical semantics of the verbs (Vendler, 1957/1967; Dowty, 1979).

(5) a. *Telic*: Jessica 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 (6), and consequently restrict the interpretation of the event to telic or atelic.

(6) a. *Telic*: The boy scrubbed the pots **in five minutes**.

b. *Atelic*: The boy scrubbed the pots **for five minutes**.

The fact that telicity is not a property of lexical items per se but depends on larger constituents, yet the semantic contribution of telicity may be localized to a single structural element (in the present study, temporal adverbials), provides us with a tool to investigate semantic processing at

the sentential level, and in turn, to distinguish between the *complete* and *partial interpretation hypotheses* (Frazier & Rayner, 1990).

1.3 Processing of telicity

Previous experimental work on the processing of aspectual events consistently reports that the semantic processing of telicity is associated with a processing cost that is often delayed (e.g., McElree et al., 2001; Pickering et al. 2006, Pylkkänen & McElree, 2006; Malaia, Wilbur, & Weber-Fox, 2009; Katsika, Braze, Deo, & Piñango, 2012; O'Bryan, Folli, Harley, & Bever, 2013, Stockall & Husband, 2014; Piñango & Deo, 2016). However, it is not clear why there should be a processing cost.[6] In order to explain their experimental findings that certain semantic interpretations can be delayed, Frazier and Rayner (1990) assume that making a semantic commitment to a particular interpretation is costly. However, 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 processing cost and its delay is surprising.[7] In order to account for the processing cost, we follow theoretical and experimental literature that argues for an additional semantic process that anchors temporal (or spatio-temporal properties) of the event and often requires the introduction of an additional semantic operator that is absent in the syntactic representation (see, for instance, the role of quantization in Krifka, 1989; 1992; 1998, maximalization in Filip, 2004 or Rothstein, 2008, a maximal stage operator in Artshuler, 2014, or placing the event in the spatial expanse and in the time interval's world in Guéron, 2008 and Dahl, 2013; cf. also the discussion of processing cost in Piñango & Deo, 2016).[8]

[6] O'Bryan et al. (2013) specifically associate the higher cost of atelic events with particular syntactic properties in their study (and its predecessors), namely, the structure of reduced relative clauses. However, as discussed in Piñango and Deo (2016), the processing profile extends beyond this particular structural configuration.

[7] We would like to thank two anonymous reviewers for raising the question about costliness of semantic interpretations.

[8] We remain agnostic as to which proposal for semantic computation of telicity is the most adequate one. For the

We argue that the processing cost systematically attested in aspectual experimental studies reflects this additional semantic component that is necessary for the semantic processing of telic properties of a VP event.

However, if the additional semantic building at the VP level were the complete explanation, we would expect the effects of telicity in processing to always occur late. Yet, early processing effects of telicity have been found in several studies (e.g., Malaia et al. 2009; O'Bryan et al. 2013; Philipp et al., 2017).[9] A direct comparison of these studies is not possible as they differ in the type of predicates they use (some compare lexically telic and atelic predicates, some allow for ambiguous predicates), and the syntactic complexity of tested structures (often reduced relative clauses). Yet, the studies provide evidence that under certain circumstances the processing cost can occur early.

Interestingly, theoretical work on telicity proposes that the telic interpretation corresponds to an additional syntactic structure (e.g., Folli, 2001; Borer, 2005; Ramchand, 2008), or that the telic inducing temporal adverbial (in-X-time) in and of itself requires an additional structure (Giorgi & Pianesi, 2001, following Larson, 1999). We hypothesize that there should be structures in which the parser receives explicit structural cues to build the additional structure required for telic interpretations, and structures in which there is no overt cue to guide the corresponding semantic interpretation. That is, the time course of semantic processing of telicity is more fine-tuned, and offers itself as a testing ground for contrasting the *complete* versus *partial interpretation hypotheses* (Frazier & Rayner, 1990).

The current study builds on the general assumption that if we isolate the constituent that brings in information about telicity, we can make predictions about how the processing of telic and atelic events unfolds. Our main question is thus, at which point does the parser commit to a

---

purposes of the current study, it suffices to assume that there is an additional semantic operation that can be associated with the processing cost.

[9] We thank an anonymous reviewer for bringing the importance of these studies to our attention.

semantic interpretation? Does the parser interpret the event as telic or atelic as soon as it encounters a structural element that brings in information about telicity, or can the interpretation be delayed? Before we lay out our experimental design, we will briefly summarize previous experimental work on telicity, introducing 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 VP is processed. The VP has been argued to be the relevant aspectual domain such that the locus of aspectual information is exclusively in the VP (e.g., Verkuyl, 1972; 1993; Dowty, 1979; Pustejovsky, 1991). 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 whether a potential disruption of aspectual processing leads to processing difficulty. This question has been investigated using *aspectual coercion* (Pustejovsky, 1991)[10] contexts where a VP with a particular aspectual profile (telic or atelic) is presented with an adverbial phrase (AdvP) that has a mismatching aspectual profile (e.g., Brennan & Pylkkänen, 2008; Townsend, 2013) as well as in garden-path sentences (created by reduced relative clauses; e.g., Malaia et al., 2009; O'Bryan et al., 2013). These mismatches are predicted to lead to a processing cost resulting either from semantic re-interpretation (in the case of aspectual coercion), or syntactic re-interpretation (in the case of garden path sentences). For example, Brennan and Pylkkänen (2008) examined aspectual coercion in sentences like (7a) and (7b). They predicted a higher processing cost in (7a), which contains an AdvP and a VP that mismatch in telicity, i.e., an atelic adverbial combined with a telic predicate, compared to (7b), in

[10] 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.

which the AdvP and VP have the same aspectual profile. The authors found that the VP was indeed read more slowly in aspectual coercion conditions, (7a), compared to control conditions, (7b). Brennan and Pylkkänen (2008) argue that these results reflect a higher processing cost associated with mismatching aspectual profiles, compared to matching profiles.

(7) 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 enforcing re-analysis. Stockall and 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 (8a)-(8c), or to an atelic interpretation, as in (8d).

(8) 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 and Husband (2014) found longer reading times on the direct object, more precisely, on a prepositional phrase modifying the nominal, i.e., *on*, in the unspecified (atelic) condition in (8d), compared to this same word in the three other conditions. 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. Note that technically the cost occurs before the complete direct object is processed (the cost is on the

modifier of the object). However, from the processing point of view, the processing cost occurs after the properties of the direct object that are relevant for telicity (the determiner and the head noun, here bare plural) have already been syntactically processed.[11]

In a second self-paced reading experiment, Stockall and Husband (2014) compared unspecified verbs, such as *inspect*, and verbs lexically specified as atelic, such as *roam*, as in (9a) and (9b). 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. The results of Stockall and Husband (2014), 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.

(9) a. *Unspecified*: The local horticulturist inspected (the) gardens in the neighborhood.
b. *Atelic*: The local horticulturist roamed (the) gardens in the neighborhood.

Their result is fully consistent with other studies that have demonstrated a higher cost for the atelic condition (e.g., McElree et al., 2001; Pickering et al., 2006; Pylkkänen & McElree, 2006; Malaia et al., 2009; Katsika et al., 2012; O'Bryan et al., 2013; Piñango & Deo, 2016). One needs to be careful, however, about the interpretation of the additional cost, as in these experiments, there was always an explicit lexical or syntactic trigger for telicity. Under the assumption that telic interpretations can be aided by syntactic structure building, the observed cost may be a side-effect of atelic interpretations lacking an explicit syntactic processing trigger.

To summarize, there are three main findings from previous experimental work on telicity that are relevant for the current study: i) adverbials affect the semantic processing of telicity (aspectual coercion), ii) no difference in the processing profiles could be detected between sentences containing verbs lexically specified as atelic and unspecified verbs, (iii) an atelic/unspecified interpretation appears to incur more of a cost than its telic counterpart. These

[11] Malaia et al. (2009) observe that the exact timing of the processing cost correlates with syntactic proficiency of the participants. In their study, the effect of telicity brought in by the lexical properties of the predicate appears earlier for high syntactic proficiency participants. Stockall and Husband didn't test for syntactic proficiency.

findings clearly suggest that aspectual information affects processing, that the VP is relevant for aspectual processing, and that the semantic processing of VPs can differ depending on the lexical and syntactic content of the VP.

2 Current study

The current study builds on Stockall and Husband's (2014) conclusion that telic and atelic (unspecified) events have distinct aspectual processing profiles. We also follow Brennan and Pylkkänen (2008) in assuming that AdvPs can affect the aspectual interpretation of verbs, and therefore, the semantic processing of events. Unlike previous studies, we specifically use adverbials as a tool to uncover a more precise time course of the processing of telicity. Similarly to Frazier and Rayner (1990), we assume that making a semantic commitment to a particular interpretation is costly (albeit with the clarification that the cost arises from the added semantic component attested with telicity, instead of regular semantic composition; see the discussion in section 1.3). 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. Moreover, we assume that the parser can only make an early commitment to a telic interpretation as only telic interpretations are specified. Once such an early commitment occurs, we do not expect any later processing cost. In contrast, if the parser does not encounter any clear indication that would aid the telic interpretation, then we expect a processing effect to occur later. Namely, we predict that the later processing effect should occur whenever the parser no longer expects additional telicity related structure building (i.e., when there is a clear indication of the VP having been completed, which typically coincides with the end of the clause). We used a working memory task to detect those processing costs, and we manipulated the position and the type of temporal adverbial in our experimental design.

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 processing. In the current study, we ask: Do *in-X-time* and *for-X-time* adverbials play a role in the processing of aspectual information? First, since temporal adverbials can attach either at the VP level or at the clause level (TP), we ask whether the processing of telicity can extend beyond VPs, and if that is the case, at what point of processing we see a processing cost. Second, according to Stockall and Husband's findings, there should be no processing difference between atelic and unspecified predicates. Yet, lexically unspecified predicates can combine with adverbials that enforce a particular semantic interpretation (*in-X-time* vs. *for-X-time*). We ask whether adding such an adverbial will yield a processing cost. A cost would be expected if such a structure (unspecified predicate combined with an adverbial that forces a particular interpretation) technically constitutes aspectual coercion. An alternative is that we will see a processing cost associated with telic predicates (that is, no effect of coercion is expected with truly unspecified predicates). 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 and that they do so even when they are attached at the clause level, and that they do not trigger an aspectual coercion cost when combined with unspecified predicates.

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 (4), and (ii) the assumption that *in-X-time* adverbials restrict the aspectual interpretation of sentences containing ambiguous predicates to telic, while *for-X-time* adverbials yield the atelic, i.e., unspecified, interpretation, as previously shown in (6) and repeated here as (10).[12]

(10) a. *Telic*: The boy scrubbed the pots **in five minutes**.

b. Atelic: The boy scrubbed the pots **for five minutes**.

[12] We thank Bridget Copley for suggesting that we use a minimal pair design with ambiguous verbs and temporal adverbs to examine the processing of telic versus atelic events.

As the ambiguous predicates are unspecified with respect to telicity, they are fully compatible with both types of AdvPs. Therefore, we predicted that any potential processing differences between (10a) and (10b) 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 (11a), or before the VP (at TP), as in (11b). The sentence-initial position of the AdvP (called *adverbial first*) allowed us to investigate whether the AdvP can trigger an immediate semantic commitment of 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 VP region.

(11) a. *Canonical*: The boy scrubbed the pots **in/for five minutes**.

b. *Adverbial first:* **In/for five minutes** The boy scrubbed the pots.

Furthermore, as discussed above, we predicted that only *in-X-time* forces a specific interpretation of the event (telic); *for-X-time* yields unspecified interpretation.

Using two distinct syntactic positions for the adverbials in combination with Frazier and Rayner's (1990) 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. According to the *immediate complete interpretation hypothesis*, the parser must interpret each phrase as soon as it is encountered. That is, we expect a processing cost immediately after the *in-X-time* adverbial is encountered, irrespective of its sentential position, because this adverbial forces a telic interpretation of the event. When the *for-X-time* adverbial is in sentence initial position, we expect the processing cost to occur after the verb and its object (that is when the VP-based computation of telicity can, and therefore, must be completed). Finding this adverbial on its own is not strong enough indication to the parser about how the event needs to be interpreted. In contrast, when the *for-X-time* adverbial appears in canonical position, we expect the processing

cost to appear after the adverbial (as the indicator of the relevant part of the VP being completed). We do not expect any downstream effects. In contrast, following the *immediate partial interpretation hypothesis*, the parser may delay making semantic commitments to a particular interpretation in some circumstances. We predict that 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.[13] 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, Kuperman, & Service, 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)

[13] An anonymous reviewer asks whether we also predict any downstream effects. The idea is that if the parser commits to a semantic interpretation early (as in the case of telic adverbials at the beginning of the sentence), perhaps processing later in the sentence (further to the right) should be easier. This is an excellent question and at this point, we cannot commit to such a prediction. Our study specifically investigated local effects and is not an online study of sentence processing. A future study investigating online effects (self-paced reading, eye-tracking) would be better able to address this question.

read and comprehend sentences, and ii) remember memory words for later recall in the order they were presentedto 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, Gibson, & Rohde, 2006; Gordon, Hendrick, & Levine, 2002, among others). The adapted versions of the task, also referred to as *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 (e.g., object- versus subject-extracted relative clauses).

Chapman 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, and Service (2016); Chapman and Service (2016) investigated sentences containing subject- or object-extracted relative clauses, see (12a) and (12b), respectively, and found that participants were less likely to recall memory words placed within more complex sentences, i.e., object relative clauses, suggesting that this variant of complex span is sensitive enough to capture effects of syntactic complexity.

(12) 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, Vasishth, & Van Dyke, 2006; Staub, 2010; Traxler, Morris, & Seely, 2002; Chapman, Kuperman, & Service, 2016; Chapman & Service, 2016, among others).

We propose that Chapman et al.'s (2016) 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 naive as to the purposes of the present study.

2.2.2 Stimuli

Forty-five predicates[14] 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's (2014) study, stimuli consisted of minimal pairs in which the sentences differed only based on which adverbial was used, as in (13a) for the telic interpretation and (13b) 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's (2014) experiment, stimulus sentences always consisted of a full NP subject, a verb in the past tense, an NP argument of the verb (in most cases),[15] and the critical temporal adverbial phrase.

(13) a. *Telic*: The boy scrubbed the pots **WORD1** in five minutes. **WORD2**

b. *Atelic*: The boy scrubbed the pots **WORD1** for five 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 (14a)-(14d). In addition to the original conditions, we also included two conditions in which the critical adverbial phrase appeared before the VP, as in (14c)-(14d).[16] Memory words were presented in three locations in order to examine the effect of memory encoding on processing

[14] All predicates are so-called scalar predicates, that is, predicates denoting events that can denote a change, a necessary factor for the event to be able to culminate.

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

[16] In the conditions in which the AdvP appeared first, a pronoun subject was used instead of a full NP, compare (14a)-(14b) to (14c)-(14d). 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.

incrementally. Doing so allowed us to investigate local effects and, specifically, examine whether the aspectual properties of the adverbials affect memory encoding. Across all conditions, the memory word appeared after the VP, shown as WORD1 in (14a)-(14b) and as WORD2 in (14c)-(14d). Memory words could also appear after the AdvP, as WORD2 in (14a)-(14b) and WORD1 in (14c)-(14d). 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.

(14) a. *Telic, AdvP after VP*: The boy scrubbed the pots **WORD1** in five minutes **WORD2** at the restaurant. **WORD3**

b. *Atelic, AdvP after VP*: The boy scrubbed the pots **WORD1** for five minutes **WORD2** at the restaurant. **WORD3**

c. *Telic, AdvP before VP*: In five minutes **WORD1** he scrubbed the pots **WORD2** at the restaurant. **WORD3**

d. *Atelic, AdvP before VP*: For five minutes **WORD1** he scrubbed the pots **WORD2** at the restaurant. **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, (14a), (ii) atelic sentences with canonical word order, (14b), (iii) telic sentences with the fronted temporal adverbial phrase, (14c), and (iv) atelic sentences with the fronted temporal adverbial phrase, (14d). 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 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 (for example, the first trial in one list would be presented as the last trial in another list). and counterbalance of trials from the first three lists (for example, if one list contained atelic adverbial first, the corresponding list contained atelic canonical word order).

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 (14a)-(14d) 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.[17]

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 (14a), the statement was: "The boy scrubbed the pots at home." (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 had appeared early in the set of 5.[18]

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 position. 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, (15a) and (15b) 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

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

[18] As an anonymous reviewer points out, due to the nature of the task, this experiment was essentially a dual memory task: i) remember the memory words for later recall, and ii) remember the sentences for the control statement task. We attribute the low accuracy in the control statement task to all of these factors.

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 a white background in the middle of the screen (Lucida Grande font, size 36 pt). Memory words appeared for 1000 ms in bold red capital letters (Lucida Grande font, size 48 pt) in the centre of the screen in one of the three memory word locations within a sentence.

(15) a. *Canonical order - regions*: / The boy / scrubbed the pots / in five minutes / at the restaurant./

b. *Adverbial-first order - regions*: / In five minutes / he / scrubbed the pots / at the restaurant./

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. We outlined our predictions at the beginning of section 2. We will now briefly relate those predictions to the complex span task. As previously discussed, *in-X-time* restricts the aspectual interpretation of the event such that it must be interpreted as completed. Consequently, 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. Thus, we predict that encountering the *in-X-time* adverbial should be costly.

The dependent measure in this task is memory word recall. We predict that we should observe lower memory word recall in locations in which we predict a processing cost. Thus, 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. 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 *in-X-time* 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 (completed, or telic). 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, if the adverbial appears after the VP, the parser will not be able to adopt a particular aspectual interpretation until later when it has processed the adverbial.

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.

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 (Baayen, Davidson, & Bates, 2008; 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, Levy, Scheepers, & Tily, 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 these adverbs 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 et al., 2016). Note that 0 in our model represents sentence number 3. We further 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/scaled 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 (*in-X-time*) had numerically higher recall scores compared to those with atelic adverbials (*for-X-time)*, this effect was not statistically significant (β = 0.045, SE = 0.087, z = 0.518, p = 0.604, model not shown[19]). Note that a difference based on adverbial type in and of itself 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). We predicted a difference between the two adverbials based on their location in the structure and thus an overall main effect of adverbial type was 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 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

[19] 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.

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.

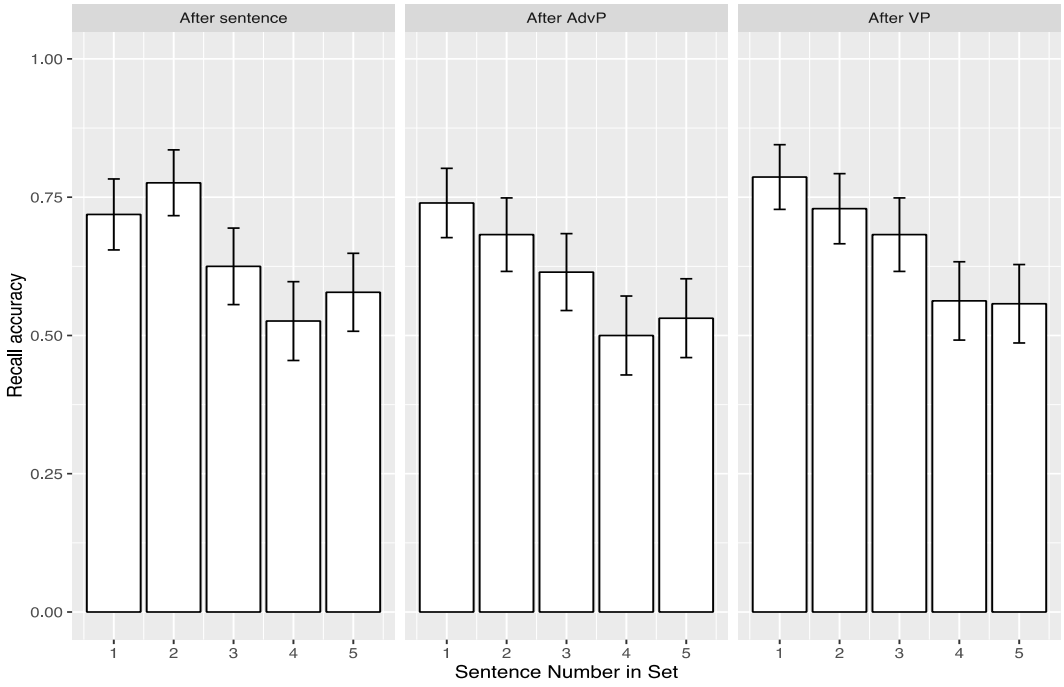

Fig. 1. Recall accuracy by sentence number in set and memory word location. Error bars represent 95 % confidence intervals.

The most interesting effects are interactions. There were significant or marginal interactions between the location of the memory word in the sentence and the syntactic position of the adverbial (canonical vs. adverbial first), as shown in Figure 2 and as reported in Table 1. 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 1, and compared to After VP: β = 0.363, SE = 0.213, z = 1.703, p = 0.09, model not shown).

Table 1

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).

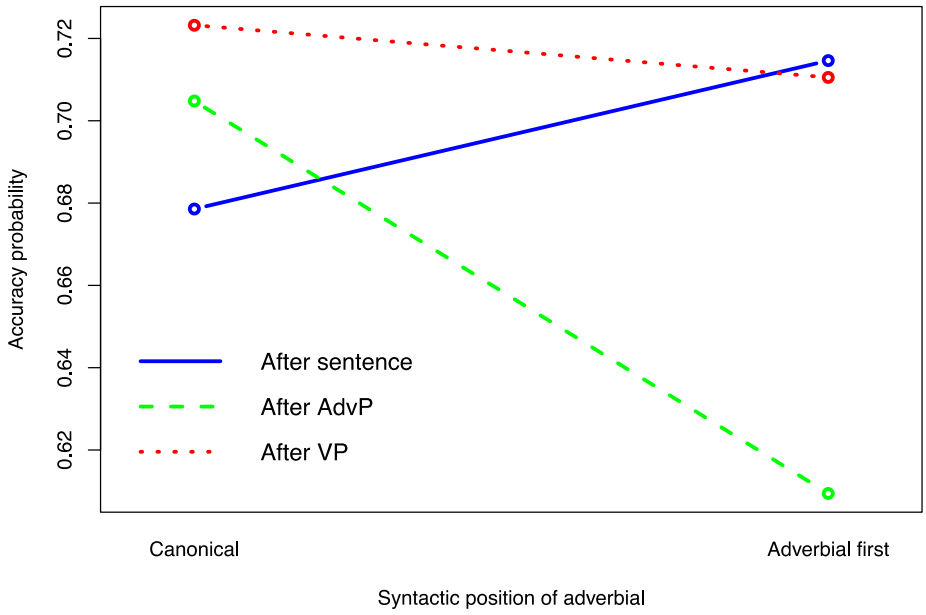

Fig. 2. Interaction plot showing an effect of memory word location by the syntactic position of the adverbial.

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.

2.4.1 In-X-time

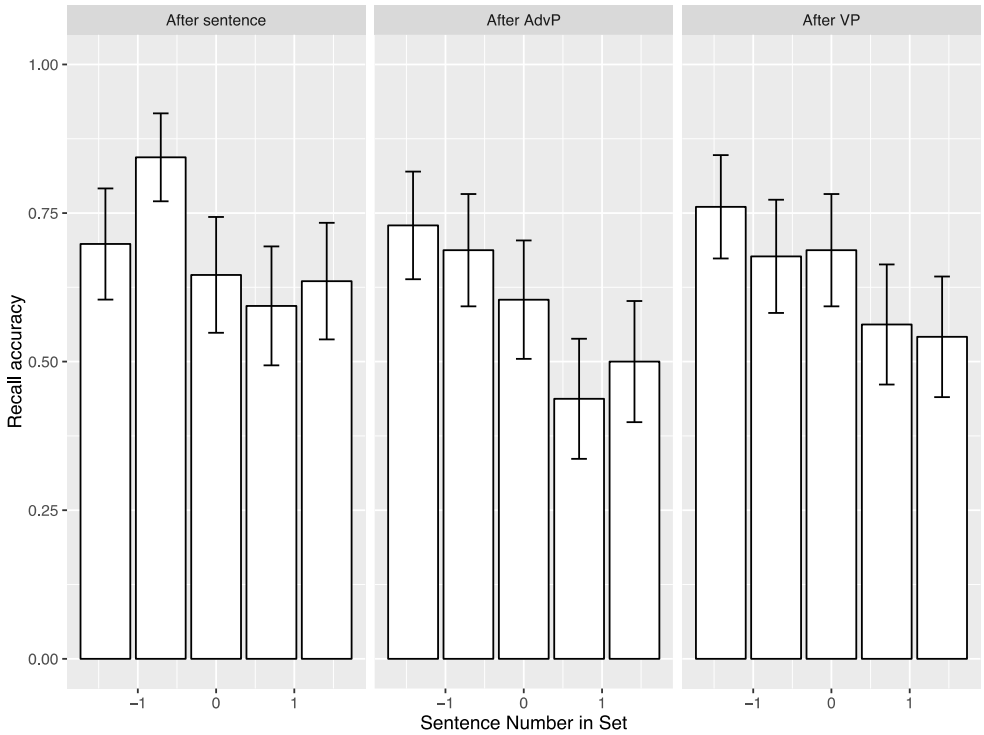

Fig. 3. Recall accuracy by sentence number in set and memory word location for sentences containing the telic adverbial (*in-X-time*). Error bars represent 95 % confidence intervals.

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 boy scrubbed the pots in five minutes at the restaurant./In five minutes he scrubbed the pots at the restaurant.*), which restricts the interpretation to telic. 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), as shown in Figure 3. 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 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 in canonical position ($p < 0.01$, Table 2). 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 2

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).

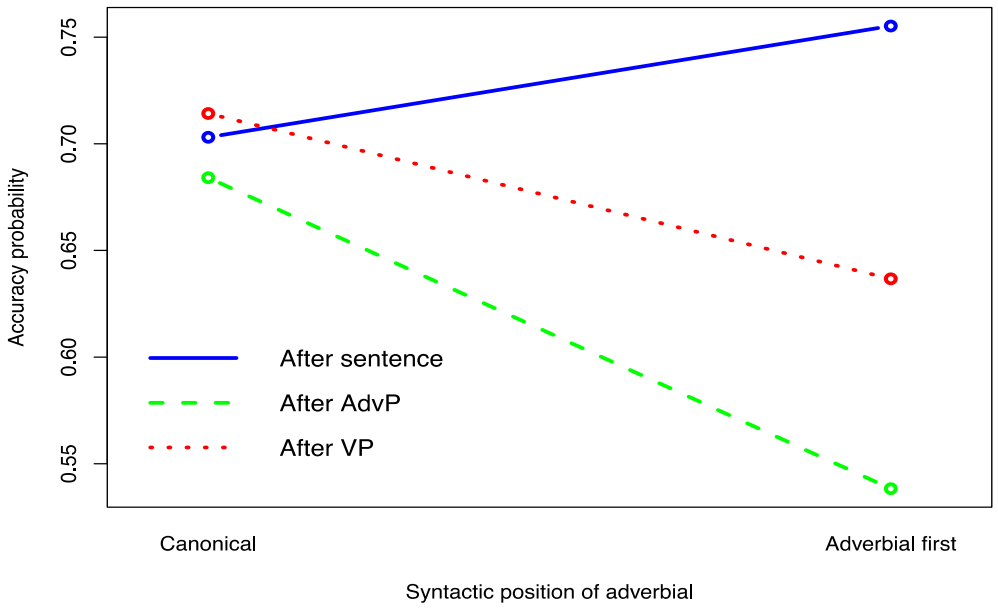

Fig. 4. Interaction plot showing an effect of the syntactic position of the adverbial on recall accuracy by memory word location for telic sentences.

2.4.2 For X-time

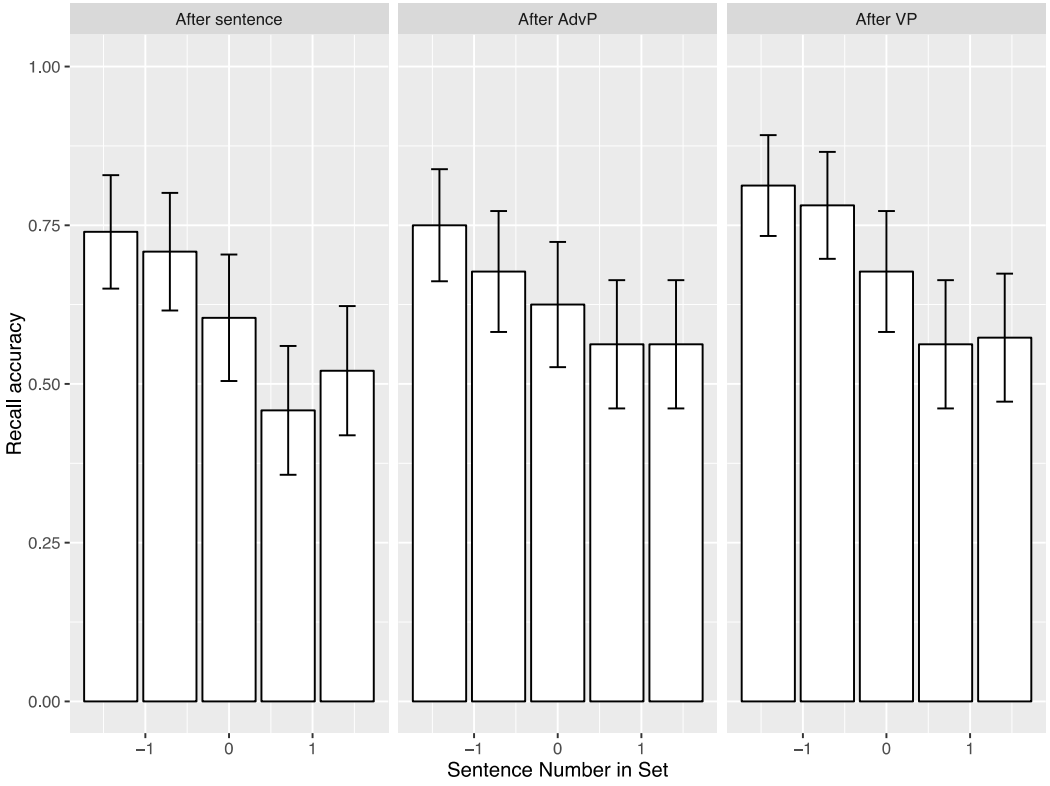

Fig. 5. Recall accuracy by sentence number in set and memory word location for sentences containing the atelic adverbial (*for-X-time*). Error bars represent 95 % confidence intervals.

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 boy scrubbed the pots for five minutes at the restaurant./For five minutes he scrubbed the pots at the restaurant.*), which does not force a particular aspectual interpretation of the sentence. 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 and Table 3. Recall was also numerically less likely After the sentence compared to After the AdvP but this effect was not significant. 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, as shown in Figure 6. We will discuss the theoretical implications of these results in the General Discussion section.

Table 3

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 sentence, 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).

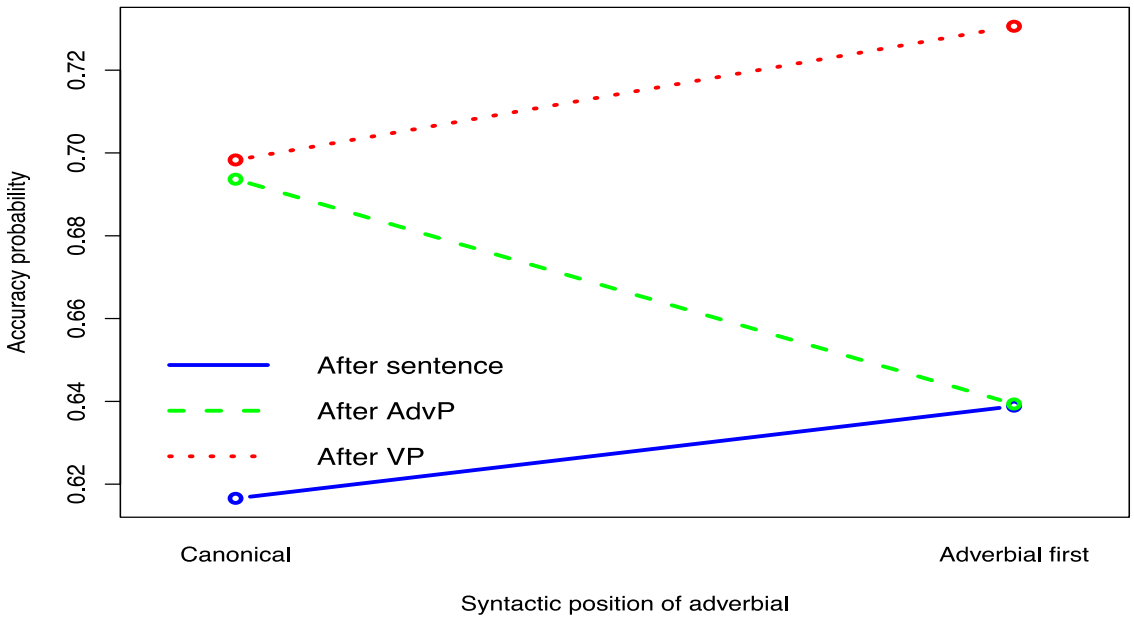

Fig. 6. Interaction plot showing an effect of the syntactic position of the adverbial on recall accuracy by memory word location for atelic sentences. This interaction was not significant at the 5% threshold.

3 General Discussion

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 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 directly support the *complete interpretation hypothesis* under which we would expect that the parser will commit to an atelic interpretation after the VP when the adverbial appears first (i.e., before the VP) but only after the adverbial has been processed when it appears in canonical position. Our results suggest that recall was more difficult at the end of the sentence in atelic sentences, irrespective of the position of the adverbial.

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 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. This leads to processing costs at the beginning of

the sentence, when the commitment has to be made but then processing is easier (higher word recall) later in the sentence. In contrast, when the adverbial appears after the VP, the parser 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.[20] In atelic sentences, we found that word recall was worse in the After sentence word location compared to the two other word locations. We interpret this effect as reflecting 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 (O'Bryan et al., 2013 Brennan & Pylkkänen, 2008; Malaia et al., 2009; Stockall & Husband, 2014; Philipp et al., 2017), 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 the 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.

[20] An anonymous reviewer asks whether we have any evidence that the parser makes a commitment at all in atelic sentences. We would like to suggest that processing difficulty (lower word recall) at the end of the sentences in atelic sentences suggests that this is when the parser interprets the sentence. Thus, we find no evidence of commitment to a semantic interpretation until the whole structure has been processed. Whether there might be effects at other earlier regions in the sentence in more fine-grained online processing measures remains an open question that we leave for future work.

4 Conclusions, Limitations, and Future Directions

Here we outline the current study's limitations and possible future directions. Malaia et al. (2009) found different effects based on participants' syntactic proficiency. In particular, an effect of telicity was found earlier for high proficiency participants. In our sentence span experiment, we did not test for participants' syntactic proficiency. It is possible that grouping our participants based on their syntactic proficiency may have led to different results. Following Malaia et al. (2009), we might predict that participants with high syntactic proficiency might show an effect of telicity (as measured by word recall) earlier than participants with lower syntactic proficiency. It is also possible that syntactic proficiency may not have played a role in our experiment. We used simple sentences whereas Malaia et al. used complex sentences (reduced relative clauses). We leave this question for future work.[21]

Moreover, our study measured processing of telicity using an offline measure (memory word recall). Previous studies used online methodologies, such as self-paced reading, eye-tracking, and ERPs. Offline and online measures of processing have different levels of sensitivity. Our working memory task might not have been sensitive enough to measure processing at all levels of structure, and consequently, we might not have seen fine-grained effects due to this methodology. However, even though we do not see effects at each word or within smaller phrases (for example, we did not measure word recall after the preposition in the AdvP), we did find larger effects at the end of larger constituents, i.e., once more structure had been built. We hypothesize that the difference between online and offline methodologies underlies the reason why our results appear to be at odds with those of Stockall & Husband (2014). They did not find a strong effect at the end of the sentence, which is where we predict the semantic commitment to an atelic interpretation has to be made. The effect found in their self-paced reading task was earlier (within the NP at the preposition). We suggest that the cost they found may only have

[21] We thank an anonymous reviewer for making us aware of this study and its importance to our own work.

been tapping into part of the effect. In a self-paced reading experiment, reading times measured at the end of the sentence are considered wrap-up effects and are generally not considered reliable for understanding sentence processing effects. Online measures cannot target the end of the sentence for measuring processing costs. In contrast, the end of the sentence is often used as a measure in sentence span studies. As a result, even though our working memory task may not have been sensitive to fine-grained effects, we suggest that this methodology allowed us to better understand more refined domains of semantic commitment. Future work should investigate whether it might also be possible to find such effects using real-time measures.[22]

Finally, there are also limitations stemming from the theory of semantic processing we used to guide the current work. There are many sentence processing models that we have not considered in the current paper. However, the nature of our task (working memory) limits what predictions can be made, which in turn affects what theories make meaningfully different predictions for the task at hand. We leave the extension to other processing theories for future work.

In such work, our stimuli could be tested using an online measure of sentence processing. It would also be beneficial to think more deeply about other theories of semantic and syntactic processing and how such results can be accounted for. As previously mentioned, Stockall and Husband (2014) were not able to examine online processing at the end of the sentence because this region is subject to wrap-up effects and therefore, reading time is not a reliable measure of processing there. However, future work could consider adding a later online measure of processing to investigate whether the results reported in the current paper can be replicated online. It might be possible to increase the length of the sentences by adding optional phrases and investigating whether processing costs are observed later also in online measures.

In conclusion, our results from a working memory task suggest that adverbials can also contribute to semantic processing. Moreover, by controlling for when this information is given to the parser (either before the VP or afterwards), we were able to tap into timing processes that may not have been detectable otherwise.

[22] We thank an anonymous reviewer for excellent questions that made us think more deeply about these issues.

Author Note

References

de Almeida, R. (2004). The effect of context on the processing of type-shifting verbs. *Brain and Language*, 90, 249–61. 10.1016/S0093-934X(03)00438-3

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. 10.1016/j.jml.2007.12.005

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. 10.1016/j.jml.2012.11.001

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

Borer, H. (2005). *Structuring sense, vol. 2: The normal course of events*. Oxford, UK and New York: Oxford University Press. 10.1093/acprof: oso/9780199263929.001.0001

Bott, O., & Hamm, F. (2014). Cross-linguistic variation in the processing of aspect. In Hemforth, B., Mertins, B., & Fabricius-Hansen, C. (Eds.). *Psycholinguistic approaches to meaning and understanding across languages*, pp. 83-109. New York: Springer.

Brennan, J., & Pylkkänen, L. (2008). Processing events: Behavioral and neuromagnetic correlates of aspectual coercion. *Brain and Language*, *106*(2), 132–143. 10.1016/j.bandl.2008.04.003

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). Ms. *Does simple sentence processing tax working memory? Evidence from complex span*.

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

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.

Dowty, D. (1979). *Word meaning and Montague grammar*. Dordrecht: Reidel. DOI: http://dx.doi.org/10.1007/978-94-009-9473-7

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. 10.1016/j.jml.2005.12.006

Filip, H. (2004). The telicity parameter revisited. In *Proceedings of Semantics and Linguistic Theory*, *14*, pp. 92-109.

Folli, R. (2001). Constructing telicity in English and Italian. Doctoral dissertation. University of Oxford, Oxford, England.

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.

Guéron, J. (2008). On the difference between telicity and perfectivity. *Lingua, 118*, 1816–1840. 10.1016/j.lingua.2007.04.005

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*(1), 1–76. 10.1016/S0010-0277(98)00034-1

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

Giorgi, A., & Pianesi, F. (2001). Tense, attitudes and subjects. In Hastings, R., Jackon, B., & Zvolenszky, Z. (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.

Katsika, A, Braze, D., Deo, A., & Piñango, M.M. (2012). Complement coercion: distinguishing between type-shifting and pragmatic inferencing. *The Mental Lexicon, 7,* 58–72. 10.1075/ml.7.1.03kat

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

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*(3), 1126–1177. 10.1016/j.cognition.2007.05.006

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. 10.1016/j.tics.2006.08.007

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

Malaia, E., Wilbur, R. B. and Weber-Fox, C. (2009). ERP evidence for telicity effects on syntactic processing in garden-path sentences. *Brain and Language*, *108*(3), 145-158. 10.1016/J.BANDL.2008.09.003.

McElree, B., Traxler, M., Pickering, M., Seely, R., & Jackendoff, R. (2001). Reading time evidence for enriched composition. *Cognition 78:*B, 17–25. 10.1016/S0010-0277(00)00113-X

O'Bryan, E.L., Folli, R., Harley, H., & Bever, T.G. (2013). Evidence for the use of verb telicity in sentence comprehension. In Folli, R., Sevdali, C, & Truswell, R. (Eds.), *Syntax and its limits*, pp. 80-104. Oxford, UK: Oxford University Press. 10.1093/acprof:oso/9780199683239.003.0005

Philipp, M., Graf, T., Kretzschmar, F., & Primus, B. (2017). Beyond verb meaning: Experimental evidence for incremental processing of semantic roles and event structure. *Frontiers in Psychology, 8*. 10.3389/fpsyg.2017.01806.

Pickering, M., McElree, B., & Traxler, M. (2005). The difficulty of coercion: A response to de Almeida. *Brain and Language*, *93*, 1–9. 10.1016/j.bandl.2004.07.005

Piñango, M. M., & Deo, A. (2016). Reanalyzing the complement coercion effect through a generalized lexical semantics for aspectual verbs. *Journal of Semantics*, *33*, 359–408. 10.1093/jos/ffv003

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.

Pylkkänen, L., & McElree, B. (2006). The syntax-semantics interface: On-line composition of sentence meaning. In Traxler, M., & Gernsbacher, M. A. (eds.). *Handbook of psycholinguistics*, pp. 537–77. New York: Academic Press.

Ramchand, G. (2008). *Verb meaning and the lexicon: A first phase syntax*. Cambridge, UK: Cambridge University Press. 10.1017/CBO9780511486319

Rothstein, S. (2004). *Structuring events*. Oxford: Blackwell.

Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition*, *116*(1), 71–86. 10.1016/j.cognition.2010.04.002

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. 10.1006/jmla.2001.2836

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

Vendler, Z. (1967). Verbs and times. *Linguistics in Philosophy*, 97–121. Ithaca: Cornell University Press. 10.2307/2182371

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