Locality Domains in Syntax: Evidence from Sentence Processing

Stefan Keine

Abstract. One of the main discoveries of generative syntax is that long-distance extraction proceeds in a successive-cyclic manner, in that these dependencies are comprised of a sequence of local extraction steps. This paper provides support for this general picture by presenting parsing evidence for intermediate landing sites created by successive-cyclic movement, building on prior work by Gibson & Warren (2004). It furthermore uses this parsing evidence to investigate the distribution of intermediate gaps. The central findings of this paper are that (i) there is evidence that successive-cyclic movement targets the edge of CPs, and (ii) that there is no comparable evidence for an intermediate landing site at vP edges. These findings are fully consistent with the classical view of successive cyclicity, according to which only finite clause edges host intermediate landing sites. In the context of phase theory, these results receive a straightforward explanation if CPs are phases, but vPs are not. The processing evidence presented here thus provides a novel diagnostic for the distribution of phases and new evidence for their active role in online sentence processing.

Keywords. successive cyclicity; phases; sentence processing; filler–gap dependencies; long A′-movement

1. Introduction

One of the central discoveries about the syntax of natural languages is the locality of the dependencies it creates. While displacement of an element is unbounded in principle, there is considerable evidence that long movement consists of a sequence of smaller movement steps (Chomsky 1973, 1977, et seq.). Any element leaving a finite clause must first land in the highest specifier of this clause (Spec,CP in current terminology). A second movement step then moves this element from this position to its landing site in the next higher clause, as illustrated in (1). If more than one finite clause boundary is crossed, this process applies iteratively. The apparent unboundedness of displacement on this view merely reflects the unbounded number of short movement steps that such sequences may comprise.

(1) [$_{CP}$ Who$_i$ does [$_{TP}$ John think [$_{CP}$ t_i that [$_{TP}$ Mary married t_i in secret]]]]

The view that long movement proceeds successive-cyclically has been widely adopted in the generative tradition and a considerable amount of empirical evidence supporting it has been accumulated, including complementizer shift in Irish (McCloskey 2002), *wh*-copying (McDaniel 1989, Horvath 1997, Bruening 2006), reconstruction to intermediate trace positions (Barss 1986, Lebeaux 1988), quantifier float in West Ulster English (McCloskey 2000), and verb movement in Belfast English (Henry 1995), French (Kayne & Pollock 1978), and Iberian Spanish (Torrego 1984).

The standard minimalist explanation for the existence of successive-cyclic movement comes from the strictly cyclic nature of syntactic derivations in the form of phases. According to phase theory (Chomsky 2000, 2001), syntactic structure is periodically transferred to the interfaces and thereby rendered inaccessible for outside operations. The *Phase Impenetrability Condition* regulates that only the complement of a phase head is rendered inaccessible, whereas material in the specifier of the phase head remains accessible. Movement past a phase head must thus proceed successive-cyclically through the specifier of that phase head in order for the moving element to remain accessible.

Because the concept of a phase does not in and of itself determine which heads are phasal, the distribution of phases has been the subject of considerable interest in the recent literature. Chomsky (2000, 2001) proposes that both C and (transitive) v are phase heads. By the above logic, movement past either of these nodes must be successive-cyclic and both Spec,CP and Spec,vP must be filled with an intermediate trace (an idea that goes back to Chomsky 1986). In other words, it is not only crossclausal movement that is successive-cyclic, but also intra-clausal movement. The view that there are intraclausal landing sites in addition to Spec,CP is widely adopted in the recent syntactic literature, and a number of empirical arguments have been made in its favor (see, e.g., Fox 1999, Legate 2003, Aldridge 2008, Abels 2012, van Urk & Richards 2015, van Urk 2016).

At the same time, this empirical support for vP phases is somewhat weaker than the evidence for CP phases. Chomsky's (2000, 2001) original reasons for treating v and C as phase heads have frequently been called into question (e.g. Bošković 2002, 2014, Epstein & Seely 2006, Boeckx & Grohmann 2007). Moreover, several empirical arguments in favor of a v phase, as well as diagnostics for phases in general, have been disputed (e.g., Matushansky 2005, den Dikken 2006, Dayal 2017).

The present paper contributes to our understanding of successive cyclicity and phases by exploring the distribution of successive cyclicity in real-time sentence processing. Building on the pioneering work of Gibson & Warren (2004), this paper presents the results of three self-paced reading experiments, which provide evidence that the effects of successive-cyclic movement—and hence phases—may be observed in online sentence processing and that such evidence may be fruitfully used to probe its distribution. This opens up a novel and largely uncharted empirical domain that can be brought to bear on long-standing questions in theoretical syntax more broadly.

Experiment 1 provides reading-time evidence for the existence of an intermediate landing site in Spec,CP, precisely what is expected if CP is a phase. Experiments 2 and 3 then compare the processing effects of CPs and vPs that are crossed by movement. The results indicate a systematic difference between the two, which can be immediately accounted for if CPs are targeted by intermediate movement, but vPs are not. By the above logic, this suggests that CP is a phase, but vP is not.

The general link between the grammatical notion of successive cyclicity and evidence from language processing assumed throughout this paper is rather direct: If XP is a phase, it requires successive-cyclic movement through Spec,XP. In order for the parser to establish a grammatically licit syntactic parse, it has to postulate this intermediate trace. Evidence showing whether or not the parser postulates such a trace in the edge of a given domain thus constitutes evidence about whether this domain requires successive-cyclic movement or not.

This paper is structured as follows: Section 2 will discuss previous psycholinguistic investigations into successive cyclicity. It shows that previous evidence that has been argued to support successive cyclicity through Spec,CP is potentially amenable to alternative explanations in terms of active gap filling, a question that motivates Experiment 1 in section 3. Against this background, Experiments 2 and 3 in sections 4 and 5 then investigate whether vPs pattern with CPs in hosting intermediate landing sites, as predicted if both CPs and vPs are phases. The results indicate that this is not the case. I then show that the pattern of results receives a straightforward explanation if CP is a phase but vP is not.

2. Successive cyclicity in online sentence parsing

Despite the immense impact of successive cyclicity on theoretical syntax, there are few direct investigations into the existence of successive cyclicity in sentence comprehension. The only designated attempts to address this question are Frazier & Clifton (1989) and Gibson & Warren (2004), both of which argue in favor of it. However, while the experimental results of these studies are compatible with successive cyclicity, they may also be attributed to unrelated but independently motivated properties. This section will review previous results and discuss their limitations, which will form the basis for Experiment 1, reported in section 3.

2.1 Previous results

The first study that explicitly argues for successive cyclicity in online sentence processing is Frazier & Clifton (1989), who argue that an unassigned filler[1] remains active across clause boundaries and that crossclausal movement dependencies are harder than monoclausal ones. Frazier & Clifton argue that these findings can be explained if (i) an intermediate landing site of the filler is responsible for carrying it across clause boundaries, and (ii) that the construal of the intermediate landing site necessary in crossclausal dependencies increases the demand such dependencies pose to the parser.

While these interpretations are consistent with the observed results, they are not necessary. First, a filler being held active across a clause boundary does not, in and of itself, require the filler to be linked to an intermediate gap at the edge of the lower clause. Second, movement over a clause boundary incurring more difficulty than intraclausal movement may be attributed to a variety of factors unrelated to successive cyclicity. One plausible alternative is that it is the increased distance between the filler and its gap in the two-clause condition that delays the establishment of the movement dependency and thereby increases the difficulty of parsing this structure.

A more recent investigation into processing reflexes of successive-cyclic movement is Gibson & Warren (2004) (henceforth G&W). The rationale underlying G&W's experiment is as follows: There is independent evidence that the distance between the moved element and its gap is positively

[1]What the syntactic literature refers to as 'moved element' (or 'antecedent') and its 'trace' is usually termed the 'filler' and its 'gap', respectively, in the processing literature. These terms will be used interchangeably here.

correlated with the difficulty with which the postulation of the gap and its semantic integration take place. In other words, the greater the distance between the filler and its gap, the harder it is to associate the filler with its gap (e.g., King & Just 1991, Gibson 1998, 2000, Warren & Gibson 2002, Grodner & Gibson 2005, Staub 2010, Bartek et al. 2011).[2] That the distance between the moved element and its gap should affect processing speed at the gap position is commonly attributed to the fact that the semantic and morpho-syntactic features of the moved element have to be retrieved in order to successfully associate it with the gap. The greater the distance to this antecedent, the harder this retrieval process will be. How exactly distance is measured is subject to considerable debate in the literature, but the choice is inconsequential for the discussion here.[3] Against this background, G&W investigate the processing of structures like (2) using self-paced reading.

(2) a. *CP structure*
  (i) *Movement*
  The manager *who*$_i$ the consultant claimed **[**$_{\textbf{CP}}$ $\boxed{t_i}$ that the new proposal had <u>pleased</u> t_i **]** will hire five workers tomorrow.
  (ii) *Control*
  The consultant claimed that the new proposal had <u>pleased</u> the manager who will hire five workers tomorrow.

 b. *DP structure*
  (i) *Movement*
  The manager *who*$_i$ **[**$_{\textbf{DP}}$ the consultant's claim about the new proposal **]** had <u>pleased</u> t_i will hire five workers tomorrow.
  (ii) *Control*
  The consultant's claim about the new proposal had <u>pleased</u> the manager who will hire five workers tomorrow.

In (2a.i) the relative pronoun *who* is moved across a finite CP clause boundary; in (2b.i) the same

[2] It is, for instance, well-known that object relative clauses are harder to process than subject relative clauses (King & Just 1991, Gordon et al. 2001, Traxler et al. 2002). One line of account attributes this contrast to the fact that the distance between the moved element and the gap is greater in object than in subject relative clauses (e.g., Just & Carpenter 1992).

[3] Metrics that have been proposed include the number of intervening words (Hawkins 1994), the number of intervening similar elements (Gordon et al. 2001, Lewis 1996), the number of intervening discourse referents (Gibson 1998, 2000, Warren & Gibson 2002, Grodner & Gibson 2005) and the time elapsed between the processing of the filler and the gap in combination with interference from similar constituent encodings (Lewis & Vasishth 2005, Vasishth & Lewis 2006). These possibilities are not mutually exclusive, of course.

element is moved over a complex DP subject, but this movement does not cross a clause boundary. I will refer to these two structures as *CP structures* and *DP structures*, respectively. The sentences in (ii) constitute the respective control structures, in which the object of the embedded verb *pleased* is not moved. Of interest are the reading times at the verb that hosts the gap of the moved element—*pleased* in (2). To obtain a measure of the difficulty of associating the filler with its gap, G&W compared the reading times at *pleased* in the movement condition in (i) to the reading times in the corresponding non-movement structure in (ii). Because the lexical content and the immediately preceding syntactic context are identical between the two versions, an increase from the non-movement structure to the movement structure must reflect the cost of retrieving the filler and associating it with the gap.

G&W reason as follows: If successive-cyclic movement proceeds through the specifier of each CP that it crosses, there will be an intermediate gap in the CP structure (boxed in (2a.i)). Because no CP is crossed in the DP structure (2b), no intermediate gap exists in this structure. Recall that the difficulty of integrating a gap depends on the distance to its antecedent. If there exists an intermediate gap in the CP condition but not in the DP condition, the distance to the antecedent should be smaller in the CP structure than in the DP structure. This in turn predicts that associating the filler with its ultimate gap should be easier in the CP condition than in the DP condition. On the other hand, if no intermediate gap exists in the CP condition, then integrating the ultimate gap should either be equally hard in the two conditions (if it is the linear distance between the gap and the filler that matters) or the integration should be harder in the CP condition than in the DP condition (if structural distance is the decisive factor). Of interest, then, is whether the integration of the gap is facilitated by a silent intermediate gap in the CP condition.[4]

G&W's results confirm the prediction of successive cyclicity. In addition to an overall reading-time increase in the movement conditions compared to the non-movement conditions, this increase was significantly smaller in the CP structure (2a) than in the DP structure (2b). G&W take this result to support the existence of an intermediate landing site in Spec,CP: The reactivation of the filler at

[4]As Chuck Clifton (p. c.) has pointed out to me, one notable difference between Frazier & Clifton (1989) and Gibson & Warren (2004) is that Frazier & Clifton (1989) reason that the presence of an intermediate landing site makes processing of a structure harder, whereas Gibson & Warren (2004) take it to make processing at the final gap site easier.

the intermediate gap site in the CP structure aids subsequent retrieval of this filler at the ultimate gap site. This effect has subsequently been replicated by Marinis et al. (2005).[5]

2.2 Limitations

While suggestive and entirely novel, G&W's findings are not completely conclusive. As it turns out, their results are amenable to an analysis in terms of independently motivated parsing strategies that are unrelated to successive-cyclic movement. This alternative account takes the following form: A crucial feature of G&W's experimental design is that the movement dependency crosses a verb (*claim* in (2a)) in the CP structure (2a.i), but not in the DP structure (2b.i). One general concern is that the parser might have initially postulated the gap of *who* as the object of the higher verb *claim* in the CP structure. Note that clear evidence that the filler originates from a lower clause becomes available only when the complementizer *that* is encountered. While thus ultimately incorrect, it is possible that the parser construes *who* as the object of the verb *claim* in its initial parse of this region. It is then furthermore possible that this intermediate, though incorrect, reactivation of *who* renders it more salient in memory and thereby facilitates its retrieval further downstream when its actual gap position is encountered. This facilitation would then be reflected as a smaller reading-time increase compared to the DP condition, where no such erroneous intermediate reactivation takes place.

On this alternative account, the effect observed by G&W would be entirely due to the particular mechanisms and decision procedures underlying the parser, and have nothing to do with successive-cyclic movement. Rather, this effect would merely be due to incorrect structural analyses temporarily pursued by the parser on the construction of a subsequent, and correct, representation. I will henceforth refer to this account as the *premature gap filling account,* as it crucially involves the premature postulation of a gap that later turns out to be incorrect.

What makes this account a viable alternative to successive cyclicity is that the pieces it comprises are independently motivated in the literature on sentence parsing. First, it is well-known that the parser pursues an *Active Filler Strategy* when scanning an input string for a gap position:

[5]Marinis et al. (2005) also tested L2 speakers of English (L1: Chinese, Japanese, German or Greek) in addition to native speakers. Interestingly, they replicated G&W's crucial effect for L1 speakers, but not for L2 speakers. Marinis et al. (2005) argue from this result that L2 speakers underuse syntactic information in online sentence processing.

(3) *Active Filler Strategy* (Frazier & Clifton 1989: 95)
When a filler has been identified, rank the option of assigning a gap above all other options.

The active filler strategy states that the parser is extremely eager to terminate an open movement dependency by postulating a gap at the earliest grammatically licit position. Crucially, it does so even in the absence of evidence from the input that a gap is present. In some cases, the parser may hence postulate a gap prematurely, i.e., in a position that will subsequently turn out to be filled by a lexical element, giving rise to 'filled gap effects', first observed by Stowe (1986). In addition to filled gap effects, a variety of other experimental paradigms have yielded support for the active filler strategy.[6]

With regard to the structures in (2), these findings are fully in line with the alternative account just sketched. Because the movement dependency crosses a higher verb (e.g, *claim*) in the CP structure, it is conceivable that the parser initially postulates the gap of *who* as the object of this higher verb. This gap would then induce reactivation of *who* and thereby facilitate its subsequent retrieval at the actual gap site. By contrast, in the DP condition, the complex subject is an island. In light of evidence that the parser does not postulate gaps in islands (Stowe 1986, Traxler & Pickering 1996, McElree & Griffith 1998), it is expected that the parser does not posit a premature gap in the complex subject in the DP structure and hence that no intermediate filler reactivation takes place.

G&W briefly address this concern (p. 64), noting that the verbs they used were strongly biased towards a clausal complement and that, to the extent that these verbs were compatible with a nominal object, they required their object to be inanimate. As the filler in all of their experimental items was animate, G&W conclude that these properties provide sufficient cues to prevent the parser from temporarily postulating a gap of the moved element at the higher verb.

Unfortunately, it is not clear at present to what extent the Active Filler Strategy can be mitigated by factors such as frequency, animacy or real-world plausibility. There is evidence that the frequency of a subcategorization frame does not affect whether a gap is postulated as the argument of

[6] Traxler & Pickering (1996), Phillips (2006), Staub (2007) and Wagers & Phillips (2009), for instance, find reading-time increases if a verb semantically mismatches an unassigned filler even if that filler is not syntactically related to the verb in the ultimately correct parse of the input string. There is moreover converging evidence from the stops-making-sense task (Boland et al. 1995), cross-modal lexical priming (Nicol & Swinney 1989, Nicol et al. 1994) and anticipatory eye movement (Sussman & Sedivy 2003).

an incoming verb (Frazier & Clifton 1989, Staub 2007) or whether a gap is postulated for an optionally transitive verb whose argument could in principle have been extraposed (Staub et al. 2006). Furthermore, Pickering & Traxler (2003) found that animacy and plausibility do not prevent the parser from construing an otherwise licit gap. In light of these findings, it is entirely possible that the parser postulates a gap of the moved element when it encounters the higher verb even if frequency and plausibility disfavor such a decision.[7] While G&W's results are thus fully compatible with the interpretation that the facilitation at the ultimate gap site in the CP condition is due to successive-cyclic movement, their results could potentially also be attributed to premature gap filling at the intermediate verb. Experiment 1 seeks to probe these two lines of explanation by investigating whether the effect is modulated by properties of the higher verb.

3. Experiment 1

In order to gain a better understanding of the relationship between G&W's effect and the properties of the higher verb, Experiment 1 manipulates the subcategorization frame of the higher verb. Whether subcategorization information modulates active gap filling is controversial. While Staub (2007) presents evidence that the parser does not postulate an object gap with obligatorily intransitive verbs, work by Omaki et al. (2015) indicates that active gap filling is not suspended even if the verb's subcategorization information is incompatible with a DP object (also see Phillips et al. 2005: 426). In Omaki et al.'s (2015) terminology, active gap filling on this view is *hyper-active*.

Let us consider the expectations that these various analytical options give rise to with respect to the relationship between the effect at the ultimate gap site and the properties of the embedding verb.[8] If active gap filling is suspended with intransitive verbs (Staub 2007), the filler should not be reactivated if the embedding verb's subcategorization frame is incompatible with a DP object. If the facilitation at the gap site is exclusively due to premature gap filling at the embedding verb, this

[7]In addition to the general concerns just mentioned, the particular properties of G&W's stimuli also deserve some remarks. First, seven of the nine verbs used by G&W are similar to *claim* in that they productively take nominal objects (*predict, claim, conclude, imply, confirm, realize, state*). Of these seven verbs, at least four appear compatible with animate objects, as a cursory search on Google reveals: *predict* ('Jesus predicted the prophet Mohammed'), *claim* ('The queen claimed the slaves as her own property'), *confirm* ('The Senate confirmed Robert Hanna as Superior Court judge') and *realize* ('He created a trade that reached to all parts of the Union, and realized him a large fortune' [NY Times]).

[8]I am indebted to two reviewers for very helpful comments and suggestions.

view predicts that no facilitation at the actual gap site should be observed with embedding verbs that are incompatible with a DP object. By contrast, if gap filling is hyper-active and applies irrespective of the subcategorization frame of the higher verb (Omaki et al. 2015), the predictions are more nuanced. While premature gap filling is then expected to take place regardless of the type of the higher verb, verb type might still modulate the facilitation downstream. First, thematic integration of the filler with the higher verb is possible only with verbs that are able to host DP objects. Second, the information that shows that the premature gap parse is incorrect is available earlier if the embedding verb is incompatible with a DP object. In this case, this initial parse can be abandoned as soon as the subcategorization information of the embedding verb becomes available—i.e., at the embedding verb itself. By contrast, if the embedding verb is in principle compatible with a DP object, then the information that shows that this parse is incorrect is available only later, at the complementizer, presumably after thematic integration into the higher verb has taken place or been attempted.[9] It is then possible that the filler receives a greater reactivation boost from verbs that can take a DP object than from verbs that cannot, and this might attenuate the facilitation downstream.

By contrast, if G&W's effect is the result of successive cyclicity, then filler reactivation should be independent of the higher verb and it should obtain regardless of its subcategorization properties. Experiment 1 investigates the relationship between the higher verb and the facilitation effect.

3.1 Method

Materials: Fifty-six sets of sentences like the one in (4) were constructed, incorporating materials adapted from Gibson & Warren (2004) and Marinis et al. (2005). Due to the fact that the semantic relations in the sentences differ by condition, a plausibility norming study was carried out to ensure that the eight conditions matched each other in plausibility. The details of this norming study are described in Appendix A in the supplementary materials. Forty-eight sets of sentences with closely matching overall plausibility ratings were selected and the remaining eight discarded.

The experiment manipulated INTERVENER TYPE (*CP* vs. *DP*) and MOVEMENT ([+*move*] vs.

[9] Evidence for rapid semantic integration comes from plausibility mismatches between a filler and a potentially gap-hosting verb, which is typically observed at the verb itself in online reading measures (Traxler & Pickering 1996, Pickering & Traxler 2003, Phillips 2006).

[–*move*]) in a way parallel to G&W's original study. In addition, the TYPE OF THE VERB preceding the complementizer was manipulated. One set of verbs productively took DP direct objects (e.g., *claim*). These verbs will be referred to as *CP/DP-verbs* as their subcategorization allows for either a CP or a DP object. The second class of verbs was incompatible with a DP object and only allowed a clausal object (e.g., *boast*). This second class will be referred to as *CP-verbs.* Membership in the two classes was determined by the sentence completion results reported in Trueswell et al. (1993) and a subcategorization database developed by Sabine Schulte im Walde, based on the British National Corpus and containing data for over 3000 verbs (Schulte im Walde 1998).[10] Verb type was crossed with the other two factors, yielding a total of eight conditions. Because of the relative difficulty of the stimuli, the target sentences were preceded by a theme-setting context sentence. This context sentence gave lexical information about the lower verb in the relative clause in the target sentences (e.g., *hurt* in (4)). The rationale for including this context was to decrease the processing load in the critical region. Informal consultation of native speakers of American English confirmed that this context sentence reduced the perceived difficulty of the target stimuli. All target sentences were followed by a comprehension question about the semantic relations in the sentence the participant had just seen. No question targeted the context sentence. All questions were multiple choice, with both answers being referents present in the sentence. Answers were presented in random order.

(4) *Context:* Groundless allegations really could hurt people in our company.

a. ***CP structure***

(i) **[+*move*]:** The secretary *who*$_i$ the lawyer $\left\{ \begin{array}{ll} \textit{CP/DP:} & \text{claimed} \\ \textit{CP:} & \text{boasted} \end{array} \right\}$ **[$_{CP}$** that the accusa-

[10]The 15 CP/DP-verbs used in the stimuli are *assert, assume, claim, conclude, confirm, decide, declare, demonstrate, guess, illustrate, imply, predict, prove, recall* and *state.* These verbs had an average DP object rate of .21 in the British National Corpus. The 14 CP-verbs used are *agree, argue, boast, comment, dream, hint, hope, hypothesize, insist, pretend, remark, speculate, theorize* and *think.* Of these, *agree, boast, dream, hint, hope, insist, pretend, remark* and *think* were selected because their DP object rate in Trueswell et al.'s (1993) sentence completion experiment was 0%. Five additional verbs (*argue, comment, hypothesize, speculate, theorize*) were included to decrease the amount of repetition in the stimuli. Overall, these CP-verbs had an average DP object rate of .09 in the British National Corpus. It should be noted that this number is likely inflated because two verbs (*boast* and *theorize*) had a very high DP object rate (.41 and .30, respectively), due to the highly literary character of some of the materials in this corpus. That *boast* did not elicit a single DP object completion in Trueswell et al. (1993) makes it very unlikely that these usages of these two verbs have a significant impact on parsing decisions in the population of interest. To confirm this conclusion, all analyses reported here were also conducted with all items containing these two verb eliminated. The critical effects remained unchanged in these analyses. With these two verbs excluded, the average DP completion rate of the CP-verbs in the Corpus was .05.

tions had hurt t_i **]** was fired from her job.

(ii) **[–*move*]:** The lawyer {claimed/boasted} that the accusations had hurt the secretary who was fired from her job.

b. ***DP structure***

(i) **[+*move*]:** The secretary *who*$_i$ **[**$_{\mathbf{DP}}$ the lawyer's $\left\{\begin{array}{ll} \textit{CP/DP:} & \text{claim} \\ \textit{CP:} & \text{boast} \end{array}\right\}$ about the accusations **]** had hurt t_i was fired from her job.

(ii) **[–*move*]:** The lawyer's {claim/boast} about the accusations had hurt the secretary who was fired from her job.

Comprehension question: Who made a {claim/boast}? *the lawyer – the secretary*

The difference between CP-verbs and CP/DP-verbs hence only affects the subcategorization properties of the higher verb, not the type of structure that this verb occurs in. In addition to the 48 target items, 48 additional items were created as fillers, which matched the target sentences in syntactic complexity and length. 24 of these were part of another experiment and the additional 24 were a haphazard collection of sentences.

Participants: 130 native speakers of American English were recruited via Amazon Mechanical Turk. All were naïve to the purpose of the experiment. Each received a compensation of USD 1.

Procedure: The experiment was conducted using the online experiment platform Ibex (Drummond 2013) and employed a region-by-region self-paced noncumulative moving-window task (Just et al. 1982).[11] The regioning of the target sentences followed the general schema in Table 1.[12] At the beginning of each trial participants saw the theme-setting context sentence, which was displayed

[11]One might wonder about the reliability of reading data gathered online. Use of online platforms for this type of study is becoming increasingly mainstream among researchers (see Wagers & Phillips 2013 for a recent example). There is so far no indication that this methodology produces results that are qualitatively different from data elicited in a more traditional lab setting (e.g., Wagers & Phillips 2013, Enochson & Culbertson 2015). Moreover, with regard to the three experiments reported here, it is noteworthy that all three extend the experimental design used by Gibson & Warren (2004) and Marinis et al. (2005) in a traditional lab setting. A measure of the reliability of the online data reported here is thus whether they replicate the key effects observed in these two studies. As emphasized in the discussion of the respective results, all three experiments here do replicate these previous findings. It is therefore very unlikely that the effects reported here are merely an artifact of the experimental methodology.

[12]This regioning is identical to the one employed by G&W in their analysis with one difference. In G&W's materials, relative pronouns were grouped inconsistently. In [+move] conditions the pronoun was grouped with the head noun phrase (*the secretary* in Table 1), while it was grouped with the remainder of the relative clause in the [–move] condition. This inconsistency was irrelevant for G&W as they used a word-by-word presentation and the words were grouped into regions only for the purposes of the analysis. Since, by contrast, the segmentation affects the units of presentation in the present experiment, a consistent regioning was employed and relative pronouns were presented with the head nominal throughout.

Table 1. Regioning of stimuli in Experiment 1 ([±mv] = [±move])

			Region						
			1	2	3	4	5	6	7
CP-verb	CP	[–mv]		The lawyer boasted	that	the accusations	had hurt	the secretary who	was fired from her job
		[+mv]	The secretary who	the lawyer boasted	that	the accusations	had hurt	was fired	from her job
	DP	[–mv]		The lawyer's boast	about	the accusations	had hurt	the secretary who	was fired from her job
		[+mv]	The secretary who	the lawyer's boast	about	the accusations	had hurt	was fired	from her job
CP/DP-verb	CP	[+mv]	The secretary who	the lawyer claimed	that	the accusations	had hurt	was fired	from her job
		[–mv]		The lawyer claimed	that	the accusations	had hurt	the secretary who	was fired from her job
	DP	[+mv]	The secretary who	the lawyer's claim	about	the accusations	had hurt	was fired	from her job
		[–mv]		The lawyer's claim	about	the accusations	had hurt	the secretary who	was fired from her job

in its entirety. Upon pressing the space bar, the context sentence was replaced by dashes masking the regions of the target sentences. Pressing the space bar caused the dashes in the first region to be replaced by the actual content of the region. When the space bar was pressed again this region reverted back to dashes and the next region appeared. Participants traversed through the entire sentence by repeatedly pressing the space bar. Pressing the space bar after the final region had been displayed caused the dashes to disappear and a comprehension question accompanied by two possible answers appeared on the screen. Participants selected the answer on the left by pressing the 'f' key and the one on the right by pressing the 'j' key. No feedback on answer accuracy was given. After every twelve trials the participant had to take a ten-second break and could rest for longer if they so desired. Altogether, the experiment contained seven of these mandatory breaks. The experimental trials were preceded by a screen collecting general demographic data, three screens of instructions and six practice trials. The experiment took about 40 minutes. Throughout the entire experiment a progress bar was displayed. The experiment used a Latin square design: items were arranged on eight different lists such that each list contained one instance of every item and each of the eight conditions of every item appeared on a different list. Each participants was assigned to one of the eight lists. The 48 target sentences were interspersed with the 48 fillers and the order of presentation was randomized for each participant.

Analysis: Unless noted otherwise, the data analysis procedures were identical for the experiments reported here and are discussed here in the context of Experiment 1. All data analysis was carried out in the R software environment (R Core Team 2014). All reading times were logarithmically transformed and then entered into linear mixed effects (LME) models; answer accuracy was analyzed by means of logistic mixed effects modeling. Models were fitted using the lme4 package (Bates et al. 2014). All models were maximal in the sense that they incorporated random intercepts for participants and items and random slopes for all fixed effects and their interactions for both participants and items, following the recommendations in Barr et al. (2013). The estimate of the regression slope β and the corresponding t/z-statistics will be reported. To obtain p-values for the model coefficients, the Satterthwaite approximation to the degrees of freedom associated with a coefficient's t-value was employed, using the lmerTest package (Kuznetsova et al. 2014). All contrasts used in the analyses were orthogonal and will be specified for each experiment individually.

The following general exclusion procedure was applied: Participants who indicated that their native language was not American English were excluded from analysis. If a participant took the experiment multiple times, only the first time was included in the data analysis. Because of the difficulty of the materials and the subtlety of the effect of interest, only particants whose overall accuracy over the entire experiment (including all target and filler items) was at least 75% were included in the data analysis. For the statistical analyses of the results without accuracy-based exclusions, see Appendix C. Reading-time data less then 200 ms or greater than 5000 ms were taken to not reflect the process of interest and hence rejected as outliers. Finally, log reading times that were more than 3 standard deviations away from the condition mean in that region were discarded.

To adjust for difference in participants' reading speed and the substantial differences in the length of the various regions, residual reading times were calculated. Using all target and filler items used in the experiment, a mixed-effects model was fitted that predicted raw reading times from the number of characters in a region and included random intercepts and slopes for participants (see Ferreira & Clifton 1986 for discussion of this procedure). At every region the reading time predicted by the regression was subtracted from the actual measured reading time. This difference constitutes the residual reading time. A positive value hence indicates slower reading times than were predicted by the model while negative values predict faster reading times. Residual reading times were submitted to mixed-effects model analyses only for regions whose length differed across conditions.

3.2 Results

In order to streamline the presentation and discussion of the results, the main text will focus attention on the critical regions of the relevant experiment. A comprehensive overview of the analyses of all regions and a discussion thereof is provided in Appendix B of the supplementary materials.

Of the 130 participants who took part in Experiment 1, 6 were excluded for taking the experiment twice. An additional 30 were discarded for falling below the accuracy threshold.[13] Outlier elimination based on the absolute thresholds excluded less than 0.4% of the data. The z-score-based rejection of outliers resulted in the exclusion of 0.8% of the observations. For the analysis of the reading time data the results were subjected to a 2×2×2-factorial LME model crossing the factor MOVEMENT ([+*move*] vs. [–*move*]), INTERVENER (*CP* vs. *DP*) and VERB TYPE (*CP-verbs* vs. *CP/DP-verbs*). Accuracy data were analyzed using a parallel logistic mixed effects model. All covariates were coded numerically using sum-coding (MOVEMENT: [–*move*] = −.5, [+*move*] = .5; INTERVENER: *CP* = −.5, *DP* = .5; VERB TYPE: *CP-verb* = −.5, *CP/DP-verb* = .5). Because one item was regioned inconsistently across conditions, it was excluded from all analyses.[14]

Comprehension question response accuracy: The mean answer accuracy over all test items and conditions was 85.3%. The proportion of correct responses by condition is given in Table 2. Logistic LME modeling revealed a main effect of *movement* such that accuracy in the [+move] condition was lower than in the [–move] condition ($\hat{\beta} = -1.25, z = -11, p < .001$). No other effect

[13]The critical statistical results reported below also hold if no accuracy-based exclusion of participants is applied (see Appendix C).

[14]Eliminating this item did not introduce a plausibility confound (all ps > .5, cf. Appendix A).

Table 2. Mean answer accuracy computed over participants in Experiment 1

	CP-verbs		CP/DP-verbs	
	CP	DP	CP	DP
[–move]	.92	.93	.92	.91
[+move]	.78	.76	.83	.78

was significant.

Reading times: The mean reading times by region and condition computed across participants as well as residual reading times for region which differed in length between conditions are given in Table 3. An overview of the by-condition standard error as a measure of the response variability is provided in Appendix B. Table 4 provides an overview of the LME analyses of the log-transformed reading times for the crucial regions. Of central importance are the verb+gap region (region 5) and the spillover region (region 6). In addition, because the manipulation of the type of the higher verb should affect whether the parser prematurely posits a gap or not, the reading times in the region following this verb (region 3) were analyzed. In the interest of space, Table 4 only reports the analyses for these three regions. A comprehensive analysis of all regions is provided in Appendix B.

Analysis of region 3 reveals main effects of *movement* and *intervener.* In addition, there was a significant three-way interaction between all factors. Because the length of the region differed between conditions, an additional analysis of the residual reading times was performed, which replicated the two main effects ($p < .001$) and the three-way interaction ($\hat{\beta} = -65, t = 2.7, p < .01$). In addition,

Table 3. Mean raw reading times by condition and region in ms computed over participants for Experiment 1. For regions that differed in length between conditions, residual reading times are additionally provided in parentheses. ('Vt' = Verb type, 'Intv' = Intervener, 'Mvmt' = Movement, [±mv] = [±move])

			Region						
Vt	Intv	Mvmt	1	2	3	4	5	6	7
CP	CP	[–mv]	—	911	536 (–37)	775	708	812 (–48)	1288 (68)
		[+mv]	731	1032	598 (25)	797	836	801 (59)	868 (–64)
	DP	[–mv]	—	964	561 (–29)	824	761	796 (–62)	1213 (–10)
		[+mv]	735	1216	679 (89)	921	860	863 (120)	866 (–64)
CP/DP	CP	[–mv]	—	896	526 (–48)	765	738	840 (–19)	1279 (58)
		[+mv]	736	1066	631 (57)	790	893	820 (79)	888 (–44)
	DP	[–mv]	—	1018	562 (–20)	824	736	815 (–47)	1219 (–2)
		[+mv]	710	1215	657 (74)	892	903	845 (102)	887 (–43)
					complementizer/ preposition		*verb + gap*	*spillover*	

Table 4. Coefficient estimates and corresponding t-value for linear mixed effects model analyses of log reading times in critical regions of Experiment 1. *Mvmt:Intv* refers to the interaction between *Movement* and *Intervener, Mvmt:Vt* to the interaction between *Movement* and *Verb type* and *Intv:Vt* refers to the interaction between *Intervener* and *Verb type*. *Mvmt:Intv:Vt* refers to the three-way interaction of all predictors. Cells with $p < .05$ are shaded. For statistical analyses of all regions, see Appendix B.

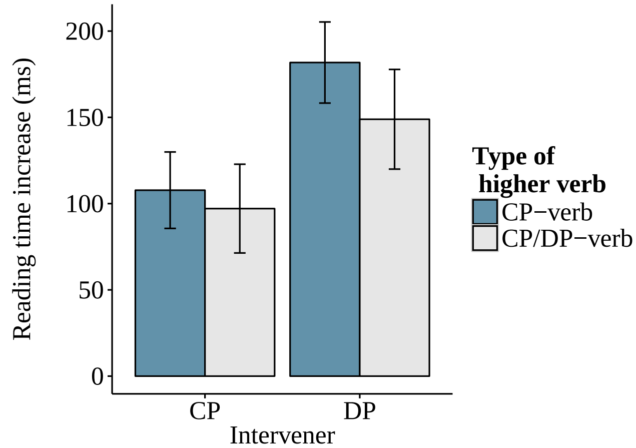
	Region 3		Region 5		Region 6	
	$\hat{\beta}$	t	$\hat{\beta}$	t	$\hat{\beta}$	t
Movement	0.127	11.33	0.112	5.03	–0.009	–0.38
Intervener	0.063	6.23	0.031	2.66	0.015	1.17
Verb type	0.000	0.03	0.023	1.99	0.012	1.11
Mvmt:Intv	0.029	1.78	–0.004	–0.17	0.060	2.55
Mvmt:Vt	0.013	0.75	0.018	0.80	–0.026	–1.27
Intv:Vt	–0.024	–1.47	–0.049	–2.37	–0.017	–0.79
Mvmt:Intv:Vt	–0.075	–2.31	0.031	0.70	–0.018	–0.43

this analysis indicated an interaction between *movement* and *intervener* ($\hat{\beta} = 25, t = 2.0, p < .05$), which is, however, uninterpretable given the higher-level interaction.

To further investigate the three-way interaction in region 3, a second LME model was fitted to the data, which nested the predictors *movement* and *verb type* under the levels of *intervener*, with the full random effects structure of the original model (including the three-way interaction) being preserved. The rationale for doing so is the question whether CP and DP structures differ in whether or not the reading-time increase induced by movement is modulated by the type of the verb (in the CP structure) or noun (in the DP structure). For DP structures, the model detected a main effect of *movement* ($\hat{\beta} = .14, t = 10, p < .001$) but neither a main effect of *verb type* ($\hat{\beta} = -.01, t = -1.0, p = .3$) nor an interaction between the two ($\hat{\beta} = -.02, t = -1.1, p = .3$). For the CP structure, on the other hand, the model revealed a main effect of *movement* ($\hat{\beta} = .11, t = 8, p < .001$), no main effect of *verb type* ($\hat{\beta} = .01, t = 1.0, p = .3$), but an interaction between both factors ($\hat{\beta} = .05, t = 2.1, p < .05$) such that the reading-time increase was higher for CP/DP-verbs than for CP-verbs. The type of the verb/noun in region 2 thus affects the reading-times increase in region 3 in the CP structures but not the DP structures. The same pattern emerged in the analogous analysis of the residual reading times. It will be considered in greater detail in the discussion section.

In the gap region (region 5), the DP structure was read more slowly than the CP structure and the [+move] conditions were read more slowly than [–move]. There is, in addition, a main effect of *verb type* such that CP/DP-verb conditions were read more slowly than CP-verb conditions. Finally, there

Figure 1. Mean increase in residual reading times from [–move] to [+move] conditions computed over participants and by-participant standard errors in region 6 of Experiment 1 by intervener and verb type



was an interaction between *verb type* and *intervener* such that the reading-time difference between CP and DP structures was smaller for CP/DP-verbs than for CP-verbs. There was no interaction between *movement* and *intervener*.

The spillover region (region 6) showed a significant interaction between *intervener* and *movement* such that the DP [+move] structure led to higher reading times relative to its [–move] control than the CP structure. There was no interaction with verb type. Because the length of this region differed between [+move] and [–move] conditions and to confirm the reliability of these critical findings, residual reading times were analyzed. This model showed a main effect of *movement* ($\hat{\beta} = 132, t = 7, p < .001$) and an interaction between *movement* and *intervener* ($\hat{\beta} = 62, t = 2.4, p < .05$), but no three-way interaction ($\hat{\beta} = -26, t = -.6, p > .5$). That the main effect of movement was restricted to the residual reading times is unsurprising because region 6 was systematically longer in the [–move] than in the [+move] condition. Higher proportional reading times in the movement condition hence are only detectable when the length of the region is factored out.

Figure 1 plots the crucial increase in the residual reading times from [–move] to [+move] structures in the spillover region. The analysis just presented reveals the following key components: (i) There is a positive reading time increase incurred by the presence of a movement dependency in all conditions, (ii) this increase is greater in DP structures than in CP structures, and (iii) the pattern of this increase does not differ across the two verb types.

3.3 Discussion

The results of Experiment 1 replicate G&W's crucial finding, both in the analysis of the log-transformed and of the residual response times: The difference between [–move] and [+move] conditions was greater in DP structures than in CP structures when the position of the gap was encountered. Filler retrieval was hence more difficult in the DP structures than in the CP structure. While the region that the effect surfaced in differed (region 6 in the present experiment but region 5 in Gibson & Warren 2004 and Marinis et al. 2005), the shape of the effect is identical.[15] Because it is common in self-paced reading experiments for an effect to surface in the spillover region, this discrepancy is not surprising. It is also worth noting that the lexical material in region 6 differed between the [+move] and [–move] conditions.[16] While this difference in length clearly affects the results for the main effect of *movement,* it is orthogonal to the reading-time increase between [–move] and [+move] in these structures, because the lexical change between [–move] and [+move] conditions is identical across intervener types and verb types. As a result, the length difference does not plausibly affect the comparison of the critical reading-time increase between [–move] and [+move] across structures. Also see Experiment 3 below for a more direct assessment.

The central finding of Experiment 1 is that the increased difficulty of retrieving the filler in the DP structure relative to the CP structure is not modulated by the type of the higher verb (in both the log-transformed and the residualized reading times). This result is consistent with the successive cyclicity account, hence G&W's interpretation of the effect: If the facilitation in the CP structure is the result of an intermediate gap at the edge of the embedded clause, the lack of interaction with the matrix predicate is accounted for. The experiment did not find support for the alternative view that the facilitation in the CP structure is exclusively due to premature gap filling at the higher verb,

[15]A reviewer raises the possibility that the CP/DP-contrast does not reflect retrieval difficulty per se, but rather temporal differences in filler retrieval, e.g., that the filler is integrated later in the DP condition (not at the verb itself) than in the CP condition. It is difficult to assess this possibility within the confines of the results of Experiment 1, but it should be noted that an account in terms of temporal delay is not incompatible with the conclusions drawn here: It would still require an account of why filler retrieval differs in the two structures, and successive cyclicity in the CP condition provides one. Note also that there is no indication in the evidence here that the retrieval penalty in the DP structure in the spillover region is offset by a relative facilitation in another region, as one might expect if it is the temporal distribution of the retrieval difficulty that differs across structures. A more comprehensive evaluation of this analytical possibility might hence require reading-time measures with a higher temporal resolution.

[16]I am grateful to two reviewers for very helpful comments about this potential confound.

though the results are compatible with the existence of hyper-active gap filling at the higher verb.

The three-way interaction observed in region 3 (the region containing the complementizer/preposition) is particularly instructive. The pattern of this effect is that the reading-time increase between [–move] and [+move] conditions is sensitive to the verb type in the preceding region, but only in the CP condition. More specifically, the reading-time increase is greater for CP/DP-verbs like *claim* than for CP-verbs like *boast.* In the DP condition, on the other hand, this increase is not affected by the type of the the preceding noun. This pattern is plausibly a filled gap effect: If the verb is the CP/DP-kind, the moved element is initially construed as its object and the parser incorrectly postulates a gap of the moved element when the higher verb is encountered. The complementizer in the following region makes it clear that this parse is incorrect and that reanalysis is required. This reanalysis amplifies the reading-time increase compared to the non-movement baseline. The fact that this increase is less pronounced if the higher verb is the CP-type (i.e., incompatible with a DP object) can be interpreted in at least two ways. First, it might indicate that no incorrect gap is postulated in this case (i.e., if subcategorization information prevents active gap filling, see Staub 2007). Alternatively, it might show that a gap is postulated in both cases (Omaki et al. 2015), but that reanalysis is faster if the verb is of the CP-type than if it is of the CP/DP-type, plausibly because the verb's subcategorization information rules out such a parse in the case of CP-verbs. Importantly, no analogous difference was observed in the DP condition. Here, no gap can be postulated in either region 2 or 3 regardless of whether the head noun is *claim* or *boast.* As such, no filled gap effect arises regardless of the type of the head noun. This accounts for the observed three-way interaction in region 3. This finding has important repercussions: It demonstrates that the two groups of verbs (*CP/DP* vs. *CP*) indeed differ in their subcategorization frames and that premature gap filling or reanalysis is affected by these frames. Significantly, the critical effect in region 6 is independent of these subcategorization frames. In fact, the advantage of the CP condition relative to the DP condition at the ultimate gap was numerically, but not significantly, greater for CP-verbs than for CP/DP-verbs. This provides further indication that the crucial effect at the gap-hosting verb is not reducible to premature gap filling at the higher verb. Note again that these results do not constitute evidence against hyper-active gap filling at the higher verb as such, and in fact the two accounts are not mutually exclusive. Experiments 2 and 3 will provide additional evidence that the facilitation effect is not plausibly solely due to premature gap filling.

Furthermore, there was an interaction between *intervener* and *verb type* in the gap region such that the reading-time difference between CP and the DP structures is smaller for CP/DP-verbs than for CP-verbs. Because this pattern does not involve the reading-time increase between [–move] and [+move] structures, it is orthogonal to the crucial effects just discussed.[17]

The gap region also showed main effects of *movement* and *intervener.* There is indication that they are not related to gap formation and filler retrieval, and hence independent of the crucial domain of interest here. As shown in Appendix B.1, these main effects are observed throughout regions 2 to 5. This timecourse suggests that the main effect of *movement* reflects the general cost of retaining a filler in working memory (Wanner & Maratsos 1978, Gibson 1998, 2000, Fiebach et al. 2002, Grodner et al. 2002, Chen et al. 2005), rather than the cost of retrieving a filler per se. The main effect of *intervener* plausibly reflects the processing cost of the complex subject in the DP condition, irrespective of the presence of movement. It is plausible, therefore, that the reason that the main effects arise one region earlier than the critical interaction is because these main effects reflect processing costs incurred before the gap site.[18]

Before proceeding, it is worth considering the role of *structural distance* in the results of Experiment 1. Taking into account structural distance does not affect the interpretation of the present results. First, suppose that no successive-cyclic movement took place in the CP condition. Because movement is crossclausal in the CP condition but intra-clausal in the DP condition, the structural distance between the filler and the gap, and hence the reading-time increase, would be predicted to be larger in the CP condition, the opposite of the observed pattern. Appeal to successive cyclicity is

[17]It is possible that this effect is a spillover from the preceding region and can be accounted for with reference to the definitional properties of the two verb classes: While CP/DP-verbs allow for a DP object, CP-verbs do not and it is natural to assume that this property of the verb is preserved under nominalization. In the DP structures, the noun in region 4 has to be conceptually integrated with the head noun (*claim*/*boast*). Due to subcategorization properties, this integration is likely easier for nouns based on CP/DP-verbs than for noun based on CP-verbs. In the CP condition, by contrast, the noun in region 4 is not an argument of the higher verb, and verb type should hence not matter. As a result, the reading-time increase from the CP to the DP condition is greater for CP-verbs than for CP/DP-verbs.

[18]I would like to thank a reviewer for raising this question.

hence necessary even if structural distance is factored in. Moreover, if successive-cyclic movement through Spec,CP takes place, the ultimate gap is equally far away from the closest antecedent position in both conditions. Structural distance hence does not affect the conclusion that successive cyclicity takes place in the CP condition but not the DP condition.[19]

4. Experiment 2

G&W's results and those of Experiment 1 provide evidence for intermediate filler reactivation when movement crosses a CP, and this reactivation seems to be independent of the properties of the embedding verb. By the above reasoning, successive cyclicity through Spec,CP provides an explanation for this finding. This aligns well with traditional syntactic evidence, a welcome result. Experiment 2 uses processing evidence to ask whether intermediate gaps are limited to Spec,CP (the traditional view) or whether they also occur in Spec,vP (the standard phase-based view). I will refer to the former view as the *CP Hypothesis* and to the latter position as the *CP+vP Hypothesis*.

(5) a. *CP Hypothesis*
Only CP is a phase. Consequently, intermediate landing sites are created only in Spec,CP.

[19]It is noteworthy that there is a second parsing-based account of the facilitation effect in the CP condition, not considered here so far. According to this account, the disproportional difficulty in the DP condition is due to a super-additive interaction between the presence of a complex subject and the existence of a movement dependency across it, the combination of which creates a computational bottleneck. Under this view, the crucial effect does not reflect particularly fast reading times in the CP condition but unexpectedly long reading times in the DP condition. In order to evaluate this account, a post hoc test was performed. Because, by hypothesis, the interaction is the super-additive combination of the relative complexity of the DP structure and the movement dependency, the size of the critical effect at region 6 should positively correlate with the complexity of the DP structure relative to the CP structure in the absence of movement. In other words, for each item, the greater the relative difficulty of the DP structure in the absence of extraction the greater the size of the interaction at region 6 is predicted to be, assuming that the complexity contributed by the movement dependency is roughly equal for all items. To evaluate this prediction, the residual reading times per structure type for regions 2 though 4 for the no-move conditions of each item were summed up. This yields a measure of the respective difficulty of the CP and DP structures in the regions preceding the critical verb in the absence of a movement dependency. To arrive at the difficulty of the DP condition relative to the CP condition, the difference of the two sums was calculated for each item. In addition, to estimate the size of the critical interaction between extraction and intervener type at region 6, the increase in log reading times from the non-extraction to the extraction condition for CP structures was subtracted from the same difference of the DP structures for each item. According to the alternative account outlined above, the size of the critical effect in region 6 should be positively correlated with the relative difficulty of the DP condition in regions 2 through 4. To assess this prediction, a linear model was devised that regressed the structural difficulty scores against the interaction scores. The model did not revealed any reliable relationship between the two measures ($r = -0.07, \textit{adjusted } R^2 = -0.02, \mathrm{F}(1,45) = 0.22, p > .5$). It is worth noting that the numerical trend was in the direction opposite of what the structural complexity account predicts. The results of Experiment 1 therefore do not support the view that the G&W effect is simply a non-cumulative combination of the two main effects, though as a reviewer points out it might be the case that integration difficulties incurred by complex subjects surface at the verb (Van Dyke & Lewis 2003).

b. *CP+vP Hypothesis*
CP as well as vP are phases. Consequently, intermediate landing sites are created in Spec,CP and Spec,vP.

Returning to the structures investigated by G&W and in Experiment 1, the two hypotheses predict a different number of intermediate landing sites in the CP and in the DP structures. (6) and (7) give the schematic structures and the movement steps necessary to derive them under each hypothesis. The dashed lines indicate the movement path under the CP hypothesis: there is a single intermediate gap in the CP condition and none in the DP condition. The solid lines indicate movement under the CP+vP hypothesis: there are three intermediate gaps in the CP structure, and one in the DP structure.

(6) *Relative clause structure in CP condition in Experiment 1*

CP hypothesis

. . . $[_{CP}$ *who* the lawyer $[_{vP}$ __ claimed $[_{CP}$ __ that the accusations had $[_{vP}$ __ hurt __]]]] . . .

CP+vP hypothesis

(7) *Relative clause structure in DP condition in Experiment 1*

CP hypothesis

. . . $[_{CP}$ *who* $[_{DP}$ the lawyer's claim about the accusations] had $[_{vP}$ __ hurt __]] . . .

CP+vP hypothesis

The results of Experiment 1 do not necessarily allow us to distinguish between the two hypotheses in (5). The results are clearly compatible with the predictions of the CP hypothesis: Because there exists an intermediate gap in the CP structure but not the DP structure, retrieval of the moved element at the final gap position is easier in the CP structure, thus giving rise to a smaller reading-time increase than in the DP structure.

Whether or not they are also compatible with the CP+vP hypothesis depends on the mechanics of how the moved element is reactivated over the course of the dependency. For the sake of the argument, consider the simplest version: When the ultimate gap site is encountered, the parser has to retreat to its closest antecedent—either the moved element itself, or an intermediate gap. On the CP+vP hypothesis, the most recent intermediate gap in both the CP and the DP structures lies in the vP edge immediately above the verb that hosts the gap. In other words, the distance between the ultimate gap and the closest intermediate gap is identical in the two structures in (6) and (7). If retrieving the antecedent amounted to a search for the most recent gap, there should be no difference between the two structures, contrary to fact. On this simple view of antecedent retrieval, the results of Experiment 1 favor the CP hypothesis.

There is, however, reason to believe that this picture of how the antecedent is retrieved is overly simplistic. Based on anti-locality effects in processing, Vasishth & Lewis (2006) argue that retrieval speed is also affected by the number of times an element has been previously activated (also see Lewis & Vasishth 2005). Simplifying somewhat, on their account, retrieval of an item is easier the higher the activation level of this item is. Each retrieval of an item from memory boosts that item's activation level and thereby facilitates subsequent retrievals. Applied to the structures at hand, if the representation of the moved item is retrieved from memory at each intermediate landing site, each such landing site should reactivate it. The difficulty of integrating the filler at the ultimate gap site should then be inversely related to the number of intermediate gaps along the movement path. In other words, the more intermediate gaps exist between the overt position of the moved item and its ultimate gap, the faster the postulation and interpretation of this gap proceeds. This view thus contrasts with the simpler one just considered in that intermediate landing sites have a *cumulative* effect on the processing speed at the gap position.[20]

Given this cumulative view of filler retrieval, let us consider again the predictions of the CP and the CP+vP hypothesis for the structures in (6) and (7). The predictions of the CP hypothesis remain unchanged: Because there is only a single intermediate gap in the CP structure and none in the DP structure, cumulative and non-cumulative reactivation are indistinguishable. The CP+vP hypothesis, by contrast, predicts that the antecedent is intermediately reactivated three times in

[20] In addition to the empirical evidence by Vasishth & Lewis (2006), this view is very plausible in light of current accounts of the mechanisms underlying retrieval processes in online parsing (see, among others, Stevenson 1994, McElree et al. 2003, Lewis et al. 2006, Wagers & Phillips 2009, 2013, Bartek et al. 2011). According to this family of theories, the ease with which an item is retrieved from memory is affected by this element's activation level and interference from similar encodings. An element's activation level is subject to decay over time. Retrieval of an element boosts its activation level and thereby counteracts the effects of decay and interference. As mentioned in the text, reactivation is crucially taken to be cumulative so that a sequence of reactivations boost an element's activation level to a greater extent than a single reactivation, thus aiding subsequent retrievals.

the CP structure (corresponding to the three intermediate landing sites), but only once in the DP structure. If reactivation is cumulative, the antecedent's activation level will be greater in the CP than in the DP structure at the point at which the ultimate gap is encountered, and integration of this gap will as a result be faster in the CP structure than the DP structure. As both hypotheses thus derive the critical finding of Experiment 1, there is so far no evidence to distinguish between them if reactivation is assumed to be cumulative.

Experiment 2 is intended to differentiate between the two hypotheses by comparing the CP and DP structure to a third structure which contains an additional vP layer, but *no* additional CP layer. The CP hypothesis and the CP+vP hypothesis differ in the effect that an additional vP is predicted to have on retrieval times. Consider first the CP hypothesis: Because no intermediate gap is created at vP edges on this account, a vP will not lead to filler reactivation, and hence should not facilitate filler retrieval at the gap site. By contrast, the CP+vP hypothesis requires Spec,vP to contain an intermediate gap. This gap should reactivate the filler and facilitate subsequent filler retrieval at the ultimate gap site. On the CP+vP hypothesis, then, a vP layer crossed by movement should have effects entirely parallel to a CP layer. To assess these predictions, Experiment 2 investigates extraction out of *exceptional case marking* (ECM) constructions like (8), which is modeled after the sample stimuli in (6)–(7) above. The central characteristic of ECM constructions is that the embedded clause is not a CP, but a TP. I will henceforth refer to structures of this type as *TP structures.* Due to the absence of a CP, the CP hypothesis predicts no intermediate landing site in (8). As two vPs are crossed, the CP+vP hypothesis predicts two intermediate landing sites.

(8) *Relative clause structure in TP condition*

CP hypothesis

. . . [$_{\text{CP}}$ *who* the lawyer [$_{\text{vP}}$ __ claimed [$_{\text{TP}}$ the accusations to have [$_{\text{vP}}$ __ hurt __]]]] . . .

CP+vP hypothesis

The number of intermediate gaps postulated by the two hypotheses for each of the three structures are summarized in (9). On the null assumption that all intermediate gaps have identical effects on filler retrieval, the two hypotheses make different predictions about the difficulty of integrating the filler in the three constructions. On the CP+vP hypothesis, the additional vP layer should have the same effect as a CP layer: It should reactivate the filler and thereby facilitate its retrieval at the ultimate gap site. This hypothesis thus predicts three intermediate gaps in the CP structure, two in the TP structure and one in the DP structure. The ease of integrating the gap should mirror this order: The reading-time increase should be smallest in the CP structure, larger in the TP structure and largest in the DP structure. On the CP hypothesis, on the other hand, vP layers should critically differ from CP layers in that only the latter facilitates filler retrieval downstream. Filler retrieval should therefore be facilitated only in the CP structure and in neither DP and TP structure, as movement does not cross a CP in the latter two.

(9) *Predicted number of intermediate gaps per structure*

	CP hypothesis	CP+vP hypothesis
CP structure (6)	1	3
DP structure (7)	0	1
TP structure (8)	0	2
Predicted reading-time increase:	{DP, TP} > CP	DP > TP > CP

In sum, the critical prediction of the CP hypothesis is that movement in the CP structure should be easier than in the other two, as reflected in a reading-time increase. The critical prediction of the CP+vP hypothesis is that the movement in the TP structure should be easier than in the DP structure.

4.1 Method

Materials: Thirty-six set of sentences like the one in (10) were constructed, out of which a subset of thirty with closely matching plausibility ratings between conditions was selected, the remaining six being discarded. The details of the norming procedure are described in Appendix A. The experiment manipulated INTERVENER TYPE (*CP* vs. *DP* vs. *TP*) and MOVEMENT ([+*move*] vs. [–*move*]). To decrease the processing load at the critical verb and to increase comparability to Experiment 1, a general scene-setting context sentence preceded each target sentence. As in Experiment 1, this context sentence gave lexical information about the embedded verbal region in the corresponding target sentence. All sentences were followed by a comprehension question about the semantic relations in the sentence. All questions were multiple choice and both possible answers occurred in the target sentence. The order of the answer options was randomized.

(10) *Context:* At the press conference last Monday several people became very agitated.

a. ***CP structure***

(i) **[+*move*]:** The journalist *who*$_i$ the union member believed **[**$_{\mathbf{CP}}$ that the tax policy had intensely agitated t_i **]** was planning a series of articles.

(ii) **[–*move*]:** The union member believed that the tax policy had intensely agitated the journalist who was planning a series of articles.

b. ***DP structure***

(i) **[+*move*]:** The journalist *who*$_i$ **[**$_{\mathbf{DP}}$ the union member's beliefs about the tax policy **]** had intensely agitated t_i was planning a series of articles.

(ii) **[–*move*]:** The union member's beliefs about the tax policy had intensely agitated the journalist who was planning a series of articles.

c. ***TP structure***

(i) **[+*move*]:** The journalist *who*$_i$ the union member believed **[**$_{\mathbf{TP}}$ the tax policy to have intensely agitated t_i **]** was planning a series of articles

(ii) **[–*move*]:** The union member believed the tax policy to have intensely agitated the journalist who was planning a series of articles.

Comprehension question: Who believed something regarding the tax policy?

the union member – the journalist

Some general remarks about the design of the stimuli are in order. Most items contained the auxiliary *have* (*to have* in the TP condition). Other items contained the future auxiliary *will* (*to* in the TP condition). Because Experiment 2 seeks to test whether there is an intermediate landing site at the vP edge, it is crucial to dissociate the region containing the ultimate gap from the one containing the left edge of the vP. It is uncontroversial that finite auxiliaries and the infinitival *to* occupy T. An intermediate landing site in the vP edge can hence be reliably postulated when the parser reaches this segment of the clause. An adverb was inserted between the auxiliary and the verb hosting the gap to separate the hypothetical landing site in vP from the final gap position.

In addition to the 30 target sentences, the experiment contained 24 sentences of another experiment and 36 distractor sentences, which matched the target sentences in length and complexity.

Participants: 162 participants were recruited via Amazon Mechanical Turk. All were naïve to the purpose of the experiment and each received a compensation of USD 1.

Procedure: Like Experiment 1, the experiment was conducted using the online experiment platform Ibex and employed a region-by-region self-paced noncumulative moving-window task. Stimuli were regioned as shown in Table 5. The mode of presentation was identical to that in Experiment 1. Also like in Experiment 1, the participant had to take a ten-second break after every twelve trials and could rest for longer if they so desired. The total experiment thus contained seven of these breaks. The experiment took about 35 minutes. Items were arranged on six different lists such that each list contained one instance of every item and each of the six conditions of every item appeared on a different list. Each participant was assigned to one of the six lists.

Analysis: Preprocessing and outlier rejection followed the same procedure as Experiment 1.

4.2 Results

Of the 162 participants, one was excluded for not being a native speaker of English. Because of a coding error, one item and seven participants were excluded from the analysis. 53 of the remaining participants had to be excluded from analysis for falling below the accuracy threshold.[21]

Outlier rejection based on absolute reading-time thresholds eliminated less than 0.4% of the data. The z-score based outlier rejected excluded less than 1% of the data. Reading-time data were analyzed using a 2×3-factorial LME model crossing the factors MOVEMENT ([–*move*] vs. [+*move*]) and INTERVENER (*CP* vs. *DP* vs. *TP*). Accuracy data were analyzed using a parallel logistic LME model. All covariates were numerically coded. *Movement* was sum-coded ([–*move*] $= -.5$, [+*move*] $= .5$). The predictor *intervener* used Helmert coding: The first contrast compared the CP condition to the mean of the DP and TP conditions ($CP = -2/3, DP = 1/3, TP = 1/3$). The second contrast compared the DP and the TP condition ($CP = 0, TP = -.5, DP = .5$). This coding makes sense in light of the critical

[21] As in Experiment 1, a relatively stringent accuracy criterion was chosen as the most reliable method of ensuring that the results reflect genuine properties of the parsing of the syntactic structures in questions, and not properties of shallow parsing. The critical statistical results reported here also obtain if no accuracy-based exclusion of participants is applied (see Appendix C). The excluded item did not introduce a plausibility confound (p's $> .5$), cf. Appendix A. The coding error also led to one item being somewhat less balanced across conditions. To be conservative, analyses were conducted both with and without this item and the patterns of significance were identical in all regions. I report here the results with this item included.

Table 5. Regioning of stimuli in Experiment 2 ([±mv] = [±move])

		Region								
		1	2	3	4	5	6	7	8	9
CP	[–mv]		The union member believed	that	the tax policy	had	intensely	agitated	the journalist who	was planning a series of articles
	[+mv]	The journalist who	the union member believed	that	the tax policy	had	intensely	agitated	was planning	a series of articles
DP	[–mv]		The union member's beliefs	about	the tax policy	had	intensely	agitated	the journalist who	was planning a series of articles
	[+mv]	The journalist who	the union member's beliefs	about	the tax policy	had	intensely	agitated	was planning	a series of articles
TP	[–mv]		The union member believed		the tax policy	to have	intensely	agitated	the journalist who	was planning a series of articles
	[+mv]	The journalist who	the union member believed		the tax policy	to have	intensely	agitated	was planning	a series of articles

Table 6. Mean answer accuracy computed over participants by condition in Experiment 2

	CP	DP	TP
[–move]	.95	.84	.92
[+move]	.78	.67	.72

predictions of the two hypotheses. The first contrast tests whether the CP condition was read faster than the other two (the critical prediction of the CP hypothesis) and is reported as *IntervCP–TPDP*. The second contrast tests whether the TP condition was faster than the DP condition (the critical prediction of the CP+vP hypothesis) and is reported as *IntervTP–DP*. Because the TP condition had no observations in region 3, the factor *intervener* was sum-coded in this region ($CP = -.5, DP = .5$).

Comprehension question accuracy: The mean accuracy on the comprehension questions was 81%. Accuracy by condition is given in Table 6. Logistic LME modeling revealed that (i) accuracy was lower in the [+move] conditions ($\hat{\beta} = -1.6, z = -12, p < .001$); (ii) accuracy in the CP conditions was higher than the mean of the other two conditions ($\hat{\beta} = -.7, z = -4.8, p < .001$); and (iii) accuracy in the TP conditions was higher than in the DP conditions ($\hat{\beta} = -.6, z = -3.3, p < .001$). Moreover, while the accuracy drop in the [+move] conditions did not differ between CP structures and the mean of the other two ($\hat{\beta} = .3, z = 1.2, p > .2$), this drop was greater in the TP condition than in the DP condition ($\hat{\beta} = .6, z = 2.1, p < .05$). This finding will be picked up in the discussion section.

Reading times: The mean reading times per region and condition as well as the residual reading times for regions that differed in length are given in Table 7. The standard error by condition and region as a measure of data variability is provided in Appendix B.

Table 7. Mean raw reading times by condition and region in ms computed over participants for Experiment 2. For regions that differed in length between conditions, residual reading times are additionally provided in parentheses ('Intv' = Intervener, 'Mvmt' = Movement)

		Region								
Intv	Mvmt	1	2	3	4	5	6	7	8	9
CP	[–move]	—	844	494(10)	657	493(29)	495	522	732(–46)	1270(90)
	[+move]	699	949	559(75)	717	495(31)	502	548	677(–2)	877(–1)
DP	[–move]	—	1032	520(12)	675	495(30)	499	514	705(–66)	1215(38)
	[+move]	706	1190	618(111)	770	515(50)	507	536	714(39)	903(36)
TP	[–move]	—	852	—	775	543(9)	523	525	701(–73)	1275(99)
	[+move]	691	1015	—	833	579(44)	527	593	698(26)	888(10)
				complementizer/ preposition		*auxi- liary*	*adverb*	*verb + gap*	*spillover*	

Table 8. Coefficient estimates and corresponding t-values for the linear mixed effects model analyses of log reading times in critical regions of Experiment 2. *Mvmt* refers to the sum-coded factor *movement; IntervCP–TPDP* compares the CP condition to the mean of the TP and DP conditions and *IntervTP–DP* compares the TP condition to the DP condition. *Mvmt:IntCP–TPDP* and *Mvmt:IntTP–DP* refer to the interaction between *Mvmt* and *IntervCP–TPDP* and *IntervTP–DP*, respectively. Cells with $p < .05$ are shaded. For statistical analyses of all regions, see Appendix B.

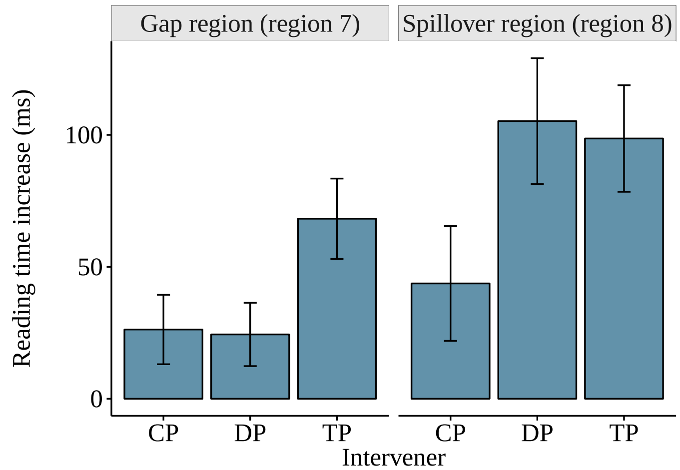
	Region 7		Region 8	
	$\hat{\beta}$	t	$\hat{\beta}$	t
Movement	0.053	4.00	–0.039	–1.11
IntervCP–TPDP	0.013	1.13	–0.000	–0.00
IntervTP–DP	–0.042	–3.65	–0.003	–0.21
Mvmt:IntervCP–TPDP	0.024	1.09	0.059	2.08
Mvmt:IntvTP–DP	–0.049	–2.10	–0.018	–0.61

To distinguish between the CP hypothesis and the CP+vP hypothesis, the crucial evidence is the pattern at the region containing the gap (region 7) and the spillover region (region 8). Of primary interest is the reading-time difference between the movement condition and the no-movement controls for each syntactic structure. Table 8 provides the results of the analyses of these two regions. In the interest of space, only the crucial gap and spillover regions are discussed in the main text. Analyses and discussions of all other regions are provided in Appendix B.

In region 7 (the gap region), reading times in the TP condition were greater than in the DP condition while the CP condition did not differ from the mean of the other two. Crucially, there was an interaction between the reading-time increase due to movement noted above and the intervener type: The reading-time increase was greater in the TP condition than in the DP condition. Pairwise comparisons made it clear that this pattern was driven by an exceptionally high reading-time increase in the TP condition: The increase was greater than in the CP condition, marginally significant by participants ($t_1(100) = 1.89, p = .06$) and fully significant by items ($t_2(28) = 2.8, p < .01$) but did not differ between the CP and DP conditions ($t_1(100) = .2, p > .5;\ t_2(28) = .3, p > .5$).

In region 8 (the spillover region), the only significant effect was an interaction such that the reading-time difference between movement and non-movement structures was significantly smaller in the CP condition than in the mean of the other two. The latter two did not differ from each other. Because the length of the region differed between movement and non-movement structures, an additional LME analysis of the residual reading times was carried out to validate this effect. This analysis corroborated the interaction ($\hat{\beta} = 59, t = 2.5, p < .05$). It also showed that the [+move]

Figure 2. Mean increase in residual reading times from [–move] to [+move] conditions in gap and spillover regions computed over participants and by-participant standard errors in Experiment 2 by intervener



conditions were read more slowly than the [–move] ones ($\hat{\beta} = 81, t = 5.2, p < .001$). Just like in Experiment 1, the fact that the region was longer in the [–move] condition than in the [+move] one renders the effect of movement visible only in the residual reading times.

Figure 2 plots the crucial reading-time increase from the [–move] to the [+move] conditions in the gap and spillover regions. As the analyses just described reveal, (i) there is a positive increase in all conditions, (ii) the increase is significantly larger for TP structures in the gap region and (iii) the increase is significantly smaller for CP structures in the spillover region.

4.3 Discussion

CPs vs. vPs: Recall from (9) the predictions of the two hypotheses: The critical prediction of the CP hypothesis is that a CP crossed by movement will lead to an intermediate reactivation of the filler, while a vP has no such effect. On the CP+vP hypothesis, both CPs and vPs are expected to reactivate the filler. The CP hypothesis thus predicts that the reading-time increase in the CP condition is smaller than in the DP and the TP condition, whereas the CP+vP hypothesis additionally predicts that the reading-time increase is greater in the DP condition than in the TP condition.

The results of Experiment 2 show facilitation in the CP structure relative to the other two, both in the log-transformed and of the residualized reading times. In line with the findings of Gibson & Warren (2004) and Experiment 1 above, this result provides evidence that crossing a CP layer reactivates the filler, indicating the presence of an intermediate landing site in Spec,CP. Crucially, however, there was no indication that crossing a vP has a similar effect. Not only was the retrieval speed in the TP condition (where two vPs were crossed) not faster than in the DP control structures (where only one vP was crossed), retrieval times were in fact *slower* in the TP structure than in the DP structure (again both in the log-transformed and the residualized reading times). This indicates that the addition of a vP layer in the TP structure does *not* aid filler retrieval at the gap site, but in fact incurs a greater difficulty of filler retrieval, a result to which we will return shortly. The crucial finding of the experiment is thus that CPs and vPs differ substantially with respect to the processing of a filler–gap dependency: While a CP layer leads to reactivation of the filler, a vP layer does not.

This differential effect of CPs and vPs on filler reactivation receives a straightforward explanation if only CPs, but not vPs, host intermediate landing sites, i.e., under the CP hypothesis. The observed contrast between CPs and vPs is not predicted by the CP+vP hypothesis, on the other hand, precisely because this hypothesis treats CPs and vPs on par. Thus, the hypothesis that CPs are phases and that vPs are not offers a principled explanation for the results. Alternative accounts are considered below.

The results of Experiment 2 also indicate that there is no successive cyclicity in ECM constructions more generally.[22] In other words, these results not only cast doubt on successive cyclicity through Spec,vP, they also suggest that there is no clause-level cyclicity in TP clauses. This finding challenges recent claims that the highest projection of an extended projection is phasal, regardless of its category (e.g., Bošković 2014). To the extent that these conclusions are on the right track, they converge with those above that only C is a phase head (at least within the verbal domain).

In addition to the critical interaction, the critical region also exhibit greater reading times in [+move] structures than in [–move] structures and likewise in TP structures than in DP structures. The former effect plausibly reflects the cost of retaining an unassigned filler in memory. The latter is difficult to interpret given the higher-level interaction.

Finally, because the movement dependency crosses a verb that is able to take a DP object in both the TP and the CP structures, the CP–TP contrast provides additional evidence against exclusively attributing this facilitation to premature gap filling at the higher verb, because such premature gap filling would apply in CP and TP structures equally. The CP–TP difference hence argues in favor of

[22]Thanks for Grant Goodall for pointing this out to me.

successive cyclicity at CP edges. This is consistent with the interpretation of Experiment 1 above.

The role of structural distance: While the results thus bear out the critical prediction of the CP hypothesis because only an embedded CP layer correlates with facilitated filler retrieval, the overall difficulty of the TP condition—both in the reading time and accuracy measures—was unexpected under both hypotheses and thus deserves discussion. Note that this increased difficulty of the TP structures relative to the DP structures directly contradicts the predictions of the CP+vP hypothesis that filler retrieval is easier in TP structures than in DP structures. By contrast, the CP hypothesis does not by itself make a commitment with regard to the relation between the TP and DP structures. The failure of the CP hypothesis to predict the TP–DP contrast hence does not put it on a par with the CP+vP hypothesis: Whereas the former merely fails to predict an observed contrast, the latter critically predicts the opposite of the observed results. Yet the question remains as to why filler retrieval should be exceptionally hard in TP structures. One plausible source of this difficulty is *structural distance*. Although the materials used in Experiment 2 controlled the *linear* distance between the overt filler and its gap, the *structural* distance between the two differed across conditions. Concretely, in the DP condition, the moved element was in the same clause as its gap (as no clause boundary was crossed), as shown in (11) below. In the TP condition, on the other hand, the moved element was in a higher clause than its gap (see (12)). Retrieval of the filler thus has to traverse a clause boundary (or, equivalently, another thematic domain) in the TP condition. It is then possible that traversing a clause boundary for filler retrieval incurs an additional processing load. It is worth noting that this line of account converges with the processing study of Dillon et al. (2014), who found effects of clause boundaries in the processing of the Mandarin long-distance anaphor *ziji*.[23]

Importantly, this structural distance-based account of the TP–DP difference does not affect the argument for the CP hypothesis, nor is it able to reconcile the CP+vP hypothesis with the results. While crossing a clause boundary induces a penalty in the TP condition relative to the DP condition, it must not do so in the CP condition: If movement in the CP condition were crossclausal in the

[23]Because Dillon et al.'s (2014) experimental design involved clausal embedding in Mandarin Chinese, their results do not reveal whether there is a distinction between finite and nonfinite clauses. As a reviewer points out, on the view put forth here, similar effects as those observed by Dillon et al. should hold for nonfinite clauses as well.

same way that movement in the TP condition is, they should pattern alike in being harder than the DP condition, contrary to fact. The ease of postulating the gap in the CP condition therefore shows that the structural distance is shorter in this condition than in the TP condition. This is, of course, precisely what successive cyclicity through Spec,CP claims. In other words, the penalty for crossclausal movement is absent in the CP condition precisely because there exists an intermediate gap within the same clause (see (13)). An appeal to structural distance hence does not obviate the need for successive cyclicity through Spec,CP.[24]

(11) *Movement in DP structure under CP hypothesis* ⇝ *intermediate*

[$_{\text{CP}}$ *who* [$_{\text{DP}}$ the union member's belief about the tax policy] had [$_{\text{vP}}$ intensely agitated *t*]]

(12) *Movement in TP structure under CP hypothesis* ⇝ *most difficult*

[$_{\text{CP}}$ *who* the union member [$_{\text{vP}}$ *t* believed [$_{\text{TP}}$ the tax policy to have [$_{\text{vP}}$ intensely agitated *t*]]]]

clause boundary

(13) *Movement in CP structure under CP hypothesis* ⇝ *least difficult*

[$_{\text{CP}}$ *who* the union member [$_{\text{vP}}$ believed [$_{\text{CP}}$ *t* that the tax policy had [$_{\text{vP}}$ intensely agitated *t*]]]]

clause boundary

Thus, I have suggested that the difficulty of filler retrieval in the TP structure can be given a plausible explanation if successive cyclicity through CP is combined with considerations of structural distance. It is due to the fact that (i) the filler is not intermediately reactivated (due to the lack of successive cyclicity) and that (ii) the parser has to access a higher clause to retrieve the closest antecedent of the gap. In the DP structure, (ii) does not arise. In the CP structure, neither (i) nor (ii) arises.

As mentioned above, equally importantly, a structural distance account in conjunction with the CP+vP hypothesis does not alter the incorrect prediction of this hypothesis. To see why, consider

[24]From the standpoint of current theories of sentence parsing, it is interesting to note that the difference between the TP and the DP conditions cannot be solely attributed to activation decay and similarity-based interference (e.g., McElree et al. 2003, Van Dyke & Lewis 2003, Lewis & Vasishth 2005, Lewis et al. 2006, Bartek et al. 2011). First, because the DP and TP structures do not differ with respect to the linear distance between the moved element and its gap, the effect of decay should be identical. Second, because the DP structure involves an additional noun (*beliefs* in (10)), similarity-based interference would predict filler retrieval to either be harder in the DP condition than in TP condition, the inverse of what was observed, or for there to be no difference at all. This indicates that there must be some other factor at play. Crossing of a clause boundary is one viable hypothesis.

again the structures in (6–8) above. On the CP+vP hypothesis, the intermediate gap closest to the ultimate gap site is at the edge of the embedded vP in all three structures. As a consequence, the structural distance between the gap and its closest antecedent is *identical* in all three structures and no clause boundary intervenes between the gap and its closest antecedent in any of the three structures. Thus, integrating structural distance into the analysis does not affect the predictions of the CP+vP hypothesis. In particular, it has no impact on the incorrect prediction that the integration of the gap should be easier in the TP structure than in the DP structure.

In sum, while reference to structural distance does provide an account for the difficulty in the TP condition, it does not modify the conclusion drawn above: The results of Experiment 2 are in line with the CP hypothesis because the application of structural distance is successful only on this hypothesis. As a reviewer notes, one consequence of this interpretation of the results is that locality domains for processing do not necessarily coincide with locality domains for the grammar: TP clauses constitute a domain for retrieval but not for phasehood or successive cyclicity.[25]

Alternative interpretations: I have argued that it is the grammatical representation of the movement dependency, specifically the representation of intermediate landing sites, that underlies the reading-time results in Experiment 2. There are a number of potential alternative explanations in terms of processing principles, which warrant consideration.

First, while the results of Experiment 2 show that only CP layers lead to facilitated retrieval of the filler at the gap site, it is a priori possible that both CPs and vPs host intermediate landing sites, but that only gaps in Spec,CP reactivate the filler. A difference between CP and vP gaps along this line would reconcile the CP+vP hypothesis with the results of Experiment 2 as it would effectively render gaps in Spec,vP irrelevant for the purposes of filler retrieval. While the experimental designs

[25]It is likewise irrelevant for the interpretation of the results whether filler reactivation is cumulative or not. As discussed on p. 24 above, cumulative reactivation refers to the hypothesis that a sequence of reactivations eases an element's subsequent retrieval to a greater extent than a single reactivation. As also discussed there, the CP+vP hypothesis is compatible with the results of G&W and Experiment 1 only if cumulative reactivation is adopted. However, even under cumulative reactivation the CP+vP hypothesis fails to account for the results of Experiment 2, as it predicts filler retrieval to be faster in the TP condition than in the DP condition (see (9)). Consequently, the CP+vP hypothesis makes incorrect predictions regardless of whether or not cumulative reactivation is adopted. The CP hypothesis, by contrast, makes correct predictions for both experiments reported here regardless of whether cumulative reactivation is adopted or not. This is trivially the case because on this hypothesis no structure investigated here involves more than a single intermediate landing site. Cumulative and non-cumulative reactivation are hence indistinguishable.

here do not allow us to assess this account directly, the extent to which this account constitutes a feasible alternative to the CP hypothesis depends on whether there is independent motivation for the claim that intermediate gaps in CP and vP differ in whether or not they lead to filler reactivation. It is not evident that there is. First, such a difference could not plausibly emerge from the *syntax* of successive cyclicity because successive-cyclic movement through CP and through vP are formed by the same syntactic mechanism. Second, it is not clear that there is independent *processing* motivation for such a contrast. In fact, standard parsing accounts that incorporate fluctuating activation explicitly assume that retrieval of an item invariably leads to reactivation of this item (e.g., Lewis & Vasishth 2005, Vasishth & Lewis 2006) and it is not clear on this view how the filler could be retrieved at vP without concomitantly reactivating it.[26] I am furthermore not aware of empirical evidence for such reactivation asymmetries. In fact, clause-internal reactivation of non-moved elements is indeed attested (Vasishth & Lewis 2006), and there seems to be no reason to expect moved elements to behave differently. In sum, reconciling the results of Experiment 2 with the CP+vP hypothesis along these lines would require independent evidence for the existence of reactivation asymmetries of this type as well as general processing principles that predict intermediate gaps in CPs and vPs to differ in a way that is consistent with the results. In the absence of such independent support, I take the reading-time pattern to pose a challenge for the view that vPs are phases alongside CPs.

A second point that warrants consideration is that the higher verb in Experiment 2 was a CP/DP-verb (using the terminology of Experiment 1), a direct consequence of the experimental design. Given the results of Experiment 1, it is expected that the parser initially postulates a gap of the moved element as the object of these verbs in the TP and CP conditions, but not in the DP condition. It is then a priori possible that the filled-gap effect in the TP condition affected the results at the actual gap site. The results of Experiment 2 do not allow us to directly investigate the impact

[26]One line of analysis could be that intermediate gaps give rise to reactivation only if they reside at a clause edge, a line of account that would liken intermediate reactivation to clausal wrap-up effects. While a priori possible, it is unclear why intermediate reactivation should be principally restricted to clause boundaries. Notice, for instance, that in contrast to wrap-up effects at the end of sentences, thematic processing in the higher clause is not completed by the time the CP of the lower clause is postulated, simply because this CP is itself a thematic argument of the higher verb. Moreover, this account would appear to be at variance with evidence for retrieval-induced activation of a non-moved constituents occurring in the middle of a clause (e.g., Vasishth & Lewis 2006). Of course, only specific proposals can be conclusively evaluated.

of this confound, but there is reason to doubt that the crucial pattern at the gap and spillover regions are an artifact of premature gap filling at the higher verb. First, incorrect association of the filler with the embedding verb would not only arise in the TP condition, but also in the CP condition. Because the reading-time increase was smallest in the CP condition, this incorrect association would have had to differentially affect only the TP condition, which would seem to be surprising on such an account.[27] Second, in Experiment 1, verb type had an impact on the reading-time pattern in the region following the higher verb, but not in the gap region. This indicates that structural revision is fast and a parse temporarily pursued in the higher clause has no discernible impact on the regions of interest, in concordance with the previous literature on structural reanalysis. Third, the crucial reading-time increase is not modulated by the intervener structure in any region prior to the gap region (see Section B.2 in the supplementary materials), which equally makes it implausible that the pattern at the gap region is a sustained effect of gap filling at the higher verb. The timing of the effect thus discourages the view that the critical results in the TP condition are an artifact of premature gap filling, but a more direct exploration of this possibility would require a follow-up study.

A third noteworthy aspect of the design of Experiment 2 is that the TP structure contained a sustained ambiguity in the regions following the ECM verb. A DP following an ECM verb could be either a direct object to this verb, the subject of an ECM infinitive, or the subject of an embedded finite clause (without a complementizer). By contrast, in the CP condition, the complementizer and the following DP disambiguate the input towards the correct structure earlier. One might wonder

[27]A reviewer points out that there is evidence that the severity of a disruption may vary across garden-path situations (e.g., Fodor & Inoue 1994, Fodor & Ferreira 1998), and the difference between CP and TP structures might create such a disparity. In order to assess the relationship between the garden-path effect at the embedding verb and the crucial effect at the embedded verb, a post-hoc analysis was carried out, which calculated, for each item, the reading time increase between [–move] and [+move] structures in the TP condition (in log reading times) in region 7 (the gap region). The mean of the corresponding difference in the CP and DP conditions was then subtracted from this increase. This yields an estimate of the difficulty of movement in the TP structure relative to the other two structures in the gap region for each item. The same difference was calculated for each item in region 4 (the beginning of the embedded clause in the TP condition). If the difficulty of the TP condition in the gap region represents a sustained garden-path effect incurred at the beginning of the embedded clause, then we might expect that the difficulty in the gap region is greater for items which display a greater reading disruption at the beginning of the embedded clause. A linear model was fitted that regressed the TP difficulty score for each item at the gap region on the corresponding TP difficulty score in region 4. The model did not indicate a correlation between the two measures ($r = 0.05$, *adjusted* $R^2 = -0.03$, $F(1, 27) = 0.08$, $p > .5$) and hence does not suggest that the TP difficulty at the gap region is systematically related to the difficulty at the active-gap filling site at the embedding verb. This finding does not support the alternative account discussed in the main text.

whether the increased difficulty in the TP structure could be attributed to this sustained ambiguity. As mentioned, the experimental design makes it impossible to directly assess this alternative. But due to the factorial design employed in the experiment, effects of baseline difficulty of the three structures are 'filtered out' in the crucial interactions. That is, the critical dependent measures are reading-time increases in a movement structure relative to a non-movement baseline. Because these structural properties of ECM constructions also arise in the non-movement baselines, these properties should not affect the reading-time increase and hence the interaction.[28] Moreover, the time course of the processing difficulty is unexpected under this account. As just mentioned, the difficulty in the TP structure is limited to the gap and the spillover regions (regions 7 and 8). On the ambiguity account, the ambiguous regions of the TP condition are those immediately following the embedding verb (region 4). The ambiguity account thus faces the challenge of explaining why the difficulty purportedly created by the structural indeterminacy does not arise until after this indeterminacy has been eliminated. By contrast, the successive cyclicity account attributes the effect to filler retrieval at the ultimate gap site and thus provides a straightforward explanation for its location.

Fourth, a reviewer points out an interesting alternative interpretation of the results of Experiment 2, according to which the difficulty in the TP condition arises from the fact that vP *is* a phase and as a result has to attract the moving element to its specifier, in addition to assigning accusative case to the ECM object (see, e.g., Rezac 2013). This might then produce the observed interaction if the cost

[28]This reasoning presupposes that the costs of reanalysing the ECM object as an embedded subject and the cost of the filled gap effect are additive. As a reviewer points out, the TP [+move] conditions involve 'double reanalysis': (i) the moved element is incorrectly parsed as the object of the ECM verb, and (ii) the embedded subject is initially analyzed as the object of the ECM verb. Both parses require reanalysis, and if they are superadditive, it is conceivable that the TP difficulty at the actual gap site might be a late effect of this double reanalysis required earlier. A post-hoc analysis was carried out to assess to what extent the available evidence supports this possibility. The goal of this analysis was to determine whether the TP difficulty at the gap region correlates with (i) the filled-gap effect at the matrix verb and (ii) the object-to-subject reanalysis cost at the embedded subject. To obtain an independent measure of (i), the logRT in region 3 (the complementizer region) in the CP [–move] condition was subtracted from the corresponding logRT in the CP [+move] condition for each item. In order to gain an approximation to (ii), the by-item mean log reading times in region 5 in the CP [–move] condition were subtracted from the TP [–move] log reading times, given that object-to-subject reanalysis is required in the TP structure but not the CP structure. A linear model then regressed the by-item difficulty score in the TP [+move] condition at the gap site (derived as described in fn. 27) on (i) the by-item filled-gap effect, (ii) the by-item object-to-subject reanalysis score, and (iii) their interaction. The rationale for this model was that if the TP difficulty at the gap site is a late effect of the superadditive interaction of the two reanalyses required earlier, then items for which these two reanalyses are more difficult should also exhibit a greater TP difficulty at the gap site. However, the model did not indicate a systematic relationship with either of the two main effects or the interaction (all p's $> .5$, $\textit{adjusted } R^2 = -0.09, \mathrm{F}(3, 25) = .2, p > .5$). See also fn. 27.

of these two processes is superadditive. This alternative differs from the account proposed here in where it locates the source of the difficulty: The account proposed here attributes it to filler retrieval at the ultimate gap site; the alternative account attributes the difficulty to ‘double duty’ by matrix v. A direct comparison of this alternative line of explanation to the retrieval-based account here would require a separate investigation, but within the scope of the present results, the timing of the reading disruption might present a challenge to this alternative account. As the reviewer notes, if this disruption is caused by the matrix v, we might expect it to occur before the embedded verb is encountered, plausibly at the embedded *to* (region 5), which disambiguates the embedded clause towards an ECM structure, hence necessitating double duty by v. But as shown in Appendix B.2, the reading disruption in the TP [+move] condition is limited to the embedded verb and the spillover region (regions 7 and 8), with no indication of a similar disruption in the immediately preceding region (region 6). Based on the available evidence, then, an account in terms of filler retrieval at the ultimate gap site seems to provide a closer match to the observed timing of the effect.

In sum, the results of Experiment 2 show reactivation of a moved element if a CP is crossed, but no similar reactivation takes place if a vP is crossed. This pattern matches the critical prediction of the CP hypothesis but not that of the CP+vP hypothesis. The results can therefore be accounted for if landing sites are created in Spec,CP but not in Spec,vP. I have discussed a number of alternative interpretations of the results and demonstrated the challenges that arise for such alternatives. Investigating these alternatives directly would be a fruitful avenue for future work.

5. Experiment 3

One limitation of the design in Experiments 1 and 2 is that the spillover regions differed across the [+move] and [–move] conditions in a number of respects, including lexical material, length, syntactic complexity, etc. As noted in section 3.3, this length difference was orthogonal to the critical interactions (i.e., the reading-time increase from [–move] to [+move]), and it hence plausibly does not directly impact them. But in order to investigate this concern more directly, Experiment 3 modified the design of Experiment 2 in a way that avoided the length confound.[29]

[29] I a indebted to a reviewer for suggesting the design of Experiment 3.

5.1 Method

Materials: The thirty items used in Experiment 2 were modified as in (14). Like Experiment 2, this experiment manipulated the factors INTERVENER TYPE (*CP* vs. *DP* vs. *DP*) and MOVEMENT ([+*move*] vs. [–*move*]). The materials differed from those in Experiment 2 in two structural respects. First, the adverb that modifies the verb of the relative clause (*intensely* in (14)) follows the verb (note that this change results in a Heavy NP Shift structure in the [–move] conditions). Second, the stimuli contain an additional level of clausal embedding in the [–move] conditions. This change incorporates a suggestion by Marinis et al. (2005), who note that in G&W's stimuli, the level of embedding of the critical verb differs between [+move] and [–move] conditions, and the same is true for Experiment 1 and 2 here. Adding a level of embedding in the [–move] conditions (*the journalist claimed* in (14)) removes this disparity. Experiment 3 used the stimuli from Experiment 2 modified in this way. Additionally, some adverbs were replaced in order to license a relative-clause internal construal in the [+move] conditions, and a few lexical adjustments were made to facilitate the Heavy NP Shift structure and accommodate the additional DP in the [–move] conditions.

(14) *Context:* At the press conference last Monday several people seemed very agitated.

a. **[+*move*]:** The journalist who { *CP:* the union member believed that the tax policy had / *DP:* the union member's beliefs about the tax policy had / *TP:* the union member believed the tax policy to have } agitated intensely was planning a series of articles.

b. **[–*move*]:** The journalist claimed that { *CP:* the union member believed that the tax policy had / *DP:* the union member's beliefs about the tax policy had / *TP:* the union member believed the tax policy to have } agitated intensely the prolific news editor who was planning a series of articles.

Comprehension question: Who believed something regarding the tax policy?

the union member – the journalist

The positioning of the adverb (*intensely*) after the critical verb (*agitated*) resolves the length confound in Experiments 1 and 2 because it makes the region following the verb identical across conditions. Note that the resulting stimuli are somewhat more complex than those of Experiment 2: First, the

Heavy NP Shift structure in the [–move] conditions imposes additional processing demands (e.g., Staub et al. 2006 observe increased processing difficulty on the material that intervenes between an obligatorily transitive verb and its object in Heavy NP Shift structures). Second, the position of the adverb in the [+move] conditions creates a structural ambiguity between attachment within the relative clause and one inside the matrix clause. Furthermore, the [–move] conditions involve an additional DP referent. In response to this increased complexity of the stimuli, a slightly lower accuracy threshold was employed (see 'Analysis' below). In addition to the target items, Experiment 3 used the same 60 filler sentences as Experiment 2.

Participants: 158 English native speakers were recruited via Amazon Mechanical Turk. All were naïve to the purpose of the experiment and each received a compensation of USD 3.

Procedure: Stimuli were regioned in shown in Table 9. The crucial regions of interest are region 7 (the verb region) and region 8 (the spillover region containing the adverb), both of which are identical across conditions. The experimental procedure was identical to Experiment 2.

Analysis: In light of the additional processing complexities introduced by the modified design noted above, a slightly lower accuracy criterion was adopted, with participants whose overall accuracy across all trials was lower than 70% being excluded from analysis. Like in Experiments 1 and 2, reading times smaller than 200 ms or greater than 5000 ms were excluded, as were log reading times that were more than three standard deviations away from the condition mean in that region.

5.2 Results

If a participant took the experiment more than once, all but their first participation were excluded. This step left results from 142 participants. 35 of these participants were excluded for falling below the accuracy threshold.[30] Outlier rejection based on absolute reading-time thresholds and z-scores excluded 1.2% and 0.8% of the data, respectively. These exclusions resulted in one participant not having any reading time data in one experimental condition of region 8. This participant was excluded from analysis. Accuracy and reading-time data were analyzed as described for Experiment 2.

[30]The critical statistical results reported here also hold if no accuracy-based exclusion of participants is applied, with the exception that the marginally significant interaction between *movement* and the TP–DP comparison in region 8 was fully significant in this analysis (see Appendix C).

Table 9. Regioning of stimuli in Experiment 3 ([±mv] = [±move])

		Region									
		1	2	3	4	5	6	7	8	9	10
CP	[–mv]	The journalist claimed	that	the union member believed	that	the tax policy	had	agitated	intensely	the prolific news editor who	was planning a series of articles
	[+mv]		The journalist who	the union member believed	that	the tax policy	had	agitated	intensely	was planning	a series of articles
DP	[–mv]	The journalist claimed	that	the union member's beliefs	about	the tax policy	had	agitated	intensely	the prolific news editor who	was planning a series of articles
	[+mv]		The journalist who	the union member's beliefs	about	the tax policy	had	agitated	intensely	was planning	a series of articles
TP	[–mv]	The journalist claimed	that	the union member believed		the tax policy	to have	agitated	intensely	the prolific news editor who	was planning a series of articles
	[+mv]		The journalist who	the union member believed		the tax policy	to have	agitated	intensely	was planning	a series of articles

Table 10. Mean answer accuracy computed over participants by condition in Experiment 3

	CP	DP	TP
[–move]	.73	.72	.76
[+move]	.72	.68	.73

Comprehension question accuracy: The mean accuracy on the comprehension questions was 72%. Accuracy by condition is provided in Table 10. To ensure convergence of the LME model, random intercepts for items were dropped. Accuracy was marginally lower in DP structures than in TP structures ($\hat{\beta} = -0.24, z = -1.88, p = .06$), all other comparisons were not significant ($p > .2$). The lower accuracy in the [–move] conditions relative to Experiment 2 is likely the result of the presence of an additional DP in this condition in Experiment 3.

Reading times: The mean reading times per region and condition computed over participants are provided in Table 11. The standard error by condition and region is provided in Appendix B. Table 12 provides the results of the LME analyses of regions 7 (the verb region) and region 8 (the spillover region containing the adverb). Analyses of other regions are provided in Appendix B.

In region 7 (the gap region), [+move] structures elicited higher reading times than [–move] structures, and CP structures were read faster than the mean of the DP and TP structures. There was no interaction between the two factors.

Analysis of region 8 (the spillover/adverb region) revealed a number of effects: (i) reading times were greater in [+move] structures than in [–move] structures, (ii) reading times in the DP structures were faster than in the TP structures, (iii) the reading time increase between [+move] and [–move] structures was smaller in CP structures than in the mean of the TP and DP structures. Additionally,

Table 11. Mean raw reading times by condition and region in ms computed over participants for Experiment 3 (‘Intv’ = Intervener, ‘Mvmt’ = Movement)

		Region									
Intv	Mvmt	1	2	3	4	5	6	7	8	9	10
CP	[–move]	918	498	966	557	736	491	525	594	1105	1402
	[+move]	—	819	1164	583	752	500	592	630	745	1018
DP	[–move]	933	496	1064	585	687	516	536	545	1054	1392
	[+move]	—	797	1341	652	774	535	620	613	715	988
TP	[–move]	901	494	984	—	840	597	556	579	1116	1431
	[+move]	—	789	1116	—	995	611	652	673	730	981
					complementizer/ preposition		*auxi-liary*	*verb + gap*	*adverb (spillover)*		

Table 12. Coefficient estimates and corresponding t-values for the linear mixed effects model analyses of log reading times in critical regions of Experiment 3. *Mvmt* refers to the sum-coded factor *movement; IntervCP–TPDP* compares the CP condition to the mean of the TP and DP conditions and *IntervTP–DP* compares the TP condition to the DP condition. *Mvmt:IntCP–TPDP* and *Mvmt:IntTP–DP* refer to the interaction between *Mvmt* and *IntervCP–TPDP* and *IntervTP–DP*, respectively. Cells with $p < .05$ are shaded. For statistical analyses of all regions, see Appendix B.

	Region 7		Region 8	
	$\hat{\beta}$	t	$\hat{\beta}$	t
Movement	0.091	5.93	0.104	7.06
IntervCP–TPDP	0.049	4.22	–0.007	–0.53
IntervTP–DP	–0.027	–1.58	–0.067	–4.87
Mvmt:IntervCP–TPDP	0.027	1.05	0.054	2.13
Mvmt:IntvTP–DP	0.003	0.12	–0.052	–1.89

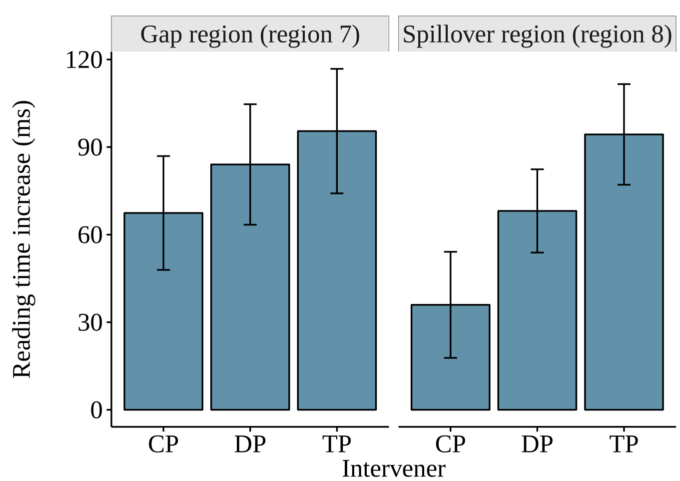
this reading time increase was numerically greater in TP structures than in DP structures, a contrast that was marginally significant ($\hat{\beta} = -0.052, t = -1.89, p = .062$).

Figure 3 plots the reading-time increase from [–move] to [+move] in the two regions of interest.

5.3 Discussion

The results of Experiment 3 are consistent with those of Experiment 2. Like in Experiment 2, the reading time increase in the CP structure was smaller than in the TP and DP structures in the spillover region. Because the spillover region did not differ in length and content across conditions in Experiment 3, this replication indicates that the effect was not an artifact of the length difference in Experiment 2. The results of Experiment 3 also replicated numerically the finding that the increase is greater in the TP condition than in the DP condition. Moreover, the materials in Experiment 3 included an additional level of embedding in the [–move] conditions, incorporating an adjustment

Figure 3. Mean increase in raw reading times from [–move] to [+move] conditions in gap and spillover regions computed over participants and by-participant standard errors in Experiment 3 by intervener



from Marinis et al. (2005). The results of Experiment 3 indicate that the key findings of Experiment 2 also arise if this adjustment is made.

Following the line of reasoning in the interpretation of Experiment 2, the results of Experiment 3 suggest that the filler is reactivated in the CP structure, but not in either the TP or the DP structure. This pattern of results can be accounted for in a principled manner if it is crossing a CP, but not crossing a vP that reactivates the filler. The hypothesis that only CP is a phase and hence only CP hosts an intermediate landing site enables an explanation of this asymmetry.

6. General discussion

This paper has explored evidence from real-time sentence processing for the existence and distribution of intermediate landing sites and hence phases, building on and extending the pioneering work of Gibson & Warren (2004). Overall, this evidence suggests that movement dependencies that cross a CP are easier to process at their gap position than dependencies that do not cross a CP. This finding is accounted for if crossing a CP requires an intermediate landing site of the moved element, which leads to reactivation of the filler, facilitating subsequent retrieval at the ultimate gap site. Experiment 1 contrasted Gibson & Warren's (2004) successive-cyclicity account with an alternative premature gap filling account, which does not resort to successive cyclicity. The results of this experiment, along with those of Experiments 2 and 3, indicate that the facilitation in the CP structure cannot be solely attributed to premature gap filling, though the two lines of explanation are not mutually exclusive. Reading-time evidence, then, provides new support for the view that CPs are phases.

Second, Experiments 2 and 3 extended this basic experimental methodology to assess whether vPs have a similar facilitatory effect on gap creation. The line of reasoning was that if vPs host an intermediate landing site analogously to CPs, they too should induce facilitation. By contrast, if vPs are not phases, they should not aid filler retrieval at the gap site and hence not induce facilitation. Experiments 2 and 3 directly contrasted these two hypotheses. The results indicate that a vP crossed by movement has no facilitatory effect on dependency completion, in direct contrast to CPs. This asymmetry between CPs and vPs is of course precisely what is predicted if intermediate landing sites are created in Spec,CP, but not in Spec,vP, hence if CPs are phases and vPs are not. Moreover, the results suggest that there is no clause-level successive-cyclic movement in TP clauses, a finding that is likewise accounted for if only CPs are phases (in the clausal domain). Taken together, the experimental results provide novel support for a type of successive cyclicity that is well-established (Spec,CP), but they found no comparable evidence for a type of successive cyclicity that is less securely established independently (Spec,vP). Attributing phase status to CP but not vP offers an account of these findings.

The investigation of structures that involve an additional vP layer without also contributing a CP layer has been limited to ECM infinitives here. ECM constructions were chosen because it is fairly uncontroversial that they lack a CP layer. If the interpretation here is on the right track, then similar patterns of results should be observed in other constructions with the relevant properties as well. However, at least obvious candidates for such extensions might be confounded. For example, control constructions likewise involve embedding of a nonfinite clause, but these clauses are often taken to contain a CP (e.g., Landau 2004). If so, then they are unsuitable to distinguish between the two hypotheses investigated here. Another potential testing case might be subject-to-subject raising construction. One limitation of these constructions is that the vP in them is traditionally assumed to be non-phasal, hence not a target for successive-cyclic movement (Chomsky 2000, 2001). As a result, whether these structures are appropriate test cases is fairly theory-dependent.

While the discussion so far has been couched within the framework of phase theory, the central conclusions have implications for other models of cyclicity as well. For example, the results here are straightforwardly compatible with the classical notion of subjacency (Chomsky 1973, 1977), according to which extraction out of CP must use Spec,CP as an escape hatch, but they are not equally compatible with more recent incarnations of subjacency that require adjunction to VP as well (e.g., Chomsky 1986). Furthermore, the results also seem to conflict with theories that adopt a very fine-grained notion of cyclicity, which requires an intermediate landing site at every projection that is crossed by movement (Manzini 1994, Takahashi 1994, Epstein & Seely 2002, Boeckx 2003, Müller 2010). The reason is that, all else equal, such an account predicts that filler retrieval is easier in the TP structure than in the DP structure, because the movement dependency crosses more projections in the TP structure, which hence requires more intermediate landing sites than in the DP structure. This would again be the opposite of what the results here suggest.

The processing evidence thus lines up closely with the substantial body of evidence from traditional diagnostics for successive-cyclic movement through Spec,CP, an encouraging result that expands the empirical evidence for successive cyclicity. On the other hand, the experiments reported here failed to detect any evidence for an analogous intermediate landing site in Spec,vP. This is overall consistent with the relative sparsity of direct evidence in favor of such landing sites, but it does seem to conflict with the existing evidence that uses traditional diagnostics in support of a landing sites in Spec,vP (see Abels 2012, Citko 2014, Georgi 2014, and van Urk 2016 for an overview of relevant arguments). I will not attempt to resolve this interesting paradox here. One possible response might be that it is the difference in the empirical domain and methodology that underlies this discrepancy. As discussed in section 4.3 above, it is a priori possible that CPs and vPs do host intermediate landing sites, which traditional tests pick up on, but that for some reason that is presently unclear, only intermediate landing sites in CP have an effect on filler retrieval. The main challenge for such an account is to provide independent evidence for reactivation mismatches of this type and to reconcile them with general principles of sentence parsing (see the discussion on p. 36–37).[31] Alternatively, it is possible that vP is not a phase, as I have suggested here, and that the traditional arguments in its favor can be explained in some other manner. The challenge for such an account is to demonstrate that this is in fact possible. I will leave an attempt to adjudicate between these and other possibilities for future work.

7. Conclusion and outlook

This paper has brought a novel type of empirical evidence to bear on the question of syntactic locality. The results presented here reveal an asymmetry between CPs and vPs with respect to filler retrieval: they indicate that a filler is reactivated if movement crosses a CP, but that no comparable reactivation

[31] Additionally, Keine (2017) argues that CP is a phase but vP is not based on locality restrictions on φ-agreement.

takes place if movement crosses a vP. In line with the logic in Gibson & Warren (2004), I have shown that this finding can be given a principled explanation if movement creates intermediate touchdowns in CPs but not in vPs, hence if CPs are phases and vPs are not. If only CPs contain intermediate landing sites, it immediately follows that only the crossing of CPs has the effect of reactivating the filler and thereby speeding up processing at the gap site. Furthermore, the experiments provide evidence against clause-level cyclicity in ECM infinitives, which lack a CP projection.

One speculative question that now emerges is *why* phases should be larger than they are usually taken to be. Although the results reported here do not directly inform this question, it is instructive to consider the claim in Chomsky (2000, 2001, 2005), who contends that the grammar cyclically transfers built syntactic structure in order to optimize computational efficiency (also see Berwick & Weinberg 1984 for a related proposal). Against this background, having *too many* phases might in fact incur an *increase* in the computational load in the presence of a movement dependency, because each transition from one phase to another requires an additional intermediate movement step (also see the related discussion about the relationship between the Preliminary Phrase Packager and the Sentence Structure Supervisor in Frazier & Fodor 1978). It is hence conceivable that a larger, C-based phase system combines the advantages of cyclic Spell-Out with a relatively small number of intermediate landing sites necessary to construct a movement dependency.

A broader conclusion of the present paper is that psycholinguistic investigations into real-time processing makes available a new domain in which the distribution of intermediate landing sites can be explored, and that such investigations contribute to our understanding of syntactic locality, in particular phasehood. As noted in section 6, there appears to be a conflict between the processing evidence here (which suggests that vP is not a phase) and more traditional diagnostics (which are usually taken to support CP and vP phases), raising further questions about the distribution of successive cyclicity. Theoretical hypotheses are therefore best addressed using a variety of different types of evidence, with a particular eye towards where they appear to conflict. Direct engagement with processing evidence contributes towards this goal.

References

Abels, K.. 2012. *Phases: An Essay on Cyclicity in Syntax*. de Gruyter, Berlin.

Aldridge, E.. 2008. Phase-based account of extraction in Indonesian. *Lingua* 118, pp. 1440–1469.

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

Barss, A.. 1986. *Chains and Anaphoric Dependence: On Reconstruction and its Implications*. Ph.D. thesis, MIT, Cambridge, MA.

Bartek, B., R. Lewis, S. Vasishth & M. Smith. 2011. In search of on-line locality effects in sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37, pp. 1178–1198.

Bates, D., M. Maechler, B. Bolker & S. Walker. 2014. lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1-7.

Berwick, R. & A. Weinberg. 1984. *The Grammatical Basis of Linguistic Performance: Language Use and Acquisition*. MIT Press, Cambridge, MA.

Boeckx, C.. 2003. *Islands and Chains: Resumption as Stranding*. Benjamins, Amsterdam.

Boeckx, C. & K. Grohmann. 2007. Putting phases in perspective. *Syntax* 10, pp. 204–222.

Boland, J., M. Tanenhaus, S. Garnsey & G. Carlson. 1995. Verb argument structure in parsing and interpretation. *Journal of Memory and Language* 34, pp. 774–806.

Bošković, Ž.. 2002. A-movement and the EPP. *Syntax* 5, pp. 167–218.

Bošković, Ž.. 2014. Now I'm a phase, now I'm not a phase: On the variability of phases with extraction and ellipsis. *Linguistic Inquiry* 45, pp. 27–89.

Bruening, B.. 2006. Differences between the wh-scope-marking and wh-copy constructions in Passamaquoddy. *Linguistic Inquiry* 37, pp. 25–49.

Chen, E., E. Gibson & F. Wolf. 2005. Online syntactic storage costs in sentence comprehension. *Journal of Memory and Language* 52, pp. 144–169.

Chomsky, N.. 1973. Conditions on transformations. Anderson, S. & P. Kiparsky (eds.), *A Festschrift for Morris Halle*, Academic Press, New York, pp. 232–286.

Chomsky, N.. 1977. On wh-movement. Culicover, P., T. Wasow & A. Akmajian (eds.), *Formal Syntax*, Academic Press, New York, pp. 71–132.

Chomsky, N.. 1986. *Barriers*. MIT Press, Cambridge, MA.

Chomsky, N.. 2000. Minimalist inquiries: The framework. Martin, R., D. Michaels & J. Uriagereka (eds.), *Step by Step: Essays in Syntax in Honor of Howard Lasnik*, MIT Press, Cambridge, MA, pp. 89–155.

Chomsky, N.. 2001. Derivation by phase. Kenstowicz, M. (ed.), *Ken Hale: A Life in Language*, MIT Press, Cambridge, MA, pp. 1–52.

Chomsky, N.. 2005. Three factors in language design. *Linguistic Inquiry* 36, pp. 1–22.

Citko, B.. 2014. *Phase Theory: An Introduction*. Cambridge University Press, Cambridge, UK.

Dayal, V.. 2017. Does Hindi-Urdu have feature-driven *wh*-movement to Spec,vP? *Linguistic Inquiry* 48, pp. 159–172.

den Dikken, M.. 2006. A reappraisal of *v*P being phasal: A reply to Legate. Ms., CUNY.

Dillon, B., W.-Y. Chow, M. Wagers, T. Guo, F. Liu & C. Phillips. 2014. The structure-sensitivity of memory access: Evidence from Mandarin Chinese. *Frontiers in Psychology* 5. Article 1025.

Drummond, A.. 2013. Ibex Farm (version 0.3) [software]. (http://spellout.net/ibexfarm/).

Enochson, K. & J. Culbertson. 2015. Collecting psycholinguistic response time data using Amazon Mechanical Turk. *PLoS ONE* 10, p. e0116946. doi: 10.1371/journal.pone.0116946.

Epstein, S. D. & T. D. Seely. 2002. Rule applications as cycles in a level-free syntax. Epstein, S. D. & T. D. Seely (eds.), *Derivation and Explanation in the Minimalist Program*, Blackwell, Oxford, pp. 65–89.

Epstein, S. D. & T. D. Seely. 2006. *Derivations in Minimalism: Exploring the Elimininination of A-Chains and the EPP*. Cambridge University Press, Cambridge.

Ferreira, F. & C. Clifton. 1986. The independence of syntactic processing. *Journal of Memory and Language* 25, pp. 348–368.

Fiebach, C., M. Schlesewsky & A. Friederici. 2002. Separating syntactic memory costs and syntactic integration costs during parsing: The processing of German wh-questions. *Journal of Memory and Language* 47, pp. 250–272.

Fodor, J. D. & F. Ferreira (eds.). 1998. *Reanalysis in Sentence Processing*. Kluwer, Dordrecht.

Fodor, J. D. & A. Inoue. 1994. The diagnosis and cure of garden paths. *Journal of Psycholinguistic Research* 23, p. 407–434.

Fox, D.. 1999. Reconstruction, Binding Theory, and the interpretation of chains. *Linguistic Inquiry* 30, pp. 157–196.

Frazier, L. & C. Clifton. 1989. Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes* 4, pp. 93–126.

Frazier, L. & J. D. Fodor. 1978. The sausage machine: A new two-stage parsing model. *Cognition* 6, pp. 291–325.

Georgi, D.. 2014. *Opaque Interactions of Merge and Agree: On the Nature and Order of Elementary Operations*. Ph.D. thesis, Universität Leipzig.

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

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

Gibson, E. & T. Warren. 2004. Reading-time evidence for intermediate linguistic structure in long-distance dependencies. *Syntax* 7, pp. 55–78.

Gordon, P., R. Hendrick & M. Johnson. 2001. Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27, pp. 1411–1423.

Grodner, D. & E. Gibson. 2005. Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive Science* 29, pp. 261–291.

Grodner, D., E. Gibson & S. Tunstall. 2002. Syntactic complexity in ambiguity resolution. *Journal of Memory and Language* 46, pp. 267–295.

Hawkins, J.. 1994. *A Performance Theory of Order and Constituency*. Cambridge University Press, Cambridge.

Henry, A.. 1995. *Belfast English and Standard English: Dialect Variation and Parameter Setting*. Oxford University Press, Oxford.

Horvath, J.. 1997. The status of wh-expletives and the partial wh-movement construction of Hungarian. *Natural Language and Linguistic Theory* 15, pp. 509–572.

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

Just, M. A., P. Carpenter & J. Woolley. 1982. Paradigms and processes in reading comprehension. *Journal of Experimental Psychology* 111, pp. 228–238.

Kayne, R. & J. Y. Pollock. 1978. Stylistic inversion, successive cyclicity, and Move NP in French. *Linguistic Inquiry* 9, pp. 595–621.

Keine, S.. 2017. Agreement and *v*P phases. LaCara, N., K. Moulton & A.-M. Tessier (eds.), *A Schrift to*

Fest Kyle Johnson, Linguistics Open Access Publications 1, University of Massachusetts Amherst, pp. 177–185.

King, J. & M. A. Just. 1991. Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language* 30, pp. 580–602.

Kuznetsova, A., P. B. Brockhoff & R. H. B. Christensen. 2014. *lmerTest: Tests for Random and Fixed Effects for Linear Mixed Effect Models (lmer Objects of lme4 Package)*. R package version 2.0-11.

Landau, I.. 2004. The scale of finiteness and the calculus of control. *Natural Language and Linguistic Theory* 22, pp. 811–877.

Lebeaux, D.. 1988. *Language Acquisition and the Form of the Grammar*. Ph.D. thesis, University of Massachusetts, Amherst, MA.

Legate, J. A.. 2003. Some interface properties of the phase. *Linguistic Inquiry* 34, pp. 506–516.

Lewis, R.. 1996. Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research* 25, pp. 93–115.

Lewis, R. & S. Vasishth. 2005. An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science* 29, pp. 1–45.

Lewis, R., S. Vasishth & J. Van Dyke. 2006. Computational principles of working memory in sentence comprehension. *Trends in Cognitive Science* 10, pp. 447–454.

Manzini, R.. 1994. Locality, Minimalism, and parasitic gaps. *Linguistic Inquiry* 25, pp. 481–508.

Marinis, T., L. Roberts, C. Felser & H. Clahsen. 2005. Gaps in second language sentence processing. *Studies in Second Language Acquisition* 27, pp. 53–78.

Matushansky, O.. 2005. Going through a phase. McGinnis, M. & N. Richards (eds.), *Perspectives on Phases*, MITWPL, Cambridge, MA, vol. 49 of *MIT Working Papers in Linguistics*, pp. 157–181.

McCloskey, J.. 2000. Quantifier float and *wh*-movement in an Irish English. *Linguistic Inquiry* 31, pp. 57–84.

McCloskey, J.. 2002. Resumptives, successive cyclicity, and the locality of operations. Epstein, S. D. & T. D. Seely (eds.), *Derivation and Explanation in the Minimalist Program*, Blackwell, Oxford, pp. 184–226.

McDaniel, D.. 1989. Partial and multiple wh-movement. *Natural Language and Linguistic Theory* 7, pp. 565–604.

McElree, B. & T. Griffith. 1998. Structural and lexical constraints on filling gaps during sentence comprehension: A time-course analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24, pp. 432–460.

McElree, B., S. Foraker & L. Dyer. 2003. Memory structures that subserve sentence comprehension. *Journal of Memory and Language* 48, pp. 67–91.

Müller, G.. 2010. On deriving CED effects from the PIC. *Linguistic Inquiry* 41, pp. 35–82.

Nicol, J. & D. Swinney. 1989. The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research* 18, pp. 5–19.

Nicol, J., J. Fodor & D. Swinney. 1994. Using cross-modal lexical decision tasks to investigate sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20, pp. 1229–1238.

Omaki, A., E. Lau, I. D. White, M. L. Dakan & C. Phillips. 2015. Hyper-active gap filling. *Frontiers in Psychology* 6. Article 384.

Phillips, C.. 2006. The real-time status of island phenomena. *Language* 82, pp. 795–823.

Phillips, C., N. Kazanina & S. Abada. 2005. ERP effects of the processing and syntactic long-distance dependencies. *Cognitive Brain Research* 22, pp. 407–428.

Pickering, M. & M. Traxler. 2003. Evidence against the use of subcategorization frequencies in the processing of unbounded dependencies. *Language and Cognitive Processes* 18, pp. 469–503.

R Core Team. 2014. R: A language and environment for statistical computing.

Rezac, M.. 2013. Case and licensing: Evidence from ECM + DOC. *Linguistic Inquiry* 44, pp. 299–319.

Schulte im Walde, S.. 1998. Automatic semantic classification of verbs according to their alternation behavior. MA thesis, Universität Stuttgart.

Staub, A.. 2007. The parser doesn't ignore intransitivity, after all. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 33, pp. 550–569.

Staub, A.. 2010. Eye movement and processing difficulty in object relative clauses. *Cognition* 116, pp. 71–86.

Staub, A., C. Clifton & L. Frazier. 2006. Heavy NP shift is the parser's last resort: Evidence from eye movements. *Journal of Memory and Language* 54, pp. 389–406.

Stevenson, S.. 1994. Competition and recency in a hybrid network model of syntactic disambiguation. *Journal of Psycholinguistic Research* 23, pp. 295–322.

Stowe, L.. 1986. Parsing wh-constructions: Evidence for on-line gap location. *Language and Cognitive Processes* 1, pp. 227–245.

Sussman, R. & J. Sedivy. 2003. The time course of processing syntactic dependencies. *Language and*

Cognitive Processes 18, pp. 143–163.

Takahashi, D.. 1994. *Minimality of Movement*. Ph.D. thesis, University of Connecticut, Storrs.

Torrego, E.. 1984. On inversion in Spanish and some of its effects. *Linguistic Inquiry* 15, pp. 103–129.

Traxler, M. & M. Pickering. 1996. Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language* 35, pp. 454–475.

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

Trueswell, J., M. Tanenhaus & C. Kello. 1993. Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19, pp. 528–553.

van Urk, C.. 2016. On the distribution of reflexes of successive cyclicity. Handout of talk at the Brussels Conference on Generative Linguistics (BCGL) 9, available at http://webspace.qmul.ac.uk/cvanurk/PhaseTheoryTalk.pdf.

van Urk, C. & N. Richards. 2015. Two components of long-distance extraction: Successive cyclicity in Dinka. *Linguistic Inquiry* 46, pp. 113–155.

Van Dyke, J. & R. Lewis. 2003. Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language* 49, pp. 285–316.

Vasishth, S. & R. Lewis. 2006. Argument–head distance and processing complexity: Explaining both locality and antilocality effects. *Language* 82, pp. 767–794.

Wagers, M. & C. Phillips. 2009. Multiple dependencies and the role of the grammar in real-time comprehension. *Journal of Linguistics* 45, pp. 395–433.

Wagers, M. & C. Phillips. 2013. Going the distance: Memory and control processes in active dependency construction. *The Quarterly Journal of Experimental Psychology* pp. 1–31.

Wanner, E. & M. Maratsos. 1978. An ATN approach to comprehension. Halle, M., J. Bresnan & G. Miller (eds.), *Linguistic Theory and Psychological Reality*, MIT Press, Cambridge, MA, pp. 119–161.

Warren, T. & E. Gibson. 2002. The influence of referential processing on sentence complexity. *Cognition* 85, pp. 79–112.