Locality domains in syntax: Evidence from sentence processing

It is widely assumed 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 the intermediate traces created by successive-cyclic movement. It furthermore uses this parsing evidence to investigate the distribution of these intermediate traces. The central findings of this paper are (i) that successive-cyclic movement targets the edge of finite clauses and (ii) that no intermediate landing sites exist at the edges of (extended) VPs. These conclusions are fully consistent with traditional characterizations of successive-cyclicity but at variance with more recent approaches that hold that even movement within a single clause is successive-cyclic. A reassessment of arguments put forth in support of the latter hypothesis reveals them to be less compelling than they are standardly taken to be, which provides converging evidence for the parsing conclusions reached here. More generally, because successive cyclicity is the result of general syntactic locality domains, this paper provides psycholinguistic evidence for such locality domains and for their distribution.

1 Introduction

One of the central discoveries of 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, despite appearance, long movement consists of a sequence of smaller movement steps (Chomsky 1973, 1977). Movement out of a finite clause provides an example: Chomsky (1973, 1977, 1981) proposes that any element leaving such a clause must first land in the highest specifer 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. If more than one finite clause boundary is crossed, this process applies iteratively. The apparent unboundedness of displacement dependencies under this view merely reflects the unbounded number of short movement steps that such sequences may comprise. Consequently, in configurations such as (1a) the dependency between the moved element *who* and its trace is mediated by an intermediate trace in Spec,CP of the embedded clause, as shown in (1b).

(1) a. Who$_i$ does John think that Mary married t_i in secret?

b. [$_{\text{CP}}$ Who$_i$ does [$_{\text{TP}}$ John think [$_{\text{CP}}$ t_i that [$_{\text{TP}}$ Mary married t_i in secret]]]]

A classic argument that crossclausal movement crucially implicates the embedded Spec,CP position comes from *wh*-islands (Ross 1967). As is well-known, *wh*-movement out of a clause in which another element has itself undergone *wh*-movement is impossible, as exemplified in (2). If the crossclausal movement of *what* directly related its surface position to its base position, the fact that the intermediate Spec,CP is occupied by *who* should not impact the movement. That it does provides evidence that the movement of *what* requires an intermediate landing site in the lower Spec,CP.

(2) *[$_{\text{CP}}$ What$_i$ did [$_{\text{TP}}$ Susan ask [$_{\text{CP}}$ who$_j$ [$_{\text{TP}}$ t_j ate t_i]]]]

The view that long movement dependencies consist of a sequence of strictly local sub-dependencies has been adopted virtually universally in the generative tradition and a considerable amount of additional empirical evidence has been accumulated. Perhaps the most striking support for successive cyclicity comes from languages that exhibit morphological reflexes of movement. A well-known example is Irish, discussed in great detail by McCloskey (1979, 2002). In Irish the form of a complementizer depends on whether or not its specifier is occupied. If it is not, the complementizer *go* is used. If it is, the complementizer surfaces as *aL*. Strikingly, *all* complementizers along the movement path have to appear in the *aL* form:

(3) an t-ainm [$_{\text{CP}}$ <u>aL</u> hinnseadh dúinn [$_{\text{CP}}$ <u>aL</u> bhí ___ ar an áit]]
the name COMP was-told to-us COMP was on the place
'the name that we were told was on the place'

That the embedded complementizer takes the *aL* form demonstrates that its specifier is filled by movement over it. This follows straightforwardly if movement has to proceed through the specifier of each CP it crosses.

There exists now a diverse body of evidence in favor of successive cyclicity through Spec,CP.

This evidence includes *wh*-copying constructions in German (Müller 1999), Romani (McDaniel 1989), Frisan and Afrikaans (Felser 2004), Hungarian (Horvath 1997), and Ancash Quechua (Cole 1982) and Passamquoddy (Bruening 2006), as well as reconstruction to intermediate trace positions (Barss 1986, Lebeaux 1988), *wh*-agreement in Chamorro (Dukes 1992, Chung 1994, 1998, Kaplan 2005), quantifier floating in West Ulster English (McCloskey 2000), verb movement in Belfast English (Henry 1995), French (Kayne & Pollock 1978) and Iberian Spanish (Torrego 1984) and tonal downstep in Kikuyu (Clements & Ford 1979, Clements 1994). The general notion that displacement dependencies are strictly local is also remarkable in that it has been widely adopted in a range of otherwise quite different syntactic frameworks, sometimes in much more radical forms, such as in Generalized Phrase Structure Grammar (Gazdar et al. 1985) and Head-Driven Phrase Structure Grammar (Pollard & Sag 1994).

In recent developments in syntax, the role played by successive cyclicity has become even more pervasive. Chomsky (2000, 2001) proposes that successive-cyclic movement is a direct consequence of the strictly cyclic nature of syntactic derivations. On this view, built syntactic structure is periodically spelled out and thereby rendered inaccessible for outside operations. Chomsky (2000, 2001) refers to such cyclic Spell-Out domains as *phases.* Crucially, 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. As a general consequence, movement past a phase head must proceed successive-cyclically through the specifier of that phase head in order to remain accessible.

In addition to this conceptual reassessment of the underlying cause of successive cyclicity, Chomsky (2000, 2001) proposes that both C as well as (transitive) v are phase heads. By the above logic, movement past either of these nodes must be successive-cyclic. In other words, successive cyclicity must target both Spec,CP as well as Spec,vP (also see Chomsky 1986, 2004, 2008). In other words, it is not only cross-clausal movement that is successive-cyclic but also intra-clausal movement. The view that there are intraclausal landing sites in addition to Spec,CP has gained considerable popularity in recent syntactic theorizing and a number of empirical arguments have been made in its favor (see, e.g., Fox 1999, Legate 2003, Rackowski & Richards 2005, Aldridge 2008, van Urk & Richards 2015) although the question remains controversial.

Because successive cyclicity is a direct consequence of the phase-based nature of syntactic derivations, investigations into the distribution of successive cyclicity directly inform the distribution of phases in general: If there is evidence that movement past a projection XP does not have to pass through the specifier of XP, this constitutes direct evidence that XP is not a phase.

Against this background, the present paper makes two main claims, based on two self-paced reading experiments. First, it provides evidence that the effects of successive-cyclic movement may be observed in online sentence processing, thus supporting the view that successive cyclicity and the concept of phases underlying it are 'cognitively real' in the sense that the structures created by the sentence parser obey it (Experiment 1). This finding is in line with current research directions that aim at reducing properties of narrow syntax to general cognitive constraints (Chomsky 2005). Second, the paper then uses sentence processing evidence to investigate the distribution of successive cyclicity. Concretely, it investigates whether successive-cyclic movement targets only the specifier of a CP or also the specifier of a verbal projection (Experiment 2). The processing evidence reported here supports the classical view that only CPs host intermediate landing sites. By the logic above, these results provide evidence that C is a phase head but v is not. Because aspects of sentence parsing have not been brought to bear on theoretical questions about the nature of successive cyclicity, the paper contributes an entirely new *type* of evidence to the discussion in the literature.

The link between the grammatical notion of successive cyclicity and evidence from language processing is rather direct: Although phases are usually thought of in derivational terms, successive cyclicity can be equally well conceptualized representationally: If a domain D requires successive-cyclic movement through its edge, no extraction dependency can be established between an element outside of D and another element embedded in D more deeply than its edge. Any such dependency requires the existence of an intermediate trace at D's edge. For the parser to establish a grammatically licit syntactic structure, it has to postulate this intermediate trace. Evidence showing whether or not the parser postulates such a trace in a given domain edge thus constitutes evidence whether this domain requires successive-cyclic movement or not.

Before moving on, a few notes on the terminology in this paper are in order. First, there is some disagreement in the syntactic literature as to the exact identity of the phasal head within the verbal projection. While it is standardly taken to be v, some authors have proposed that it is Voice or some related head. Because the identity of this head is irrelevant for the argumentation in this paper, the abbreviatory label 'VP' will be used here to refer to whatever the phasal projection within the extended verb phrase is taken to be. Thus, the label 'VP' stands for the extended verbal projection, potentially comprising functional projections in addition to the core VP. Second, there are some terminological divergences between the syntactic and the processing literature. 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.

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 is amenable to alternative explanations. Experiment 1 in section 3 contrasts an account in terms of successive cyclic movement with alternative hypothesis and provides evidence for successive-cyclic movement through Spec,CP. Against this background, Experiment 2 in section 4 then investigates whether VPs pattern with CPs in hosting intermediate landing sites, as predicted by recent phase-based approaches. The results of Experiment 2 indicate that this is not the case. Section 5 discusses the central conclusions that can be drawn from the evidence presented here and additional consequences beyond the question of successive cyclicity. Finally, section 6 concludes.

2 Successive cyclicity in the parser

Despite the immense impact of successive cyclicity on theoretical syntax, there are very few direct investigations into the existence of successive-cyclic movement 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-cyclic movement, they may also be attributed to independently motivated properties of the parser that are unrelated to successive-cyclic movement.

This section will review previous results and discuss their limitations. These limitations 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 remains active across clause boundaries. In addition, the results of two end-of-sentence acceptability rating experiments indicate that cross-clausal movement dependencies are harder than monoclausal ones, an effect also observed by Phillips et al. (2005). 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 cross-clausal dependencies increases the demand such dependencies pose onto the parser.

While these interpretations are consistent with the observed results, they are not necessary. First, that a filler is held active across a clause boundary does not, in and of itself, require that filler to be linked to an intermediate trace at the edge of the lower clause. Second, that movement over a clause boundary incurs 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 attempt to identify processing reflexes of successive-cyclic movement is Gibson & Warren (2004) (henceforth G&W). The rationale underlying G&W's experiment is the following: There is independent evidence that the distance between the moved element and its trace is positively 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 moved element and its trace, the harder it is to integrate that trace (King & Just 1991, Gibson 1998, 2000, Gordon et al. 2001, Warren & Gibson 2002, Grodner & Gibson 2005, Lewis & Vasishth 2005, Vasishth & Lewis 2006, Staub 2010, Bartek et al. 2011). 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 trace is greater in object than in subject relative clauses (e.g., Just & Carpenter 1992). That the distance between the moved element and its trace should affect processing speed at the trace position is commonly assumed to follow from the fact that the semantic and morpho-syntactic features of the moved element have to be retrieved in order to successfully construe and interpret the trace. 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 insubstantial for the present purpose.[1]

Against this general background, G&W's investigated the processing of structures like the one in (4).

[1] 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).

(4) a. *CP condition*
   i. *Movement*
   The manager *who*$_i$ the consultant claimed $\boxed{t_i}$ that the new proposal has pleased t_i will hire five workers tomorrow.
   ii. *Control*
   The consultant claimed that the new proposal had pleased the new manager who will hire five workers tomorrow.

  b. *DP condition*
   i. *Movement*
   The manager *who*$_i$ the consultant's claim about the new proposal had pleased t_i will hire five workers tomorrow.
   ii. *Control*
   The consultant's claim about the new proposal had pleased the manager who will hire five workers tomorrow.

In (4a.i) the relative pronoun *who* is moved across a finite clause boundary; in (4b.i) the same element is moved over a complex subject but this movement does not cross a clause boundary. The sentences in (b) 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 trace of the moved element – *pleased* in (4). To obtain a measure of the difficulty of integrating the trace, G&W compared the reading times at *pleased* in the movement condition in (i) to the reading times in the 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 integrating the trace.

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

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 reliably smaller in the CP condition (4a) than in the DP condition (4b). G&W take this result to support the existence of an intermediate landing site in Spec,CP. This effect has subsequently been replicated by Marinis et al. (2005).[2]

[2] Marinis et al. (2005) also tested L2 speakers of English (L1: Chinese, Japanese, German or Greek) in addition to

2.2 *Limitations*

While suggestive, G&W's findings do not provide unambiguous evidence for successive-cyclic movement. This is because 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 the design of G&W's experiment is that the movement dependency crosses a verb (*claim* in (4a)) in the CP condition (4a.i) but not in the DP condition (4b.i). One very general concern is that the parser might have initially postulated the trace of *who* as the object of the higher verb *claim* in the CP condition. 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 (and, as will be shown shortly, plausible) that the parser construes *who* as the object of the verb *claim* in its initial parse of this region. It is then furthermore plausible that this intermediate, though incorrect, reactivation of *who* facilitates its retrieval when its actual trace 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 is entirely due to the particular mechanisms and decision procedures underlying the parser and have nothing to do with successive-cyclic movement. Rather, they reflect the effect of incorrect structural analyses temporarily pursued by the parser on the construction of a subsequent, and correct, representation. This alternative account will subsequently be referred to 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 (also see de Vincenzi 1991's *Minimal Chain Principle*):

(5) 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 trace at the earliest grammatically licit position. Crucially, it does so even in the absence of direct evidence from the input that a trace is present. As a result, the parser may in some cases postulate a trace 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.[3]

(*fn. continued*)
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] 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 electroencephalography (Garnsey et al. 1989, Kaan et al. 2000, Phillips et al. 2005), 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).

With regard to the structures in (4), 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 likely that the parser initially postulates the trace of *who* as the object of this higher verb. In the DP condition, on the other hand, the complex subject is an island. In light of evidence that the parser does not postulate traces in islands (Stowe 1986, Traxler & Pickering 1996, McElree & Griffith 1998), it is expected that the parser does not construe a premature trace in the DP structure. The ultimately incorrect but temporarily entertained trace in the CP condition would then require a reactivation of *who* and thereby facilitate its subsequent retrieval at the actual site of the trace.

G&W briefly address this concern (p. 64), noting that the verbs they used were strongly biased towards a clausal complement and, to the extent that these verbs were compatible with a nominal object, required an inanimate one. 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 trace of the moved element at the higher verb.

It is not clear that this rebuttal adequately addresses the objection, however. First, seven of the nine verbs used by G&W are similar to *claim* in that they productively take nominal objects.[4] Of these seven verbs, at least four are compatible with animate objects.[5] In terms of the 15 items used in their experiment, twelve contained a verb that takes a nominal object and seven of those licitly combined with an animate object. In other words, nearly half of their material contained verbs that were syntactically and semantically compatible with a parse that incorrectly associates the filler with the higher verb. While such a parse might still be dispreferred by frequency and real world plausibility, it is clear that there were not, in fact, pervasive subcategorization constraints against an early gap parse.

Apart from these particular concerns regarding the materials used by G&W, it is quite controversial if considerations of frequency, animacy and real world plausibility can completely suspend the parser's active filler strategy and thus prevent the parser from postulating an otherwise preferred gap.

First, while subcategorization constraints have been found to affect initial parsing decisions (e.g., Staub 2007), it is less clear whether the frequency of a subcategorization frame has the same effect. The frequency subcategorization frame has been argued not to affect whether a gap is postulated as the argument of 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).

Second, it is likewise doubtful whether animacy and plausibility constraints have enough of an impact on initial parsing decisions to prevent the parser from construing an otherwise licit trace. Pickering & Traxler (2003), for example, crossed optionally transitive verbs that occur more or less frequently with a direct object with a filler that is either semantically compatible with the verb or not, and observed elevated reading times on a verb if the moved element was an implausible object

[4] The verbs used in G&W's materials are *predict, claim, conclude, imply, confirm, realize, state, think* and *hypothesize*. Of these, only the last two do not productively combine with nominal objects.

[5] A cursory search on Google reveals that at least the following verbs may take animate objects: *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]).

of that verb, indicating that a trace was postulated even if plausibility considerations disfavor such a choice. Moreover, these elevated reading times arose regardless of the verb's subcategorization frequency, thus corroborating Staub (2007)'s findings.

In sum, it is plausible in light of previous evidence that the parser posulates a trace of the moved element when it encounters the higher verb. G&W's results leave it open whether the facilitation at the ultimate gap site in the CP condition is due to successive-cyclic movement or this premature gap filling. To distinguish between these competing explanations, Experiment 1 investigates the role of the higher verb by systematically manipulating its subcategorization restrictions. The results favor the successive cyclicity account over the premature gap filling one.

3 Experiment 1

To distinguish whether the effect observed by Gibson & Warren (2004) is due to the integration of the filler into the intermediate Spec,CP – i.e., successive cyclicity – or to a temporary incorrect parse in which the filler is construed as the object argument of the higher verb – i.e., premature gap filling –, Experiment 1 manipulates the subcategorization frame of the higher verb. As reviewed above, there is evidence that the parser postulates an object trace only for verbs whose subcategorization requirement allows the verb to take a DP object (Staub 2007). Under the premature gap filling account, the moved element should hence only be reactivated at the higher verb if this verb can in principle take an object DP. Consequently, the premature gap filling hypothesis predicts that the G&W effect should be modulated by properties of the intervening verb. By contrast, if G&W's effect is the result of successive cyclicity, the reactivation of the filler is independent of the higher verb and the verb type manipulation should hence not have an impact on the effect.

3.1 *Method*

3.1.1 *Materials*

Design
Fifty-six sets of sentences like the one in (6) 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 Section A.1 of 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. [–*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 nominal direct objects (like, e.g., *claim*). Verbs in this class 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., *think*). 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).[6] 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 (6)). The rationale for including this context was to decrease the processing and integration 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. Answer options were presented in random order.

(6) *Context:* Groundless allegations really could hurt people in our company.
  a. *CP intervener*
    (i) [+move]
    The secretary *who*$_i$ the lawyer $\left\{\begin{array}{ll} \textit{CP/DP:} & \text{claimed} \\ \textit{CP:} & \text{boasted} \end{array}\right\}$ that the accusations 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 intervener*
    (i) [+move]
    The secretary *who*$_i$ 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

[6] 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 a 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. (1993)'s sentence completion experiment was 0%. Five additional verbs (*argue, comment, hypothesize, speculate, theorize*) were included to decrease the amount repetition in the stimuli. Overall, these CP-verbs had a 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), which is 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.

Fillers

In addition to the 48 target items, 48 additional items were created that matched the target sentences in syntactic complexity and length. 24 of these were part of unrelated experiments and the additional 24 were a haphazard collection of sentences.

3.1.2 *Participants*

The experiment reported here involved 130 participants, recruited via Amazon Mechanical Turk. All were native speakers of American English and naïve to the purpose of the experiment. Each received a compensation of USD 1.

3.1.3 *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).[7] The regioning of the target sentences followed the general schema in Table 1.[8] At the beginning of each trial participants saw the theme-setting context sentence, which was displayed 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 forty minutes. Throughout the entire experiment a progress bar was displayed. Items were arranged in a Latin Square design on eight different lists such that each list contained one instance of every item and all eight conditions

---

[7] 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). Moreover, with regard to the two experiments reported here, it is noteworthy that both 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, both Experiment 1 and 2 do replicate these previous findings. It is therefore very unlikely that the effects reported here are merely an artifact of the experimental methodology.

[8] This regioning is identical to the one employed by G&W in their analysis with one difference. In G&W 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

of each item appeared on one list. Participants were randomly 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.

3.1.4 *Analysis*

The data analysis procedures were identical for both 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 fit 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 suggestions 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 subjects whose overall accuracy over the entire experiment (including all target and filler items) was at least 75% were included in the data analysis. Reading time data less then 200 ms or greater than 5000 ms were taken as not reflecting 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, an LME model was fit 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 for a given participant and condition 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 LME 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 Section 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. 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 predictors were sum-coded (Movement: [−*move*] = −.5, [+*move*] = .5; Intervener: *CP* = −.5, *DP* = .5; Verb type: *CP-verb* = −.5, *CP/DP-verb* = .5).

Because one item (item 9 in Section C.1 of the supplementary materials) was regioned inconsistently across conditions, it was excluded from all analyses.[9]

3.2.1 *Comprehension question response accuracy*

The overall answer accuracy over all 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

[9] Eliminating this item did not introduce a plausibility confound (all ps > .5, cf. Section A.1 in the supplementary materials).

Table 2. Mean answer accuracy in Experiment 1

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

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 was significant.

3.2.2 *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 Section B.1 of the supplementary materials.

Table 4 provides an overview of the LME analyses of the log-transformed reading times for the regions that are crucial in evaluating the two contrasting hypotheses. Of central importance are the region containing the trace-hosting verb (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 trace or not, the reading times in the region following this verb (region 3) should be affected by the type of the preceding verb. In the interest of space, Table 4 only reports the analyses for these three regions. A comprehensive analysis of all regions is provided in Section B.1 in the supplementary materials.

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 condition, a additional analysis of the residual reading times was performed, which replicated the two main effects ($p < .001$) and the three-way interaction ($\hat{\beta} = -65$, $t = 2.7$, $p < .01$).

Table 3. Mean raw reading times by condition and region in ms 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 (−38)	775	708	812 (−49)	1288 (68)
		[+mv]	731	1032	598 (24)	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 (119)	866 (−64)
CP/DP	CP	[−mv]	—	896	526 (−48)	765	738	840 (−19)	1279 (58)
		[+mv]	736	1066	631 (57)	790	893	820 (78)	888 (−44)
	DP	[−mv]	—	1018	562 (−21)	824	736	815 (−47)	1219 (−2)
		[+mv]	710	1215	657 (74)	892	903	845 (102)	887 (−44)

Table 4. Coefficient estimates and corresponding t-value for linear mixed effects model analyses of log reading times in crucial 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 Section B.1 of the supplementary materials.

	Region					
	3		5		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

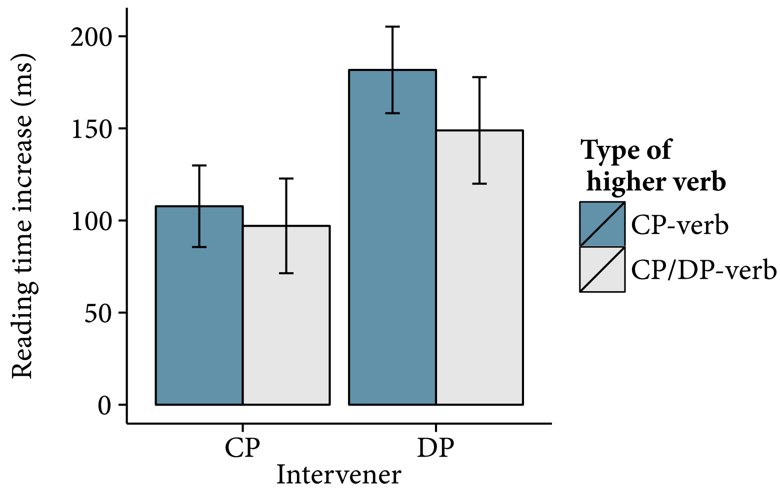
In addition, this analysis indicated an interaction between *movement* and *intervener* ($\hat{\beta}$ = 25, t = 2, $p < .05$), which is, however, uninterpretable given the higher-level interaction.

To further investigate the three-way interaction in region 3, the predictors *movement* and *verb type* were nested under the levels of *intervener* in the LME analysis or log reading times, with the full random effects structure 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* ($\hat{\beta} = .15$, $t = 8.5$, $p < .001$) but neither a main effect of *verb type* ($\hat{\beta} = .001$, $t = 0.05$, $p > .5$) nor an interaction between the two ($\hat{\beta} = -.03$, $t = -1.06$, $p = .3$). For the CP structure, on the other hand, the model revealed a main effect of *movement* ($\hat{\beta} = .09$, $t = 4.9$, $p < .001$), no main effect of *verb type* ($\hat{\beta} = -.01$, $t = -.8$, $p = .4$), and an interaction between both factors ($\hat{\beta} = .05$, $t = 2.15$, $p < .05$) 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 analyis of the residual reading times. It will be considered in greater detail in the discussion section.

In the gap region (region 5), the DP condition was read more slowly than the CP condition and the [+move] condition was 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 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) shows a significant interaction between *intervener* and *movement* such that the DP [+move] structure leads to higher reading times relative to its [−move] control than the CP structure. There is no interaction with verb type. Because the length of this

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



region differed between the [+move] and [–move] conditions and to confirm the reliability of these critical findings, an analysis of the residual reading times was carried out. This model showed a main effect of *movement* ($\hat{\beta} = 132$, $t = 7$, $p < .001$) and an interaction between *movement* and *intervener* ($\hat{\beta} = 62$, $t = 2.4$, $p < .05$). There was no three-way interaction ($\hat{\beta} = -26$, $t = -.6$, $p > .5$). That there was a main effect of movement in only 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 between [–move] and [+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: The difference between [–move] and [+move] structures was greater in DP structures than in CP structures when the position of the trace was encountered. While the region that the effect surfaced in differed (region 6 in the present experiment but region 5 in Gibson & Warren 2004's and Marinis et al. 2005's experiments), the shape of the effect is identical. Because it is common in self-paced reading experiments for an effect to surface in the spillover region, this discrepancy is not surprising.

The central finding of Experiment 1 is that this effect is not modulated by the type of the higher verb. Because the type of the higher verb is crucially implicated in the premature gap filling account but not the successive cyclicity account, this result is consistent with the latter but not the former. It is worth noting that the advantage of the CP condition relative to the DP condition was numerically, but not significantly, greater for CP-verbs than for CP/DP-verbs, the opposite of what the premature gap filling account predicts.

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 times increase between [–move] and [+move] conditions is sensitive to the type of the verb 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 whether the preceding noun was *claim* or *boast.* This pattern is very plausibly a *filled gap effect:* If the verb is of the CP/DP-kind, the moved element is initially construed as its object and the parser incorrectly postulates a trace of the moved element when the higher verb is encountered. The complementizer in the following region makes it clear this parse is incorrect and reanalysis is required. This reanalysis amplifies the reading time increase compared to the non-movement baseline. That this increase is less pronounced if the higher verb is of the CP-type demonstrates that either no incorrect trace is postulated or that it is postulated on fewer trials. No such difference was observed in the DP condition: Regardless of whether the head noun is *claim* or *boast,* no gap can be postulated in either region 2 or 3. As such, no filled gap effect results as the type of the higher noun has no impact on whether or not a trace is postulated. This accounts for the observed three-way interaction. 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 is affected by these frames. Yet crucially, the critical effect in region 6 is entirely independent of these subcategorization frames. This corroborates the conclusion that the effect at the trace-hosting verb cannot be attributed to premature gap filling as it is stable regardless of whether or not premature gap filling takes place.

There was, in addition, an interaction between *intervener* and *verb type* in the gap region such that the reading time difference between CP and 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] structured, it is orthogonal to the crucial effects just discussed. It is plausible 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 their conceptual 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 semantically construed as 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.

That the presence of a movement dependency incurred an increase in reading times in regions prior to the trace is expected under the assumption that an unassigned filler has to be held active in working memory (Wanner & Maratsos 1978, King & Kutas 1995, Gibson 1998, 2000, Fiebach et al. 2002, Grodner et al. 2002, Chen et al. 2005), a requirement that leaves fewer resources available for the processing of incoming material.

Before proceeding, it is worth considering the role of *structural distance* in the results of Experiment 1. There is some indication that the linear distance between the filler and the gap is an insufficient predictor of integration difficulty. Investigating Chinese relative clauses, Kroch (2006) and Lin & Bever (2006) found an advantage for subject relatives over object relatives despite the fact that the object position is linearly closer to the head noun in Chinese. They conclude that structural distance must also have an impact on integration cost. 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 cross-clausal 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 hence necessary even if structural distance is factored in. Moreover, if successive-cyclic movement through Spec,CP takes place, the ultimate trace is equally far away from the closest antecedent position in both condition. Structural distance hence leaves unaffected the conclusion that successive cyclicity takes place in the CP condition but not the DP condition.[10]

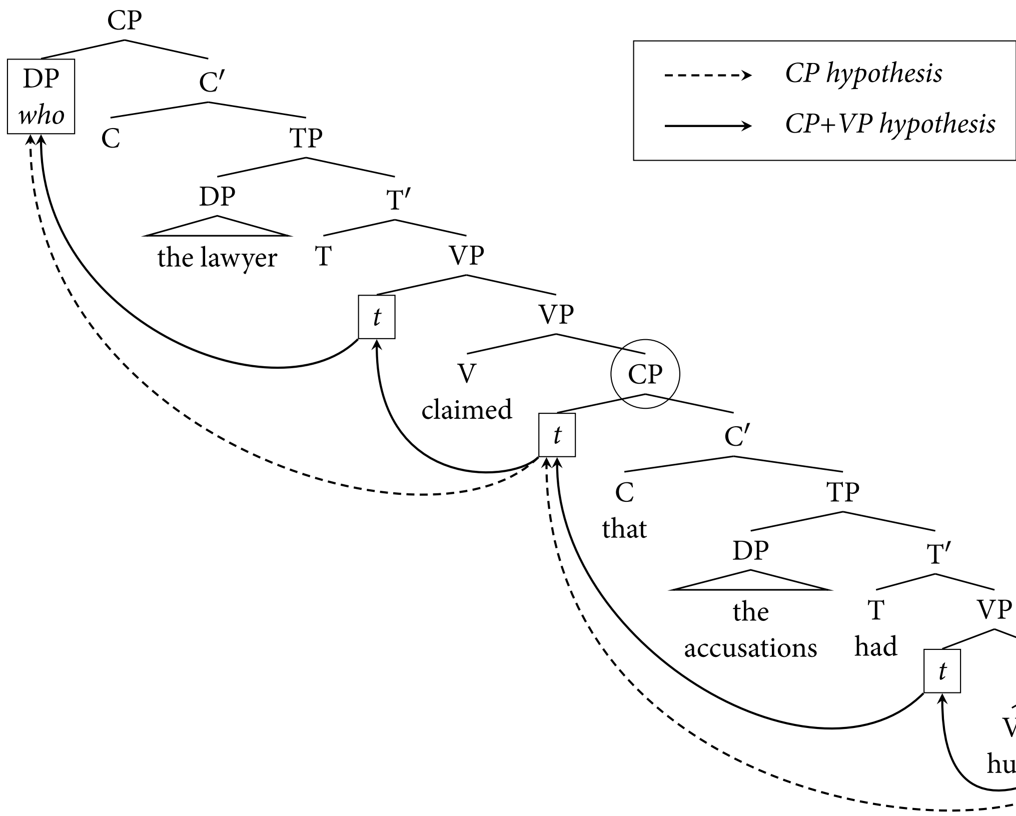
4 Experiment 2

G&W's results as well as the ones of Experiment 1 provide evidence for successive-cyclic movement through Spec,CP. These results do not, however, show whether successive cyclicity is limited to Spec,CP (as has been traditionally assumed) or also targets VP edges (as is standardly assumed in the current literature).[11] Let us refer to the traditional position as the *CP hypothesis* and to its alternative as the *CP+VP hypothesis*. 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 the DP structures. (7) and (8) give the schematic structures and the movement steps

[10] 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 extraction condition but unexpectedly long reading times in the DP extraction 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 6 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 should be positively correlated with the relative difficulty of the DP condition. To assess this prediction, a linear model was devised that regressed the structural difficulty scores against the interaction scores. The model revealed a significant negative correlation ($r = -.3$, *adjusted* $R^2 = .07$, $F(1, 45) = 4.5$, $p < .05$). Thus, the greater the baseline difficulty of the DP structure the *smaller* the critical interaction at region 6 was. This finding directly disconfirms the prediction of the alternative account. This in turn entails that the G&W effect cannot simply be reduced to a non-cumulative combination of the two main effects.

[11] Recall from section 1 above that this paper reserves the label 'VP' for the extended verbal projection, not necessarily its core VP.

(7) *Structure of CP condition in Experiment 1*



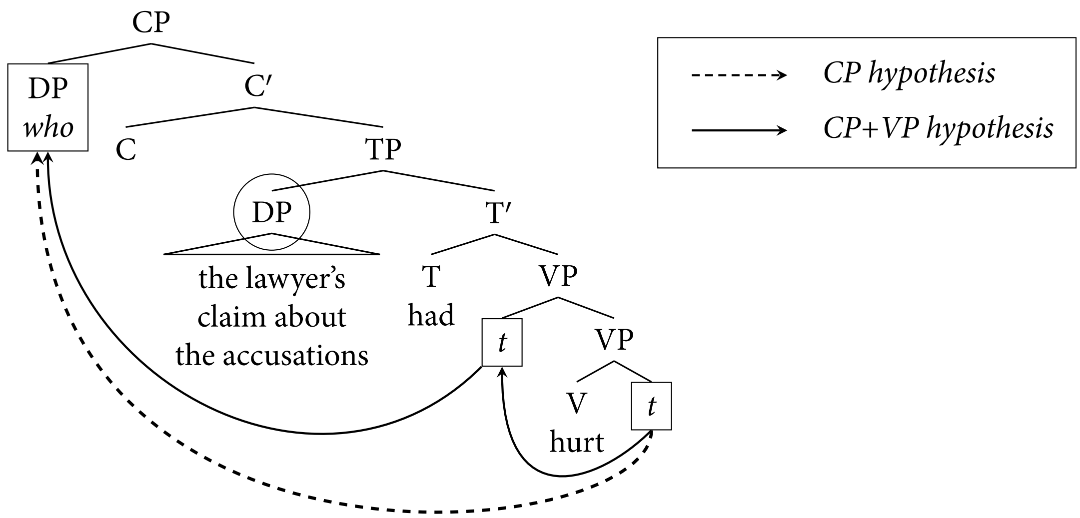
necessary to derive them under each hypothesis. The dashed lines indicate the movement path under the CP hypothesis; There is a single intermediate gap in the CP condition and none in the DP condition. The solid lines indicate movement under the CP+VP hypothesis: There are three intermediate gaps in the CP structure and one in the DP structure.

The results of Experiment 1 do not necessarily allow us to distinguish between the two hypotheses. They are clearly compatible with the predictions of the CP hypothesis: Because there exists an intermediate trace in the CP structure but not the DP structure, retrieval of the moved element at the final trace 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, suppose the simplest version: When the ultimate trace site is encountered, the parser has to retreat to its closest antecedent – either the moved element itself or an intermediate trace. On the CP+VP hypothesis, the most recent intermediate trace in both the CP and the DP structures lies in the VP edge immediately above the verb hosting the trace. In other words, the distance between the ultimate trace and the closest intermediate trace is identical in the two structures. If retrieving the antecedent amounted to a search for the most recent trace, 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

(8) *Structure of DP condition in Experiment 1*



it is not merely the the distance to the closest antecedent that affects processing speed but also the number of times an element has been previously activated. Applied to the structures at hand, it is plausible in light of Vasishth & Lewis (2006)'s findings that the retrieval of a moved element is inversely related to the number of intermediate traces of this element. In other words, the more intermediate traces exist between the overt position of the moved item and its ultimate trace, the faster the postulation and interpretation of this trace 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 trace position.[12]

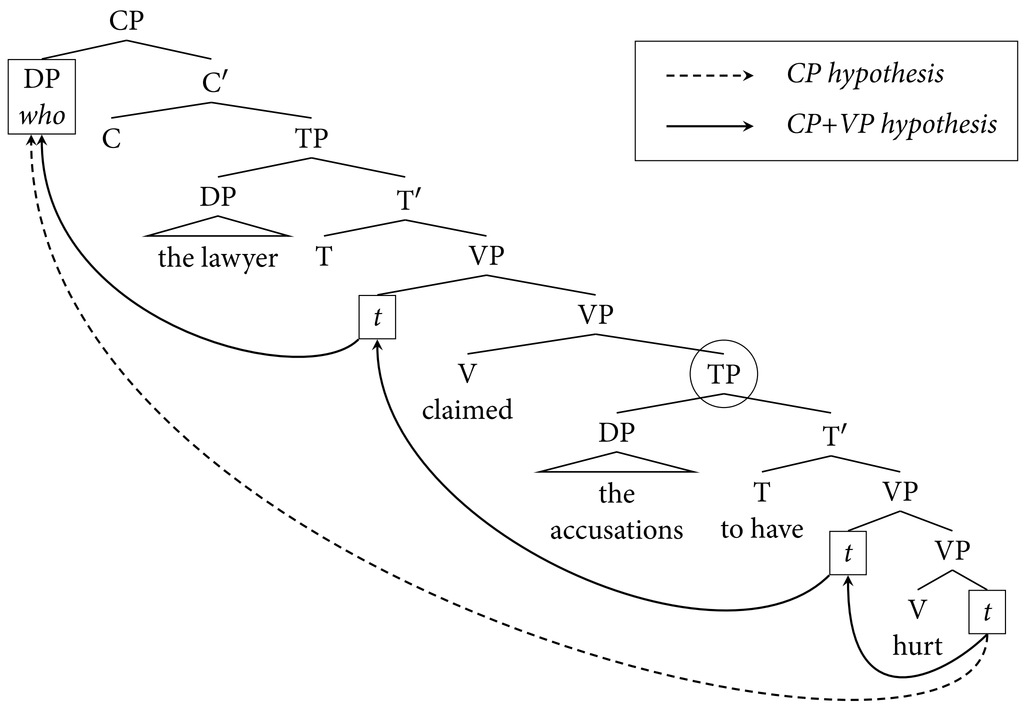
Against this view of antecedent retrieval, let us consider again the predictions the of the CP and the CP+VP hypothesis for the structures in (7) and (8). The predictions of the CP hypothesis remain unchanged: Because there is only a single intermediate trace in the CP structure and none in the DP structure, cumulative and non-cumulative reactivation are indistinguishable. The CP+VP hypothesis, by contrast, predicts that the antecedent is intermediately reactivated three times in 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 when the position of the ultimate trace is encountered and integration of this trace will be faster in the CP structure than the DP structure. As both hypotheses thus derive the critical effect, there is so far no evidence to distinguish between them if reactivation is assumed to be cumulative.

To distinguish between the two hypotheses, Experiment 2 contrasts the CP and DP structures with extraction out of a non-finite clause that lacks a CP layer. Concretely, the experiment in-

[12] 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, Gordon et al. 2001, 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. Crucially, reactivation is 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 retrieval.

vestigates extraction out of *exceptional case marking* constructions like the one in (9), which is modeled after the sample stimuli in (7) and (8) above.

(9) *Structure of TP condition*



The central characteristic of these constructions is that the embedded clause is not a CP but a TP (circled in (9)), and structures of this type will subsequently be referred to as the *TP condition*. Due to the removal of the CP, the CP hypothesis predicts no intermediate landing site in (9). As two VPs are crossed, the CP+VP hypothesis predicts two intermediate landing sites in the TP condition.

The CP and the CP+VP hypothesis make different predictions regarding the ease of integrating the trace of the moved element in the TP structure compared to the CP and the DP structures. The number of intermediate gaps postulated by the two hypotheses for each of the three structures are summarized in (10). Given cumulative reactivation, the more intermediate landing sites there are, the smaller the reading time increase at the gap site is predicted to be. According to the CP hypothesis, successive-cyclic movement is limited to the CP structure. The trace should consequently be more easily integrated in the CP structure than the other two, which should in turn be reflected in the reading times. The CP+VP hypothesis, on the other hand, predicts three intermediate traces in the CP structure, two in the TP structure and one in the DP structure. The ease of integrating the trace 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:

(10) *Predicted number of intermediate traces per structure*

	CP hypothesis	CP+VP hypothesis
CP structure (7)	1	3
DP structure (8)	0	1
TP structure (9)	0	2
Prediction:		
Reading time increase:	{DP, TP} > CP	DP > TP > CP

To summarize, the critical prediction of the CP hypothesis is that movement in the CP condition is easier than in the other two, as reflected in the reading time increase. The critical prediction of the CP+VP hypothesis is that the movement in the TP structure is easier than in the DP condition.

4.1 *Method*

4.1.1 *Materials*

Design

Thirty-six items like the one in (11) 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 Section A.2 of the supplementary materials. 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.

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

a. *CP intervener*

(i) [+move]
The journalist who the union member believed that the tax policy had intensely agitated 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 intervener*

(i) [+move]
The journalist who the union member's beliefs about the tax policy had intensely agitated 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 intervener*
  (i) [+move]
  The journalist who the union member believed the tax policy to have intensely agitated 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 trace 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 trace to separate the hypothetical landing site in VP from the final trace position.

Fillers

In addition to the 30 target sentences, the experiment contained 24 sentences of an experiment not reported here and 36 distractor sentences. All stimuli matched the target sentences in length and complexity.

4.1.2 *Participants*

162 participants were recruited using Amazon Mechanical Turk. Each participant received a compensation of USD 1.

4.1.3 *Procedure*

The experiment procedure was identical to the one in Experiment 1. Stimuli were regioned as shown in Table 5. As in Experiment 1, participants had to take a 10-second break after every twelve trials and were able to rest longer if they desired. The total experiment contained seven of these breaks.

4.1.4 *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. No participant took the experiment twice. 57 participants had to be excluded from analysis for falling below the accuracy threshold. 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 ([−

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

move] vs. [+*move*]) and *Intervener* (*CP* vs. *DP* vs. *TP*). Accuracy data were analyzed using a parallel logistic LME model. The predictor *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 ($CP = -\frac{2}{3}, DP = \frac{1}{3}, TP = \frac{1}{3}$). 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 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 region 3, the factor *intervener* was sum-coded in this region ($CP = -.5, DP = .5$).

Because one item (item 8 in Section C.2 of the supplementary materials) was coded incorrectly, it was excluded from all analyses.[13]

4.2.1 *Comprehension question accuracy*

The mean accuracy on the comprehension questions was 81.1%. Accuracy by condition is given in Table 6. Logistic LME modeling revealed that (i) accuracy was lower in the [+move] conditions ($\hat{\beta} = -.95, z = -8, p < .001$); (ii) accuracy in the CP conditions was higher than in the DP and TP ones ($\hat{\beta} = -.94, z = -3.2, p < .01$); and (iii) accuracy in the TP conditions was higher than in the DP conditions ($\hat{\beta} = -.65, z = -2.9, p < .01$). Importantly, the type of the intervener did not affect the accuracy difference between [–move] and [+move] conditions (all ps > .1).

Table 6. Mean accuracy by condition in Experiment 2

	CP	DP	TP
[–move]	.95	.85	.92
[+move]	.77	.66	.71

4.2.2 *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 Section B.2 in the supplementary materials.

To distinguish between the CP hypothesis and the CP+VP hypothesis, the crucial evidence is the pattern at the region containing the trace (region 7) and the spillover region (region 8). 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 all other regions are provided in Section B.2 of the supplementary materials.

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

[13] Eliminating this item did not introduce a plausibility confound (p's > .5), cf. Section A.2 in the supplementary materials.

Table 7. Mean raw reading times by condition and region in ms 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]	—	846	496 (11)	656	491 (27)	496	521	729 (−48)	1268 (87)
	[+move]	699	949	559 (75)	715	495 (31)	506	548	673 (−4)	879 (−1)
DP	[−move]	—	1028	521 (13)	678	496 (32)	501	514	705 (−67)	1209 (33)
	[+move]	705	1195	613 (106)	770	513 (48)	504	537	711 (33)	899 (33)
TP	[−move]	—	854	—	774	545 (12)	521	525	697 (−76)	1272 (97)
	[+move]	692	1012	—	833	579 (44)	527	593	696 (24)	885 (9)

high reading time increase in the TP condition: The increase was greater than in the CP condition, marginally significant by participants ($t_1(103) = 1.84, p = .07$) and fully significant by items ($t_2(28) = 2.8, p < .01$) but did not differ between the CP and DP conditions ($t_1(103) = .19, p > .5$; $t_2(28) = .26, p > .5$).

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 between 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 ($\hat{\beta} = 57, t = 2.5, p < .05$). It also showed that the [+move] conditions were read more slowly than the [−move] ones ($\hat{\beta} = 80, t = 5.2, p < .001$). 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.[14]

Table 8. Coefficient estimates and corresponding t-values for the linear mixed effects model analyses of log reading times in crucial regions of Experiment 2. *Mvmt* refers to the sum-coded factor *movement; InterveCP–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 Section B.2. of the supplementary materials

	Region			
	7		8	
	$\hat{\beta}$	t	$\hat{\beta}$	t
Movement	0.054	4.15	−0.040	−1.14
IntervCP–TPDP	0.013	1.16	0.001	0.04
IntervTP–DP	−0.042	−3.70	−0.002	−0.17
Mvmt:IntervCP–TPDP	0.022	1.00	0.058	2.08
Mvmt:IntvTP–DP	−0.046	−2.01	−0.021	−0.74

[14] Although not critical for the evaluation of the two hypotheses considered here, a post hoc analysis was carried

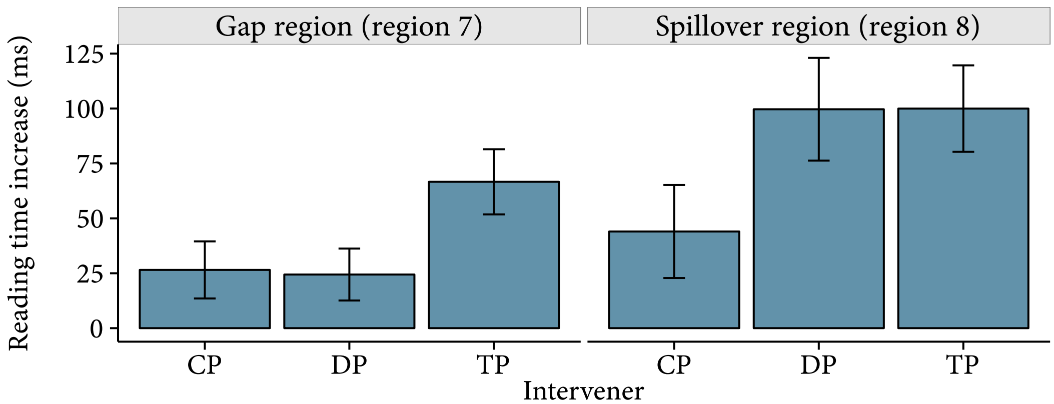


Figure 2. Mean increase in residual reading times between [–move] and [+move] conditions in gap and spillover regions and by-particpant standard errors in Experiment 2 by intervener

Figure 2 plots the crucial reading time increase between the [–move] and [+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 substantially larger for TP structures in the gap region and (iii) the increase is substantially smaller for CP structures in the spillover region.

4.3 *Discussion*

Recall from (10) the predictions of the two hypotheses: The critical prediction of the CP hypothesis is that the reading increase in the CP condition is smaller than in the DP and the TP condition. The CP+VP hypothesis, on the other hand, critically predicts that the reading time increase is greater in the DP condition than in the TP condition.

The results of Experiment 2 support the CP hypothesis. In the gap region, the reading time increase is greater in the TP condition than in either of the other two; in the spillover region, the increase in the both the TP and the DP region is greater than in the CP condition. That the advantage of the CP condition surfaces in the spillover region replicates the result of Experiment 1 above. Integrating the filler is thus easier in the CP condition than in the other two. This finding is in accordance with the critical prediction of the CP hypothesis: The reading time increase is smaller in the CP condition than in the other two. On the other hand, the critical prediction of the CP+VP hypothesis that integration should be easier in the TP than in the DP condition was not supported. In fact, the results showed the opposite, with the reading time increase being significantly greater in the TP than in the DP condition. There is, in other words, evidence that

(*fn. continued*)
out to compare the reading time increase in the CP condition to that in the DP condition in the spillover region. The rationale for this test was to replicate G&W's key finding, replicated in Experiment 1 above, that the reading time increase is greater in the DP condition than in the CP condition. Pairwise t-tests comparing the reading time difference between [+move] and [-move] conditions in CP and DP structures were carried out on the log-transformed and residual reading times. Analysis of the former revealed that the difference was marginally smaller in the CP condition than in the DP condition ($t_1(103) = 1.63, p = .11; t_2(28) = 2.03, p = .05$). This difference was fully significant in the residual reading times ($t_1(103) = 2.05, p < .05; t_2(28) = 2.60, p < .05$). The results of Experiment 2 are thus fully consistent with those of Experiment 1.

retrieval of the filler is harder in the TP than in the DP condition, which disconfirms the prediction of the CP+VP hypothesis.

While the results thus bear out the critical prediction of the CP hypothesis, it should be noted that this hypothesis by itself does not make a commitment with regard to the relation between the TP and DP structures. Put differently, as far as intermediate landing sites are concerned, neither of the two conditions should have an advantage. That the reading time increase is greater in the TP condition must thus be due to a second factor.[15] A plausible candidate is *structural distance*. There is evidence that increased structural distance between a moved element and its trace is associated with greater retrieval difficulty at the gap site (e.g. Kroch 2006, Lin & Bever 2006). 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 between conditions. Concretely, in the DP condition the moved element was in the same clause as its trace (as not clause boundary was crossed in this condition). In the TP condition, on the other hand, the moved element resided in a higher clause than its trace. Retrieval of the filler thus has to traverse a clause boundary in the TP condition. That the reading time increase was greater in the TP than in the DP condition is likely a reflex of this difference in the structural length of the dependency.

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 structural difference induces a penalty in the TP condition relative to the DP condition, it must not do so in the CP condition (or else the CP and TP condition would pattern alike). If movement in the CP condition were cross-clausal in the 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 trace 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 claims. In other words, the penalty for cross-clausal movement is absent in the CP condition precisely because there exists an intermediate trace within the same clause. Reference to structural distance hence does not obviate the need for successive cyclicity through Spec,CP.

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 again the structures in (7–9). On the CP+VP hypothesis, the intermediate trace closest to the ultimate trace site is at the edge of the embedded VP in all three structures. As a consequence, the structural distance between the trace and its closest antecedent is identical in all 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 trace 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 support the CP hypothesis because the application of structural distance is successful only on this hypothesis.

One possible concern with the results of Experiment 2 is that the higher verb is a CP/DP-verbs (using the terminology of Experiment 1). Given the results of Experiment 1, it is expected that

[15] This failure of the CP hypothesis to predict the TP–DP contrast does not put it on a par with the CP+VP hypothesis. While the former merely fails to predict an observed contrast, the latter critically predicts the opposite of the observed results.

the parser initially postulates a trace of the moved element as the object of these verbs in the TP and CP conditions but not in the DP condition. It is, however, unlikely that the crucial pattern at the gap and spillover regions are an artifact of this premature gap filling. In Experiment 1, verb type had an impact on the reading time pattern in the region following the higher verb but not in the trace region. This indicates that structural revision is fast and a parse temporarily pursued in the higher clause has no discernable impact on the regions of interest, in concordance with the previous literature on structural reanalysis. Moreover, 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 makes it very implausible that the pattern at the gap region is a sustained effect of gap filling at the higher verb.

Finally, the relevance of the notion of cumulative reactivation deserves some remarks. As discussed 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 above, 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. 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. This ensures maximal generality of the results obtained here: No stance on whether intermediate reactivation is cumulative or not is necessary for the experiments here to be interpretable in the context of the hypotheses. The CP hypothesis is uniformly corroborated whereas the CP+VP hypothesis is uniformally disconfirmed.

5 General discussion

5.1 *Parsing evidence for phases*

This paper has brought a novel type of empirical evidence to bear on the question of syntactic locality. It has presented evidence from sentence processing that corroborates the general conclusion that long movement dependencies are constructed successive-cyclically. To find evidence for successive cyclicity in online parsing is especially relevant in light of current theories about the underlying cause of successive cyclicity.

Chomsky (2000, 2001, 2004, 2008) pursues the intuition that successive-cyclic movement is the direct result of the strictly local nature of syntactic derivations (also see Berwick & Weinberg 1984 for a related proposal). He argues that syntactic derivations apply in a 'moving window' mode, where built syntactic structure is shunted at regular intervals (a process called 'Spell-Out'). Because shunted material becomes inaccessible, moving elements have to be periodically moved out of a built constituent before it is shunted. The constituents that are subject to Spell-Out are commonly called 'phases'. In conjunction with incremental structure-building, the phase-based nature of syntactic derivation necessitates successive-cyclic movement.

That phase-based cyclic Spell-Out should exist in the first place is in turn attributed by Chomsky (2000, 2001) to general constraints on working memory that limit the amount of material that can be held accessible at any given point. If working memory is taken to be the computational resources underlying sentence processing—as is the null assumption—, one naturally expects to find evidence for intermediate landing sites in online sentence comprehension, a domain in which constraints on working memory play a perennial role. That such evidence indeed exists provides rather intriguing support for the overall program of reducing properties of syntax to general cognitive constraints (Chomsky 2005's 'third factor'). As such, psycholinguistic evidence for successive cyclicity not only aligns well with other types of tests for successive cyclicity, it in addition provides important evidence as to why syntactic movement should be successive-cyclic to begin with.

5.2 *The distribution of phases*

In addition to providing evidence for successive cyclicity in sentence processing, the results presented here shed light on the question of where successive-cyclic movement takes place. We have seen evidence that the edges of finite clauses are targeted by successive-cyclic movement while extended VPs are not. These findings support the traditional view that only CPs (i.e., finite clauses) constitute locality domains that give rise to the successive-cyclic application of movement. The results hence directly contrast with the currently widely adopted view that VP edges act just like CP edges in enforcing intermediate touchdowns.

If successive cyclicity is the result of phase-based syntactic derivations and the cyclic Spell-Out they incorporate, these results have immediate implications for the distribution of phases. Quite generally, if movement over a projection XP does not have to pass successive-cyclically through the specifier of XP, then it follows that XP cannot be a phase. The evidence presented here thus supports the view that C is a phase head but also entails that there is no phase head within the extended VP domain. This conclusion challenges the widely held assumption that both the C as well as the V domain host a phase head and thus pattern alike as far as the need for successive cyclicity is concerned.

It appears, then, that the processing evidence here lines up closely with traditional diagnostics for successive-cyclic movement when it comes to the status of C but at the same time the two seem to diverge when the extended VP is considered. It is thus worth considering some of the arguments that have been adduced to support successive-cyclic movement through VP edges. While a comprehensive discussion of all arguments that have been advanced in the literature is clearly beyond the scope of this paper, it is nonetheless instructive to discuss a representative sample. As it turns out, the case these arguments make for successive cyclicity through VP edges is somewhat less compelling than it is often taken to be. In particular, as has been noticed by a number of authors, many of the empirical phenomena brought to bear on this question can be recast in a system that does not involve intra-clausal successive cyclicity.

The orginal, conceptual motivation for intermediate VP gaps stems from Chomsky (2000, 2001), who defines phases as constituents that are 'propositionally complete'. Because a VP contains a verb and all of its arguments, it counts as propositionally complete and thus constitutes a phase, thereby requiring successive-cyclic movement through its specifier. It has often been noted that this purely conceptual argument faces a number of severe problems (Bošković 2002, Epstein & Seely 2006, Boeckx & Grohmann 2007). To give just one example, Chomsky (2000) argues that finite clauses, but not non-finite ones, are phases, yet both are propositional (Bošković 2014).

More intriguing are empirical arguments that have been brought forth in support of intermediate movement to VP edges. One type of argument involves semantic reconstruction effects. Fox (1999: 174) presents the sentence in (12). In (12) a complex DP containing a proper name (*Ms. Brown*) and a pronoun (*he*) is *wh*-moved. Crucially, this movement path crosses a pronoun coindexed with the proper name and the quantified DP *every student* binding the pronoun. This configuration imposes two constraints: First, due to Principle C, the proper name must not be c-commanded by the pronoun *her*. Second, the pronoun *he* contained inside the DP must be c-commanded by *every student* in order to be bound. Fox (1999) reasons that the DP can neither be interpreted in its surface position—because *every student* does not c-command *he* in this position—, nor in its base position—as the proper name *Ms. Brown* would be c-command by *her*, creating a Principle C violation. The only position that allows the binding requirement and Principle C to be simultaneously satisfied is situated between *every student* and *her* (indicated by an underline in (12)). Because no CP boundary falls within this range, Fox (1999) concludes that there exists an intermediate trace within the VP region.

(12) [Which of the books that he$_1$ asked Ms. Brown$_2$ for] did every student$_1$ ___ get from her$_2$ $\underline{\ \ *\ \ }$?

As den Dikken (2006) points out, however, this conclusion is not necessary. Den Dikken (2006) notes that the only possible answer to (12) is a pair list answer. Kiss (1993) has proposed that pair list readings are obtained by covert movement of the quantified DP over the *wh*-expression. Following this proposal, den Dikken (2006) concludes that the LF representation for (12) is (13), where *every student* has been moved over the complex DP. This derives the correct result that the books are distributed over each student (i.e., that there is a potentially different book for each student).

(13) [every student]$_1$ [which of the books that he$_1$ asked Ms. Brown$_2$ for]$_3$ did t$_1$ get from her$_2$ t$_3$

Den Dikken (2006) then notes that no reconstruction whatsoever is necessary to fulfill the requirements of the various items in (13). In particular, because *every student* c-commands *he*, binding is possible. Moreover, because *her* does not c-command *Ms. Brown*, no Principle C violation arises. Because no resort to reconstruction to any position is thus necessary to derive the grammaticality of (12), it does not constitute evidence for an intermediate landing site within the VP domain.

A second type of diagnostic for intermediate landing sites comes from so-called *wh*-copying constructions. In these constructions, a single *wh*-element can be multiply realized—in the Spec,CP it takes scope from as well as in lower Spec,CP along the movement path. (14) provides an example from the Algonquian language Passamaquoddy (Bruening 2006).

(14) a.

[$_{CP}$ *Tayuwe*	kt-itomups	[$_{CP}$ *tayuwe*	apc	k-tol-i	malsanikuwam-ok]]?
when	2-say	when	again	2-there-go	store-LOC

‘When did you say you’re going to go to the store?’

b. [$_{\text{CP}}$*wenil* Mali wewitahamacil [$_{\text{CP}}$*wenil* kisiniskamuk]]?
who Mary remember who dance.with
'Who does Mary remember I danced with?'

Semantically and syntactically, there is only a single *wh*-element in (14a,b). In line with previous literature on related phenomena in other languages, Bruening (2006) argues that the two occurrences of the *wh*-element are distinct realizations of the same syntactic element. In other words, the lower occurrence of the *wh*-element in (14) is the realization of a trace (or copy) inside the lower Spec,CP. This in turn constitutes evidence that the extraction applies successive-cyclically, passing through the lower Spec,CP and creating a trace in this position. Similar facts have been observed for German (Müller 1999), Romani (McDaniel 1989), Frisan and Afrikaans (Felser 2004), Hungarian (Horvath 1997), and Ancash Quechua (Cole 1982).

To the best of my knowledge, there is no reported case of *wh*-copying in VP edges. A crucial case would be language which intraclausal *wh*-movement as in (15a) and crossclausal movement as in (15b), where copy instances of *who* are italicized.

(15) *CP+VP copying*

a. [$_{\text{CP}}$ Who did you [$_{\text{VP}}$ *who* see]]?

b. [$_{\text{CP}}$ Who did John [$_{\text{VP}}$ *who* say [$_{\text{CP}}$ *who* Mary [$_{\text{VP}}$ *who* saw]]]]?

On the CP+VP hypothesis, movement creates intermediate landing sites in VP and CP edges. Nothing thus rules out a language in which copies in both positions are overtly realized, as in (15). In fact, such a case should be the default because in the absence of extraneous stipulations all intermediate landing sites should behave alike. Yet not only is such a pattern clearly not the default, it completely unattested to my knowledge. That intermediate copies are widely attested in CP edges but crosslinguistically absent in VP edges constitutes rather strong evidence for the CP hypothesis and against the CP+VP hypothesis.

A third type of argument for intermediate landing sites at VPs has been put forth by Barbiers (2000) and is based on *what for*-splits in Dutch. In constructions of the form '*what for NP*', *what* may undergo *wh*-movement on its own, thereby stranding *for NP*. Thus, in (16a), *wat* is extracted out of the complex *wh*-phrase *wat voor boeken* 'what books' and moved into the matrix Spec,CP, stranding *voor boeken* 'for books' in its base position. As (16b) illustrates, it is also possible to strand *wat voor* in a position in the higher clause, immediately preceding the matrix verb.

(16) a. Wat voor-*split in base position*
Wat$_1$ had je gedacht [$_{\text{CP}}$dat Jan [t$_1$ voor boeken] zou lezen]?
what had you thought that John for books would read
'What kind of books did you think that John would read?'

b. Wat voor-*split at higher VP*
Wat$_2$ jad jij dan [t$_2$ voor boeken]$_1$ [$_{\text{VP}}$gedacht [$_{\text{CP}}$dat Jan t$_1$ zou
what had you then for books thought that John would
lezen]]?
read
'What kind of books did you think that John would read?'

Barbiers (2000) also notes that the *voor NP* remnant cannot appear immediately to the right of the matrix verb (*gedacht* in (16b)). Finally, he observes that it is independently possible to move an element out of an embedded clause into a position preceding the matrix verb. As evidence, he provides the fragment answer in (17), which he analyzes as involving movement of *Jan* into the matrix clause, followed by elision of the complement clause.

(17) *Matrix scrambling*

A: Wie denk je dat het probleem heeft opgelost?
Who think you that the problem has solved
'Who do you think solved the problem?'

Q: Ik denk Jan ~~[$_{CP}$ dat het probleem heeft opgelost]~~
I think John

Because it is thus possible to strand a *for*-phrase in VP edges but not CP edges, Barbiers (2000) concludes that successive-cyclic movement targets only the former. On this account, (16b) is derived by first moving the complex DP to the edge of the matrix VP, followed by subextraction of *wat*. That it is not possible to strand *voor boeken* after the matrix verb in (16b), is then taken as evidence that no intermediate movement to the embedded Spec,CP has taken place.

This argument is not conclusive, however. As is well-known, there are independent constraints on which positions stranding may take place in (for an overview see Bobaljik 2003). Suppose that subextraction of *wat* may only take place in positions in which a feature is checked, i.e. positions that are not purely intermediary (criterial positions, to use a term by Rizzi 2006). The edge of VP is a criterial position because, as Barbiers (2000) shows, it can be independently targeted by scrambling (17) and is thus not purely intermediary. Intermediate Spec,CPs, on the other hand, are not criterial because movement to them is motivated solely in order to escape a phase domain. On this alternative account, the reason that *voor NP* cannot be stranded in Spec,CP is precisely because movement to this position is purely intermediary. That stranding in VP edges is possible is, in turn, due to the fact that scrambling to the VP edge is not due to escape a phase domain: The derivation of (16b) involves a scrambling movement step, followed by an independent *wh*-movement step that strands a preposition inside the scambled constituent. It follows from this alternative account that the Dutch facts above do not constitute evidence for successive-cyclic movement through VP edges. Independent support for an account along these lines comes from English: As is well-known, while *wh*-movement may strand prepositions, it does not do so in intermediate Spec,CP positions. In other words, it is not possible to move a *wh*-PP to an intermediate Spec,CP and then strand the preposition there. If stranding is limited to criterial positions, this restrictions is straightforwardly accounted for.

One conclusion that can be drawn from this admittedly limited consideration of traditional arguments for intermediate landing sites is that the disparity between the evidence from processing on the one hand and from more traditional syntactic domains on the other is less severe than it appears at first glance. The evidence for intermediate traces within VP domains is less compelling than it is standardly taken to be: Either the evidence is equally compatible with a view that only recognizes such traces in Spec,CP, or the curious typological absence of patterns predicted to exist under successive-cyclic movement through VP provides indirect evidence against such an account.

The status of intermediate traces at VP edges thus remains questionable. The processing evidence presented in this paper brings to the table a new type of evidence. This evidence directly favors the CP hypothesis over the CP+VP hypothesis, a conclusion that, as just noted, converges with other diagnostics of successive cyclicity. If correct, these considerations require a major reassessment of the distribution of phases. It suggests that syntactic locality domains in general and phases in particular are larger than is commonly assumed and that the traditional view according to which only CPs constitute a locality domain is empirically favored over a system with smaller and more domains.

5.3 *Implications for theories of parsing*

The main focus of this paper was the existence and role of successive cyclicity in sentence parsing. The evidence gathered in the course of investigating this question additionally sheds light on questions regarding general parsing strategies. This section will discuss these implications for sentence parsing.

Experiments 1 and 2 investigated how the type of the verb that is crossed by a cross-clausal movement dependency affects the processing of that dependency at the gap site. This manipulation provided evidence against a premature gap filling account of G&W's effect. In addition, the manipulations of these factors sheds light on how the parser employs them to decide on where to postulate traces.

First, Experiment 1 found evidence against a strong version of hyperactive gap filling, according to which the parser postulates gaps completely regardless of subcategorization information (see Omaki et al. 2012 for the concept of hyperactive gap filling). The elevated reading times at a complementizer in this experiment make it clear that a preceding verb could not have hosted the trace of a moved DP. Crucially, this effect was significantly larger for verbs that can in principle host a DP object than for verbs that only take clausal complements. This contrast indicates that the subcategorization frame of the verb must have affected to parser's inclination to posit a trace as the object of this verb. This finding is consistent with the view that the parser obeys a verb's subcategorization constraints when deciding where to postulate traces (Staub 2007) or that subcategorization restrictions modulate the parser's inclination to posit traces. It is not consistent with strongly hyperactive gap filling, according to which the parser postulates a trace as early as possible regardless of the verb's subcategorization restrictions.

We have furthermore seen evidence that subcategorization frequency and real world plausibility have less of an impact on trace postulation. The results of Experiment 1 provide evidence that the parser postulates a DP trace after DP-taking verbs like *claim* even if considerations of frequency and plausibility disfavor such a trace. This result does not, of course, entail that frequency and plausibility information is ignored in initial parsing (see, e.g., Gibson & Pearlmutter 1998). But it indicates that frequency and plausibility does not invariably prevent the parser from postulating an otherwise licit trace.

It is also worth emphasizing once more that the experiments reported here do not offer evidence for or against cumulative reactivation of a moved element. Reactivation is cumulative if multiple reactivations of some item increase this item's activation level to a greater degree than a single reactivation. As already discussed in section 4.3, the results here are compatible with both cumulative and non-cumulative reactivation and so are the conclusions with respect to the distribution successive cyclicity. Under non-cumulative reactivation, Experiment 1 provides evidence against the CP+VP hypothesis because the closest reactivation site in Spec,*v*P is equally far away from the actual trace position in both the CP and the DP condition. If cumulative reactivation is assumed, the results of Experiment 2 constitute evidence against this hypothesis. The CP hypothesis, on the other hand, makes the correct predictions for both experiments here regardless of whether reactivation is cumulative or not, simply because in every structure there is at most one intermediate reactivation site. Because cumulative and non-cumulative reactivation make different predictions only if there is more than one reactivation site, the difference between the two is irrelevant to the predictions of the CP hypothesis. As a consequence, the experiments do not offer evidence for or against either view of reactivation and the CP hypothesis emerges as superior under either view.

Lastly, the results here have important implications for current theories of the retrieval processes involved in the online computation of long-distance dependencies. Recall from Experiment 2 that the integration of the trace incurred greater processing cost in the TP condition than in the DP condition. Because no CP was crossed in either condition, this contrast cannot be attributed to successive cyclicity. In other words, in both conditions the closest antecedent is the overt instance of the moved element. This differences in the retrieval difficulty in the TP and the DP condition is particularly illuminating against the background of current theorizing of the factors regulating retrieval processes in sentence comprehension. Traditional accounts of retrieval processes assume that elements within a local syntactic domain are easier to access than elements outside of this domain (Kimball 1973, Frazier 1978, Berwick & Weinberg 1984, Frazier & Clifton 1989, Gibson 1998, Sturt et al. 1999). Alternative accounts of retrieval processes eschew a syntactic distinction between local and non-local domains and characterize retrieval difficulties solely in terms of an element's activation decay and interference from similar other elements (e.g., Van Dyke & Lewis 2003, Lewis & Vasishth 2005, Lewis et al. 2006, Bartek et al. 2011). This family of accounts treats the availability of an element for retrieval as a function of its activation level, which is subject to decay over time, and the interference from similar elements that match a retrieval cue. In contrast to the former type of analysis, syntactic locality and domains do not enter directly into the computation of the ease of retrieving an element. Instead, recency effects are recast as the indirect result of decay and interference: The greater the structural or linear distance between the encoding of an element and the point at which it needs to be retrieved the greater the effect of decay on this element's activation level, thus increasing retrieval difficulty. Similarly, greater distance corresponds to a greater number of similar elements that interfere with retrieval.

The aforementioned contrast between the ease of retrieving the moved element in the TP and DP condition does not obviously correlate with the predictions made by decay and similarity-based interference but does follow from structural locality effects. Consider first the effects of time-based decay on the activation level of the moved element when the trace-hosting verb is encountered. The two structures do not differ with respect to the distance between the moved element and its trace so that the effect of decay should be comparable. The two structures do differ, however, with regard to similarity-based interference. As the DP condition contains an additional noun phrase (*beliefs* in (11)), retrieval of the moved element should be subject to greater interference in this condition than in the TP condition. The processing cost at the trace site should thus be greater in the DP than in the TP condition. Yet the results of Experiment 2 show the opposite. It is therefore unclear how the relative difficulty in the TP structures could be accounted for solely in terms of decay and interference.

A more promising avenue of analysis is based on syntactic locality. While the moved element is equally distant from its trace linearly in the TP and DP structures, their structural distance differs considerably. Whereas the moved element is situated in the same clause as its trace in the latter structure, it is located in a higher clause in the former. If leaving the current clause increases the cost of retrieving an element, the disadvantage of the TP structure is accounted for. This analysis involves crucial reference to the syntactic structure separating the element to be retrieved and its trace. It requires that the retrieval process be affected by properties of syntactic domains and is hence not available to theories that deny a direct impact of syntactic structures on retrieval processes.

These considerations converge in a rather striking way with evidence from the processing of reflexives reported in Dillon et al. (2014). Based on Mandarin Chinese, Dillon et al. (2014) investigated whether the retrieval of an antecedent for a reflexive is affected above and beyond the impacts of decay and interference by whether or not this antecedent is contained within the same clause as the reflexive or a higher one. Their findings support the view that clause affiliation does affect retrieval difficulty and conclude that syntactic domains have a direct impact on retrieval processes. The results here corroborate this conclusion and suggest that this conclusion is not limited to the processing of reflexives but likewise holds for movement dependencies.

6 Conclusion

Based on the results of two self-paced reading experiments, this paper has argued that there is evidence for successive cyclicity in online sentence comprehension and that this evidence favors the classical view, according to which movement out of finite clauses is successive-cyclic but movement out of non-finite clauses and movement within clauses is not. In technical terms, CP edges host intermediate landing sites but VP edges do not.

This paper took as its empirical starting point results by Gibson & Warren (2004), who presented initial evidence for intermediate reactivation of a moved element if a CP boundary is crossed. However, their experiment did not determine whether this intermediate reactivation was due to successive cyclicity (i.e., an intermediate landing site) or premature gap filling (viz., an incorrect association with a higher verb). Experiment 1 presented evidence that the reactivation they observed is indeed due to successive cyclicity.

To determine whether successive cyclicity is restricted to finite clauses or also takes place at VP edges, Experiment 2 investigated extraction out of non-finite clauses. The results show that extraction dependencies in this configuration are harder than would be predicted if VP edges hosted intermediate landing sites. Instead, the results are compatible only with the classical view that only finite clause host successive-cyclic landing sites.

These findings not only bring an entirely new type of evidence to bear on the current debate about the location and distribution of syntactic locality domains. They are also in line with approaches that view syntactic locality as emerging from syntax-external constraints on working memory (Berwick & Weinberg 1984, Chomsky 2000, 2001, 2005). Furthermore, these results have immediate implications for the theory of phases as successive cyclicity provides a diagnostics for phasehood. While the results the notion that C is a phase head, they also entail that there is is no phase head within the extended verbal projection. Although this conclusion appears to starkly contrast at first glance with more traditional diagnostics for successive-cyclic movement, there are reasons to doubt the force of these traditional arguments. At least several of the critical empirical generalizations can be accounted for with or without successive-cyclic movement through VP edges and they are hence less than conclusive. Sentence processing provides a welcome additional source of evidence, which indicates that phases are larger and less frequent structurally than is standardly assumed at present.

Supplementary materials

Supplementary materials—including a description of the plausibility-norming procedure, a comprehensive analysis of all regions of Experiments 1 and 2 and a list of the stimuli used—is available at http://people.umass.edu/keine/papers/Keine_Locality_domains_Appendix.pdf.

Acknowledgments

I am particularly indebted to Brian Dillon and Lyn Frazier for discussions, suggestions and help throughout the various stages of this project. Rajesh Bhatt, Chuck Clifton, Janet Fodor, Kyle Johnson, Ethan Poole, Shayne Sloggett and Tessa Warren provided very helpful feedback, for which I am grateful. Portions of this work were presented at CUNY 28 (University of Southern California) and WCCFL 33 (Simon Fraser University). I thank the audiences for their questions and suggestions.

References

Aldridge, Edith (2008). Phase-based account of extraction in Indonesian. *Lingua* 118: 1440–1469.

Barbiers, Sjef (2000). Intermediate landing sites. *GLOT International* 4: 15–16.

Barr, Dale, Roger Levy, Christoph Scheepers & Harry Tily (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68: 255–278.

Barss, Andrew (1986). Chains and anaphoric dependence. Ph.d. thesis, MIT, Cambridge, MA.

Bartek, Brian, Richards Lewis, Shravan Vasishth & Mason Smith (2011). In search of on-line locality effects in sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37: 1178–1198.

Bates, Douglas, Martin Maechler, Ben Bolker & Steven Walker (2014). lme4: Linear mixed-effects models using Eigen and S4, R package version 1.1-7.

Berwick, Robert & Amy Weinberg (1984). *The Grammatical Basis of Linguistic Performance: Language Use and Acquisition*. Cambridge, MA: MIT Press.

Bobaljik, Jonathan (2003). Floating quantifiers: Handle with care. In: *The Second GLOT International State-of-the-Article Book*, ed. by Lisa Cheng & Rint Sybesma, Berlin: Mouton de Gruyter, pp. 107–148.

Boeckx, Cedric & Kleanthes Grohmann (2007). Putting phases in perspective. *Syntax* 10: 204–222.

Boland, Julie, Michael Tanenhaus, Susan Garnsey & Greg Carlson (1995). Verb argument structure in parsing and interpretation. *Journal of Memory and Language* 34: 774–806.

Bošković, Željko (2002). A-movement and the EPP. *Syntax* 5: 167–218.

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

Bruening, Benjamin (2006). Differences between the wh-scope-marking and wh-copy constructions in Passamaquoddy. *Linguistic Inquiry* 37: 25–49.

Chen, Evan, Edward Gibson & Florian Wolf (2005). Online syntactic storage costs in sentence comprehension. *Journal of Memory and Language* 52: 144–169.

Chomsky, Noam (1973). Conditions on transformations. In: *A Festschrift for Morris Halle*, ed. by Stephen Anderson & Paul Kiparsky, New York: Academic Press, pp. 232–286.

Chomsky, Noam (1977). On wh-movement. In: *Formal Syntax*, ed. by Peter Culicover, Tom Wasow & Adrian Akmajian, New York: Academic Press, pp. 71–132.

Chomsky, Noam (1981). *Lectures on Government and Binding*. Dordrecht: Foris.

Chomsky, Noam (1986). *Barriers*. Cambridge, MA: MIT Press.

Chomsky, Noam (2000). Minimalist inquiries: The framework. In: *Step by Step: Essays in Syntax in Honor of Howard Lasnik*, ed. by Roger Martin, David Michaels & Juan Uriagereka, Cambridge, MA: MIT Press, pp. 89–155.

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

Chomsky, Noam (2004). Beyond explanatory adequacy. In: *Structures and Beyond. The Cartography of Syntactic Structures, Vol. 3*, ed. by Adriana Belletti, Oxford: Oxford University Press, pp. 104–131.

Chomsky, Noam (2005). Three factors in language design. *Linguistic Inquiry* 36: 1–22.

Chomsky, Noam (2008). On phases. In: *Foundational Issues in Linguistic Theory: Essays in Honour of Jean-Roger Vergnaud*, ed. by Robert Freidin, Carlos Otero & Maria Luisa Zubizaretta, Cambridge, MA: MIT Press, pp. 89–155.

Chung, Sandra (1994). Wh-agreement and 'referentiality' in Chamorro. *Linguistic Inquiry* 25: 1–44.

Chung, Sandra (1998). *The Design of Agreement: Evidence from Chamorro*. Chicago: Chicago University Press.

Clements, George (1994). Principles of tone assignment in Kikuyu. In: *Autosegmental Studies in Bantu Tone*, ed. by George Clements & John Goldsmith, Dordrecht: Foris, pp. 281–339.

Clements, George & Kevin Ford (1979). Kikuyu tone shift and its synchronic consequences. *Linguistic Inquiry* 10: 179–210.

Cole, Peter (1982). *Imbabura Quechua*. Amsterdam: North-Holland Publishing.

de Vincenzi, Marica (1991). *Syntactic Parsing Strategies in Italian*. Dordrecht: Kluwer.

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

Dillon, Brian, Wing-Yee Chow, Matthew Wagers, Taomei Guo, Fengqin Liu & Colin Phillips (2014). The structure-sensitivity of memory access: Evidence from Mandarin Chinese. *Frontiers in Psychology* 5: 1025.

Drummond, Alex (2013). Ibex Farm (version 0.3) [software]. (Retrieved from https://github.com/addrummond/ibexfarm).

Dukes, Michael (1992). On the status of Chamorro wh-agreement. In: *Proceedings of the Eleventh West Coast Conference on Formal Linguistics (WCCFL 11)*, Stanford: CSLI Publications, pp. 177–190.

Epstein, Samuel David & T. Daniel Seely (2006). *Derivations in Minimalism: Exploring the Eliminination of A-Chains and the EPP*. Cambridge: Cambridge University Press.

Felser, Claudia (2004). Wh-copying, phases and successive cyclicity. *Lingua* 114: 543–574.

Ferreira, Fernanda & Charles Clifton (1986). The independence of syntactic processing. *Journal of Memory and Language* 25: 348–368.

Fiebach, Christian, Matthias Schlesewsky & Angela Friederici (2002). Separating syntactic memory costs and syntactic integration costs during parsing: The processing of German wh-questions. *Journal of Memory and Language* 47: 250–272.

Fox, Danny (1999). Reconstruction, Binding Theory, and the interpretation of chains. *Linguistic Inquiry* 30: 157–196.

Frazier, Lyn (1978). On comprehending sentences: Syntactic parsing strategies. Ph.D. dissertation, University of Connecticut, Storrs.

Frazier, Lyn & Charles Clifton (1989). Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes* 4: 93–126.

Garnsey, Susan, Michael Tanenhaus & Robert Chapman (1989). Evoked potentials and the study of sentence comprehension. *Journal of Psycholinguistic Research* 18: 51–60.

Gazdar, Gerald, Ewan Klein, Geoffrey Pullum & Ivan Sag (1985). *Generalized Phrase Structure Grammar*. Oxford: Blackwell.

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

Gibson, Edward (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In: *Image, Language, Brain*, ed. by Alec Marantz, Yasushi Miyashita & Wayne O'Neill, Cambridge, MA: MIT Press, pp. 95–126.

Gibson, Edward & Neal Pearlmutter (1998). Constraints on sentence comprehension. *Trends in Cognitive Science* 2: 262–268.

Gibson, Edward & Tessa Warren (2004). Reading-time evidence for intermediate linguistic structure in long-distance dependencies. *Syntax* 7: 55–78.

Gordon, Peter, Randall Hendrick & Marcus Johnson (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27: 1411–1423.

Grodner, Daniel & Edward Gibson (2005). Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive Science* 29: 261–291.

Grodner, Daniel, Edward Gibson & Susanne Tunstall (2002). Syntactic complexity in ambiguity resolution. *Journal of Memory and Language* 46: 267–295.

Hawkins, John (1994). *A Performance Theory of Order and Constituency*. Cambridge: Cambridge University Press.

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

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

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

Just, Marcel Adam, Patricia Carpenter & Jaqueline Woolley (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology* 111: 228–238.

Kaan, Edith, Anthony Harris, Edward Gibson & Phillip Holcomb (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes* 15: 159–201.

Kaplan, Aaron (2005). Long-distance wh-movement in Chamorro. In: *Proceedings of the Twelfth Meeting of the Austronesian Formal Linguistics Association (AFLA XII)*, ed. by Jeffrey Heinz & Dimitris Ntelitheos, Los Angeles: UCLA, no. 12 in UCLA Working Papers in Linguistics, pp. 117–129.

Kayne, Richard & Jean Yves Pollock (1978). Stylistic inversion, successive cyclicity, and Move NP in French. *Linguistic Inquiry* 9: 595–621.

Kimball, John (1973). Seven principles of surface structure parsing in natural language. *Cognition* 2: 15–47.
King, Jonathan & Marcel Adam Just (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language* 30: 580–602.
King, Jonathan & Marta Kutas (1995). Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience* 7: 376–395.
Kiss, Katalin É. (1993). Wh-movement and specificity. *Natural Language and Linguistic Theory* 11: 85–120.
Kroch, Chien-Jer Charles LinAnthony (2006). Grammar and parsing: A typological investigation of relative-clause processing. Ph.D. dissertation, University of Arizona, Tuscon, AZ.
Kuznetsova, Alexandra, Per Bruun Brockhoff & Rune Haubo Bojesen 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.
Lebeaux, David (1988). Language acquisition and the form of the grammar. Ph.D. dissertation, University of Massachusetts, Amherst, MA.
Legate, Julie Anne (2003). Some interface properties of the phase. *Linguistic Inquiry* 34: 506–516.
Lewis, Richard (1996). Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research* 25: 93–115.
Lewis, Richard & Shravan Vasishth (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science* 29: 1–45.
Lewis, Richard, Shravan Vasishth & Julie Van Dyke (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Science* 10: 447–454.
Lin, Chien-Jer Charles & Thomas Bever (2006). Subject preference in the processing of relative clauses in Chinese. In: *Proceedings of the 25th West Coast Conference on Formal Linguistics (WCCFL 25)*, ed. by Donald Baumer, David Montero & Michael Scanlon, Somerville, MA: Cascadilla Press, pp. 254–260.
Marinis, Theodore, Leah Roberts, Claudia Felser & Harald Clahsen (2005). Gaps in second language sentence processing. *Studies in Second Language Acquisition* 27: 53–78.
McCloskey, James (1979). *Transformational Syntax and Model Theoretic Semantics: A Case Study in Modern Irish*. Dordrecht: Reidel.
McCloskey, James (2000). The morphosyntax of wh-extraction in Irish. *Journal of Linguistics* 37: 67–100.
McCloskey, James (2002). Resumptives, successive cyclicity, and the locality of operations. In: *Derivation and Explanation in the Minimalist Program*, ed. by Samuel David Epstein & T. Daniel Seely, Oxford: Blackwell, pp. 184–226.
McDaniel, Dana (1989). Partial and multiple wh-movement. *Natural Language and Linguistic Theory* 7: 565–604.
McElree, Brian, Stephanie Foraker & Lisbeth Dyer (2003). Memory structures that subserve sentence comprehension. *Journal of Memory and Language* 48: 67–91.
McElree, Brian & Teresa 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: 432–460.
Müller, Gereon (1999). Imperfect checking. *The Linguistic Review* 16: 359–404.
Nicol, Janet, Janet Fodor & David Swinney (1994). Using cross-modal lexical decision tasks to investigate sentence processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 20: 1229–1238.
Nicol, Janet & David Swinney (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research* 18: 5–19.
Omaki, Akira, Ellen Lau, Imogen Davidson White, Myles Louis Dakan & Colin Phillips (2012). Hyper-active gap filling: Pre-verbal object gap creation in English filler-gap dependency processing, Ms., John

Hopkins University, University of Maryland, University College London and Northeastern University.
Phillips, Colin (2006). The real-time status of island phenomena. *Language* 82: 795–823.
Phillips, Colin, Nina Kazanina & Shani Abada (2005). ERP effects of the processing and syntactic long-distance dependencies. *Cognitive Brain Research* 22: 407–428.
Pickering, Martin & Matthew Traxler (2003). Evidence against the use of subcategorization frequencies in the processing of unbounded dependencies. *Language and Cognitive Processes* 18: 469–503.
Pollard, Carl J. & Ivan A. Sag (1994). *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.
R Core Team (2014). R: A language and environment for statistical computing.
Rackowski, Andrea & Norvin Richards (2005). Phase edge and extraction. *Linguistic Inquiry* 36: 565–599.
Rizzi, Luigi (2006). On the form of chains: Criterial positions and ECP effects. In: *Wh-Movement: Moving On*, ed. by Lisa Cheng & Norbert Corver, Cambridge, MA: MIT Press, pp. 97–133.
Ross, John (1967). Constraints on variables in syntax. Ph.D. dissertation, MIT, Cambridge, MA, Published in 1986 as *Infinite Syntax!*, Ablex, Norwood.
Schulte im Walde, Sabine (1998). Automatic semantic classification of verbs according to their alternation behavior, MA thesis, Universität Stuttgart.
Staub, Adrian (2007). The parser doesn't ignore intransitivity, after all. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 33: 550–569.
Staub, Adrian (2010). Eye movement and processing difficulty in object relative clauses. *Cognition* 116: 71–86.
Staub, Adrian, Charles Clifton & Lyn Frazier (2006). Heavy NP shift is the parser's last resort: Evidence from eye movements. *Journal of Memory and Language* 54: 389–406.
Stevenson, Suzanne (1994). Competition and recency in a hybrid network model of syntactic disambiguation. *Journal of Psycholinguistic Research* 23: 295–322.
Stowe, Laurie (1986). Parsing wh-constructions: Evidence for on-line gap location. *Language and Cognitive Processes* 1: 227–245.
Sturt, Patrick, Martin Pickering & Matthew Crocker (1999). Structural change and reanalysis difficulty in language comprehension. *Journal of Memory and Language* 40: 136–150.
Sussman, Rachel & Julie Sedivy (2003). The time course of processing syntactic dependencies. *Language and Cognitive Processes* 18: 143–163.
Torrego, Esther (1984). On inversion in Spanish and some of its effects. *Linguistic Inquiry* 15: 103–129.
Traxler, Matthew, Robin Morris & Rachel Seely (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language* 47: 69–90.
Traxler, Matthew & Martin Pickering (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language* 35: 454–475.
Trueswell, John, Michael Tanenhaus & Christopher 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: 528–553.
Van Dyke, Julie & Richards 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: 285–316.
van Urk, Coppe & Norvin Richards (2015). Two components of long-distance extraction: Successive cyclicity in Dinka. *Linguistic Inquiry* 46: 113–155.
Vasishth, Shravan & Richard Lewis (2006). Argument–head distance and processing complexity: Explaining both locality and antilocality effects. *Language* 82: 767–794.
Wagers, Matthew & Colin Phillips (2009). Multiple dependencies and the role of the grammar in real-time

comprehension. *Journal of Linguistics* 45: 395–433.
Wagers, Matthew & Colin Phillips (2013). Going the distance: Memory and control processes in active dependency construction. *The Quarterly Journal of Experimental Psychology* : 1–31.
Wanner, Eric & Michael Maratsos (1978). An ATN approach to comprehension. In: *Linguistic Theory and Psychological Reality*, ed. by Morris Halle, Joan Bresnan & George Miller, Cambridge, MA: MIT Press, pp. 119–161.
Warren, Tessa & Edward Gibson (2002). The influence of referential processing on sentence complexity. *Cognition* 85: 79–112.