

Multiple dependencies and the role of the grammar in real-time comprehension¹

MATTHEW W. WAGERS

*Department of Linguistics
University of California, Santa Cruz*

COLIN PHILLIPS

*Department of Linguistics and Neuroscience & Cognitive Science Program
University of Maryland, College Park*

(Received 18 December 2007; revised 12 January 2009)

Wh-dependencies are known to be formed rapidly in real-time comprehension. The parser posits the location of gap sites in advance of the bottom-up evidence for missing constituents, and must therefore have a means of deciding when and where to project dependencies. Previous studies have observed that the parser avoids building ungrammatical *wh*-dependencies, for example, by restricting the search for gap sites from island domains. This paper tests the stronger claim that constraints are not merely respected, but that grammatical knowledge actively prompts the construction of some representations in advance of the input. Three self-paced reading experiments examined patterns of *wh*-dependency formation in multiple-dependency constructions: obligatory across-the-board (ATB) extraction from coordinated verb phrases, and from optional parasitic gaps in post-verbal adjunct clauses. The key finding is that comprehenders immediately enforce the requirement for extraction from coordinates, and hence actively search for multiple gap sites within a coordinate VP; but they do not search for post-verbal parasitic gaps. This difference cannot be attributed to relative differences in acceptability, as comprehenders rated both of these multiple-gap constructions equally highly, nor can it be explained by general parsing incentives to develop maximal incremental interpretations of partial strings. More plausibly, the difference reflects the deployment of detailed grammatical knowledge in a parser that is motivated to satisfy structural licensing requirements in real time.

I. INTRODUCTION

Grammars define the space of possible and impossible expressions in a language and thus capture a key component of linguistic knowledge. While it

[1] The authors acknowledge the support of grants from the National Science Foundation (BSC-0196004) and the Human Frontiers Science Program (RGY-0134). Thanks to Norbert Hornstein, Amy Weinberg, the audience of the CUNY Sentence Processing Conference (19), and the two anonymous *Journal of Linguistics* referees for helpful discussion and comment.

is generally agreed that introspective acceptability judgments are a product of lexical resources, structure-generating functions, and output constraints, grammatical models typically leave unspecified the relation of these components to the mechanisms that generate or recover linguistic representations in real time. However, that relation constitutes another crucial aspect of linguistic knowledge, and, in order to characterize it, we first must establish how tightly real-time mechanisms are linked to grammatical distinctions. The goal of this paper is to show that grammatical constraints actively direct the formation of *wh*-dependencies. We report the results of three on-line comprehension experiments that target this question by assessing whether real-time structure building is affected by (i) the necessity of across-the-board extraction in coordinate structures and (ii) the optionality of parasitic gaps. It has previously been observed in several domains that the processor avoids building ungrammatical representations in the resolution of syntactic dependencies (e.g. Stowe 1986, Sturt 2003), but we attempt here to substantiate a stronger claim. Specifically, grammatical constraints are not merely respected in real time; rather, grammatical knowledge can actively prompt the construction of representations in advance of the input.

2. THE ROLE OF THE GRAMMAR IN PROCESSING *WH*-DEPENDENCIES

2.1 *Background*

In overt movement languages like English, displaced *wh*-phrases are found in clause-edge positions where they establish scope properties, but they are assigned a thematic role within the clause in grammatically licensed positions. When the language processor encounters a *wh*-phrase, it must have a way of deciding when and where to link that phrase with its thematic role assigner. Making this decision is complicated by two aspects of unbounded dependencies. Firstly, the tail of a *wh*-dependency is (usually) only indirectly signaled by the input, for example, by the absence of a verb's subcategorized constituent. Secondly, there are numerous and diverse island constraints that restrict where *wh*-dependencies can terminate (Ross 1967; see Szabolcsi & den Dikken 2002 for a review). It is the existence of these constraints that makes *wh*-dependency formation a fruitful research target for assessing the role of grammatical knowledge in incremental comprehension.

In the past two decades, it has been established that the sentence processor attempts to form *wh*-dependencies before any direct evidence in the input signals the position of missing constituents (Crain & Fodor 1985, Stowe 1986, *inter alia*), a phenomenon commonly referred to as 'active filling' (Frazier & Flores D'Arcais 1989). Once the parser identifies a displaced element (henceforth, a *filler*), gaps are posited at each available position that

would allow the dependency to be completed.² For example, in (1), a *wh*-dependency exists between the underlined filler phrase and the verb *play*:

- (1) Which CD does the toddler like her mother to play __ before naptime?

A wide range of experimental findings indicates that, in the course of comprehending this sentence, speakers posit a direct object gap upon encountering the verb *like*. This gap assignment must then be revised upon reaching the pronoun *her*, and the ultimate gap assignment is made upon reaching the verb *play*.

This sequence of events has been established using several experimental paradigms. For example, previous work has compared strings like (1) to nearly identical ones that lack a *wh*-dependency – for example, (2a) versus (2b):

- (2) (a) The babysitter forgot which CD the toddler **likes** her mother to play __
 (b) The babysitter forgot whether the toddler **likes** her mother to play a CD ...

Word-by-word reading times have shown that a disruption begins at the constituent *her mother* in (2a) (Crain & Fodor 1985, Stowe 1986, Lee 2004). This processing disruption, called the Filled Gap Effect, suggests that comprehenders posit a gap in advance of an overt constituent. Convergent evidence comes from measures in which the semantic fit between a filler and a potential gap host (in bold) is manipulated, as in the following pair from Traxler & Pickering (1996):

- (3) (a) That's the pistol with which the heartless killer **shot** the hapless man ...
 (b) That's the garage with which the heartless killer **shot** the hapless man ...

These semantic fit or plausibility manipulations show that readers detect when a filler is an implausible argument of a verb while they are reading that verb or very shortly thereafter. A disruption is reflected either by a slowdown in reading times (Traxler & Pickering 1996; Phillips 2006), a deflection of a lexical-semantic evoked response potential (ERP) component, the N400 (Garnsey, Tanenhaus & Chapman 1989), or a sharply increased tendency to report that the sentence stops making sense (Tanenhaus, Stowe & Carlson

[2] We use the term 'gap' and the expression 'posit a gap' in a theory-neutral way, as is standard in the psycholinguistics literature. Our discussion and results do not depend on whether or not the tail of the dependency is a trace. This question has been the subject of periodic debate in psycholinguistic circles (McElree & Bever 1989, Nicol, Fodor & Swinney 1994, Pickering & Barry 1991, Gorrell 1993, Gibson & Hickok 1993). Current experimental techniques are most informative about the timing of dependency formation, but the timing facts are orthogonal to the representational hypothesis assumed (Phillips & Wagers 2007).

1985, Boland et al. 1995). These results show that comprehenders not only posit a dependency between a filler and a potential gap site, but also evaluate the semantic impact of their decision as soon as possible. They are strengthened by a wide array of findings using related methodologies and sampling a number of different languages.³

2.2 *Motivation for active dependency formation*

The parser approaches *wh*-dependency formation in a top-down manner and does not wait for bottom-up confirmation that a given phrase is actually missing any constituents.⁴ For this reason, the process is described as ‘active’. The active strategy can be seen as a way of responding to two of the properties of *wh*-dependencies: unboundedness and indirect evidence for the tail. However, the top-down strategy is not so clearly an inevitable feature of an efficient sentence processor. For example, it is likely to often lead to temporary misanalyses, as in (1). A bottom-up strategy that relies on more conclusive evidence for the position of missing constituents is a plausible alternative (Wanner & Maratsos 1978; cf. Fodor 1978). Explanations for the parser’s active approach must therefore appeal to more general incentives that prompt the parser to reach certain states as soon as possible.

One class of explanations, associated with principle-based parsing (e.g. Pritchett 1992, Weinberg 1992), identifies a pressure to satisfy grammatical licensing requirements as rapidly as possible. In the case of filler-gap dependencies, it has been suggested that there is a pressure to satisfy the Theta Criterion as soon as possible (e.g. Pritchett 1992, Aoshima, Phillips & Weinberg 2004). Until a structural relation can be established between the filler and a thematic role assigner, the parse is partially unlicensed, which is an undesirable state of affairs. This explanation can be framed in less grammaticized terms, under the assumption that the parser attempts to interpret as much of the sentence as it can as soon as possible (e.g. Altmann &

[3] CROSS-LINGUISTICALLY: Dutch (Frazier 1987, Frazier & Flores D’Arcais 1989, Kaan 1997), German (Schlesewsky, Fanselow, Kliegl & Krems 2000), Hungarian (Radó 1999), Italian (de Vincenzi 1991), Japanese (Aoshima, Phillips & Weinberg 2004), Russian (Sekerina 2003).

CROSS-METHODOLOGICALLY: electrophysiology using EEG (Garnsey et al. 1989, Kaan, Harris, Gibson & Holcomb 2000, Phillips, Kazanina & Abada 2005) and MEG (Lau, Yeung, Hashimoto, Braun & Phillips 2006), the ‘stops making sense’ task (Tanenhaus, Stowe & Carlson 1985, Boland et al. 1995), eye-tracking (Traxler & Pickering 1996), cross-modal lexical priming (Nicol & Swinney 1989, Nicol, Fodor & Swinney 1994), anticipatory eye movements (Sussman & Sedivy 2003).

[4] Information about the verb’s argument structure and facts about its usage can affect the comprehender’s decisions about gap placement. For instance, how likely a verb is to be transitive may change the likelihood with which the parser posits a gap in object position at the verb (Fodor 1978, Tanenhaus, Stowe & Carlson 1985; cf. Pickering & Traxler 2003). Crucially, though, decision-making is guided by expectations about the input, whether they be structural or lexical.

Kamide 1999, Sedivy, Tanenhaus, Chambers & Carlson 1999). The active strategy, under this view, is simply one manifestation of the parser's efforts to derive an interpretation from only partial information. By actively completing a *wh*-dependency the parser can produce a more informative interpretation from the limited input available. Finally, architectural approaches identify a pressure to keep dependencies as short as possible (Gibson 1998), as open dependencies may tax processing resources (e.g. Wanner & Maratsos 1978). Actively completing *wh*-dependencies is therefore potentially a strategy for reducing processing load.

Multiple possible accounts have been offered to explain the parser's active behaviour. The existing data, however, make it difficult to tell these accounts apart. Almost all studies to date have focused on single dependencies, and for single, open dependencies, these various pressures usually operate in the same direction. Correspondingly it is difficult to assess the contribution of the separate accounts, and thus whether grammatical knowledge plays any positive role in motivating active dependency formation.

2.3 *Islands*

Island constraints provide a natural testing ground for the question of whether grammatical knowledge plays any role in active dependency formation, because island constraints restrict the kinds of dependencies that can be formed in ways that are potentially independent of constraints on interpretation or processing. For instance, the example sentence in (4) contains a complex noun phrase in subject position. It is impossible for the filler phrase in the main clause to terminate in the NP-contained clause:

- (4) *Which babysitter did [_{NP} the revelation [_S that the toddler tormented]] frighten her mother?

Given the pressure to complete the dependency headed by *which babysitter*, the question arises whether it would ever be linked to a position inside the complex NP (i.e. with *tormented*). The notion expressed by linking the filler *which babysitter* with the verb *tormented*, as in (4), is plausible and perhaps even a likely state of affairs. Furthermore, as the discussion of (1) illustrates, in non-island domains the parser seems willing to make some mistakes and revise temporary commitments. However, linking *which babysitter* with the verb *tormented* can never turn out to be the right analysis, because of the constraint on constructing dependencies inside complex NPs. Therefore, showing that parsers do not engage in active dependency formation inside island domains would in principle constitute strong evidence that grammatical knowledge guides parsing decisions.

A number of studies have shown that experimental measures of active dependency formation are not observed in island domains (Stowe 1986, Bourdages 1992, Pickering, Barton & Shillcock 1994, Traxler & Pickering

1996; cf. Freedman & Forster 1985, Kurtzman & Crawford 1991), and many share the consensus that island constraints are respected in incremental processing (Phillips 2006). However, this position is vulnerable, exactly because the typical empirical consequence of respecting an island constraint is the absence of evidence that a dependency was ever entertained in that island.⁵ The strength of conclusions that can be drawn from a lack of evidence has raised concerns, particularly because some island domains have been argued to themselves constitute complex processing environments, whether a dependency is present or not (e.g. Deane 1991, Kluender 2004). Therefore null findings in island processing are liable to alternative interpretations that are unrelated to the parser's interaction with the grammar.

2.4 *Multiple dependencies: coordinate structures and parasitic gaps*

This article presents a new argument that grammatical knowledge plays a definite role in the active formation of *wh*-dependencies, and an argument that does not suffer from the null-effect logic of previous studies on islands in language processing. Instead of considering island constraints that absolutely restrict the formation of a *wh*-dependency inside a certain domain, we consider a related constraint on *wh*-dependency formation, the Coordinate Structure Constraint, in which extractions are permitted in certain cases. The Coordinate Structure Constraint (CSC; Ross 1967) rules out gaps within coordinate structures (5a–b), except in the case of across-the-board extraction, where one gap must occur in each coordinated phrase (5c).⁶

- (5) Phil generally dislikes the poetry ...
- (a) *that *The New Yorker* reviews authors or publishes __
 - (b) *that *The New Yorker* reviews __ or publishes interviews
 - (c) that *The New Yorker* reviews __ or publishes __

[5] Three EEG studies have demonstrated a processing disruption when the search for a gap encounters the boundary of an island domain. This disruption is reflected in a particular evoked response potential (ERP; P600: McKinnon & Osterhout 1996; LAN: Kluender & Kutas 1993, Neville, Nicol, Barss, Forster & Garrett 1991). However, these ERPs are observed at island boundaries. Therefore they are not informative about whether the parser attempts to construct a dependency, only whether an island domain is noticed. Moreover, the observed ERPs are also sensitive to processing difficulty. Consequently, while these results may reflect calculation of ill-formedness in a formal account of island constraints, they may equally well reflect increased complexity.

[6] There are several well-known classes of exceptions to this generalization (Goldsmith 1985, Lakoff 1986), as, for example, in:

- (i) What did you go to the store and buy __?
- (ii) How much can Josh drink __ and still stay sober?

These exceptions occur in specific circumstances when certain narrative relationships hold between the coordinates. All materials used in our studies were designed so as to avoid these contexts. See Postal (1998) for further discussion of these environments.

If the parser is guided by the Coordinate Structure Constraint, then there should be evidence that a second gap is actively posited in the second coordinate. Since this evidence would be positive, and not a null effect, it could thereby avoid the concerns raised by previous island studies.

The presence of continued active dependency formation in coordinate structures could be explained by the real-time application of the CSC, but it could also be explained by a mechanism that favors maximal interpretation. If a subsequent verb in a coordinate structure can take the filler as its argument, then it can satisfy its interpretive needs earlier than by waiting for an argument phrase. Continued active dependency formation in coordinate structures could simply reflect the pressure of interpreting as much of the partial expression as is possible. Therefore we consider a second kind of multiple-dependency construction, a parasitic gap inside a post-verbal adjunct clause (Engdahl 1983). Single extractions from a post-verbal adjunct clause are generally unacceptable (6a). In the presence of an extraction from direct or indirect object, the post-verbal adjunct clause can support an additional gap (6b), but, crucially, it is optional (6c).

- (6) Phil generally dislikes the poetry ...
- (a) *that *The New Yorker* reviews authors without publishing __
 - (b) that *The New Yorker* reviews __ without publishing __
 - (c) that *The New Yorker* reviews __ without publishing too much detail

A comparison of *wh*-dependency formation in coordinate structures and in parasitic gap environments, like post-verbal adjunct clauses, is therefore informative. If dependencies are actively formed in both domains, this suggests that an incentive for maximal incremental interpretations motivates the parser. Such findings would fail to isolate the role that grammatical principles may play, since the results would not reflect any grammatical distinctions. If, on the other hand, CSC environments exhibit active dependency formation, but parasitic gap environments do not, then a principle of incremental interpretation would be insufficient. If this is the case, then it must be that an important grammatical constraint is reflected in parsing routines.

Table 1 outlines the three candidate patterns of active dependency formation that might be observed in multiple-dependency constructions, and the conclusion that could be drawn from each. Firstly, it is entirely possible that active dependency formation ceases once a single, verified dependency is constructed. In this case, active dependency formation should not be observed either in second coordinates or in post-verbal adjunct clauses. We call this prediction 'ENTIRELY FILLER DRIVEN', since it corresponds to a parser that is driven solely by the requirements of the filler. We assume in this case that once the parser establishes a gap site or grammatical role for the filler, then active search is terminated. Secondly, active dependency formation might be observed in both the context of a second coordinate and of a post-verbal

ACTIVE DEPENDENCY FORMATION EXPECTED? <i>Active dependency formation principle</i>	<i>Verb position</i>	
	Second coordinate	Adjunct clause
Entirely filler driven:	×	×
Interpret or license displaced filler		
<u>Maximal Interpretation:</u>	✓	✓
Interpret as much as possible		
<u>Grammatical Licensing:</u>	✓	×
Satisfy grammatical constraints		

Table 1

Predictions for active dependency formation in multiple-dependency constructions.

adjunct clause. This prediction is called ‘MAXIMAL INTERPRETATION’, since it is what would be expected if active dependency formation reflects the parser’s drive to interpret as much of an expression as possible. Finally, active dependency formation might be observed only in second coordinates, and not in post-verbal adjunct clauses. This prediction is called ‘GRAMMATICAL LICENSING’, since it suggests that knowledge about the distinction between coordinate gaps and parasitic gaps affects parsing decisions.

A fourth logical possibility is that active dependency formation only persists in parasitic gap environments. This outcome seems unlikely, and it would be puzzling, as it would imply that the parser undertakes the effortful task of constructing an optional dependency, but not an obligatory dependency.

One previous study addresses the real-time status of the Coordinate Structure Constraint. Pickering et al. (1994) used a Filled Gap Effect design to compare sentences like:

- (7) (a) I know what you *hit* the cupboard and *broke* the mirror **with** __
 (b) I know that you *hit* the cupboard and *broke* the mirror with a ball

In sentence (a), the filler *what* is the argument of the preposition *with*, but there are two predicates that intervene between filler and gap: *hit* and *broke*. In self-paced reading, Pickering et al. (1994) found that reading times were elevated at the determiner following *hit* in (a), compared to a control sentence (b) with no *wh*-dependency. No such effect, however, was observed following *broke*. One interpretation of these data is that the parser attempted to form a dependency with *hit*, but then had to retract this analysis because there was an overt direct object. However, recognizing that it was inside a coordinate VP, the parser did not then attempt to form a dependency with

broke, since doing so would have violated the CSC. This study thus suggests that the parser can exhibit real-time sensitivity to the CSC by refraining from entertaining an analysis that violates the CSC. However, the signal to apply the grammatical constraint in this experiment comes in the form of a parsing failure followed by the coordinator *and*. Once the initial object gap site has failed, the reader might not be expected to figure out where it is possible to actively resolve the *wh*-dependency. For these reasons, the present study seeks to find evidence for application of the CSC that comes from a positive measure and to use a design where the signal to apply the constraint follows a successfully constructed representation.

2.5 *The present experiments*

In a series of three self-paced reading experiments, we address the issues outlined above: (1) whether comprehenders persist in actively constructing filler-gap dependencies after one filler-gap dependency has been successfully constructed; (2) whether comprehenders do this only for obligatory dependencies, as in coordinate extractions, or whether they search for all possible subsequent gap sites. Our results suggest that with coordinate extraction, comprehenders immediately enforce the requirement of across-the-board extraction from the coordinates by actively constructing a subsequent dependency. However, in potential parasitic gap environments, comprehenders do not search for further gaps. A decision-making strategy for comprehension that uses the grammar to actively and incrementally license structural analyses is both capable of explaining these results and consistent with the previous literature.

3. EXPERIMENT I

The goal of Experiment I was to test whether or not active dependency formation is operative in second coordinate phrases and parasitic gap environments. To do this, we created sentences containing object extractions from an initial VP, where active dependency formation is uncontroversial. Beyond the first gap, sentences had two possible continuations: (i) a coordinate VP, in which case a second gap is obligatory, as in (8); or (ii) an adjunct clause that could host a parasitic gap, in which case a second gap is optional, as in (9). As our index of active dependency formation, we manipulated the semantic fit of the filler with the second verb.

3.1 *Method*

3.1.1 *Participants*

Thirty-seven native speakers of American English from the University of Maryland community were paid \$10 to participate in an experimental session

lasting 50 minutes. All were naive to the purpose of the experiment and gave informed consent.

3.1.2 *Materials*

Experimental materials consisted of 24 sets of 4 conditions organized in a 2×2 factorial design that independently manipulated the conditions VP STRUCTURE and PLAUSIBILITY. These sentences followed the scheme in (8) and (9), each containing a relative clause with two verbs. The second verb constitutes the critical region, where active dependency formation is tested.

(8) *Coordinated VP, Plausible*

- (a) The wines which the gourmets were energetically **discussing** __ or slowly **sipping** __ during the banquet were rare imports from Italy.

Coordinated VP, Implausible

- (b) The cheeses which the gourmets were energetically **discussing** __ or slowly **sipping** __ during the banquet were rare imports from Italy.

(9) *Adjunct Clause, Plausible*

- (a) The wines which the gourmets were energetically discussing __ before slowly **sipping** the samples during the banquet were rare imports ...

Adjunct Clause, Implausible

- (b) The cheeses which the gourmets were energetically discussing __ before slowly **sipping** some wine during the banquet were rare imports ...

The semantic fit of the filler with the first verb was plausible across all conditions, so that processing would not be disrupted before the critical region. For example, in (8)–(9), one can equally well discuss wines or cheeses. The factor *PLAUSIBILITY* manipulated the semantic fit of the filler as a potential direct object of the second verb only: *Plausible* ((8a)–(9a)) or *Implausible* ((8b)–(9b)). In this example materials set, it is natural and plausible to sip wines, but not to sip cheeses. This manipulation provides a measure of dependency formation, as a slowdown is expected for implausible verb–argument combinations (e.g. Traxler & Pickering 1996). If this slowdown occurs at the verb, before any signal in the input that there is a missing constituent, then we can conclude that dependency formation is active.

The VP STRUCTURE factor manipulated the structure that contains the second verb, specifying two conditions: *Coordinate VP* vs. *Adjunct Clause*. This manipulation allowed us to compare evidence for active dependency formation in contexts that require second gaps, and those that merely allow second gaps, as outlined in table 1. *Coordinate VP* sentences always contained direct object gaps in the second VP, as required by the Coordinate Structure Constraint. The *Adjunct Clause* sentences provided environments

where a gap might be anticipated, but they did not actually include a parasitic gap (p-gap). Thus when the initial VP was followed by an adverbial clause, the parser could be lured to a p-gap analysis, but such an analysis was not confirmed. This design permitted the identification of effects due to active dependency formation, rather than bottom-up, gap-driven processing. While no p-gaps were present in our experimental target materials, 12 filler sentences were included that contained well-formed p-gaps. Since the 12 experimental items contained p-gap-hosting environments but no p-gap, the chance of finding a p-gap in a licit post-verbal position was 50% experiment-wide. There would therefore be no implicit cue within the experiment either to always expect p-gaps in p-gap positions or to only expect gaps in non-p-gap positions.

As discussed in footnote 6 there are several classes of apparent exceptions to the CSC, e.g. *What did you go to the store and buy ___?*, which involve expressions of purpose, outcome, and temporal contiguity. Since these environments seem most felicitous under conjunction with *and*, three different coordinators were used – *and*, *but*, and *or* – equally balanced across the materials to mitigate the potential confound whereby comprehenders might believe that they are in one of the CSC-exempt environments. Additionally, *and* conjuncts were constructed to avoid CSC-exempt construals by coordinating events that seemed equally felicitous in either order of mention. An additional (acceptability) rating study, reported below, shows that participants were not treating the materials as CSC-exempt. We fully balanced closed-class lexical items in two other ways: in the adjunct conditions, four prepositions were used: *while*, *without*, *before*, and *after*. Across materials, both the relative pronouns *who* and *which* were used. Analyses of the results showed that these various lexical differences had no effect on reading times, and therefore we do not discuss these manipulations further.

There were two additional constraints on the materials. First, both VPs contained adverbial modifiers before the verbs. These were included to provide a strong cue for the upcoming verb and thus allow participants sufficient time to recognize the CSC environment. Secondly, adjunct clause verbs were in the past progressive form, since simple past forms less readily host parasitic gaps. We matched both these features across all conditions, for comparability, and across both VPs, for parallelism.

The experimental materials for this and subsequent experiments are available on the *Journal's* website as Supplementary Materials accompanying the present article (see journals.cambridge.org/lin → vol. 45.2).

Seventy-two filler sentences were included. In order to prevent participants from recognizing the target structures, the fillers included syntactic features characteristic of the target items, such as progressive morphology, coordinate structures, and filler-gap dependencies. In addition, six anomalous predicate-argument combinations were included.

3.1.3 *Acceptability rating study*

In order to verify the generalization that multiple dependencies are necessary in coordinate structures but optional in parasitic gap environments, we conducted an off-line rating study. 32 participants who did not take part in the on-line study completed an acceptability questionnaire using a five-point scale, with '5' being the most acceptable and '1' the least acceptable. These participants either received extra credit in an introductory linguistics class, or payment as part of another set of experiments.

Half of the target item sets from the on-line study were included in this study. The sentences were minimally-modified versions of the on-line items from the plausible conditions, crossing the structural role of the second VP, i.e. as *Coordinate VP* or as *Adjunct Clause VP*, with the presence of a gap in the second VP.

- (10) The wines which the gourmets were energetically discussing __ ...
- (a) *Coordinate VP, Gap*
... or slowly sipping __ during the banquet were rare imports.
 - (b) *Coordinate VP, No Gap*
... or slowly sipping the samples during the banquet were rare imports.
 - (c) *Adjunct Clause VP, Gap*:
... before slowly sipping __ during the banquet were rare imports.
 - (d) *Adjunct Clause VP, No Gap*:
... before slowly sipping the samples during the banquet were rare imports.

The 12-item sets of these conditions were distributed by a Latin Square across four lists and combined with 12 filler items of similar length and complexity. Six fillers were uncontroversially acceptable sentences and six were highly unacceptable sentences. Each list was permuted in two pseudo-randomized versions.

Results for the experimental items are given in table 2. The average rating for uncontroversially acceptable filler items was 4.3 ± 0.09 (standard error), and 1.8 ± 0.07 for clearly unacceptable filler items. A repeated measures ANOVA showed a main effect of structure ($F_{1(1,31)} = 9.8$; MSE: 1.3; $p < 0.01$), a main effect of the presence of a second gap ($F_{1(1,31)} = 49.5$; MSE: 1.6; $p < 0.0001$), and, crucially, an interaction of the two factors ($F_{1(1,31)} = 11.4$; MSE: 1.3; $p < 0.01$). In the coordinate VP condition, ratings were substantially lower if there was no gap in the second coordinate ($p < 0.001$). This pattern confirms that participants were sensitive to the CSC. P-gap conditions were highly rated, and there was no difference between the two conditions with multiple gaps. This result mirrors an earlier finding for subject p-gap constructions (Phillips 2006), and neutralizes the potential concern that p-gaps are somehow marginal structures.

	GAP IN SECOND VP?		<i>C.I.</i>
	Yes	No	
Second VP			
Coordinate VP	4.2 ±0.1	2.9 ±0.1	0.49
Adjunct clause	4.2 ±0.1	3.7 ±0.1	0.35
<i>C.I.</i>	0.30	0.80	

Table 2

Experiment 1. Acceptability ratings summary.

Average ratings are given for each sentence type with standard error. 95% confidence intervals (C.I.) on mean ratings differences, across participants and items, are reported in the margins. N = 32.

Surprisingly, there was a moderate decline in ratings for adjunct clauses if no gap was present ($p < 0.05$). P-gap and non-p-gap materials were predicted to be equally highly rated, so it was unexpected that the gapless adjunct conditions were rated slightly lower than their p-gap analogues (average 4.2 vs. 3.7). It is worth emphasizing the small size of this effect relative to the large drop in ratings observed for CSC-violations (average 4.2 vs. 2.9). The mean difference, normalized against variance, between p-gap and non-p-gap continuations (Cohen’s *d*) was 0.4, much smaller than the corresponding difference in coordinate structures (*d*: 1.2). We suspect that this difference may reflect a bias in the construction of materials. P-gaps seem most felicitous when there is a close relation between the events or states expressed in the main and adjunct clauses. Creating a non-p-gap analogue in the adjunct clause meant inserting a theme argument that was necessarily closely related to the displaced theme of the main clause. This may have seemed to experimental participants, therefore, an awkward way to express an idea more naturally expressed by a p-gap or even a coordination. Irrespective of the cause of this small difference, this result strengthens the logic of the on-line study. If speakers fail to actively construct a second gap in the adjunct conditions, despite the high acceptability of a p-gap in this structure, this would show that active dependency formation is not merely motivated to derive a natural interpretation from the input, but interacts strongly with grammatical principles.

3.1.4 *Plausibility rating study*

At the conclusion of the on-line reading time study, 24 of the 37 participants were presented with a questionnaire containing 24 sentences, and asked to rate each on a five-point scale for plausibility, with ‘5’ being the most

plausible and ‘1’ the least plausible. These data were collected to confirm the effectiveness of the plausibility manipulation. The sentences were simple SVO clauses derived from the critical VPs of the on-line study (e.g. *The gourmets discussed the {wine/cheese}*, *The gourmets drank the {wine/cheese}*). The rating study used four conditions in a 2×2 design, crossing the two factors of filler type (*Plausible*, *Implausible*) and verb type (first vs. second VP in the target items). The four resulting conditions for the 24 item sets were distributed by a Latin Square across four lists, each of which was then permuted into two pseudo-randomized versions.

On-line materials had been designed such that all fillers should be plausible at the first verb position, but should differ in plausibility at the second verb position. The rating study results confirmed this manipulation, as there was a strong interaction between filler class and verb class ($F_{1(1,23)}: 70.6$; MSE: 2.3; $p < 0.0001$). Sentences containing first-position verbs were rated equally highly, regardless of whether the object corresponded to a *Plausible* or an *Implausible*-class filler (mean: 3.9, difference n.s.). Whereas the average rating for *Plausible*-class objects, as objects of the second-position verb, remained high and consistent with the first-verb ratings (mean: 3.9), the average rating for *Implausible*-class direct objects was much lower (mean: 1.8, $p < 0.001$). We conclude that the ‘filler class by verb position’ manipulation met the desired specifications.

3.1.5 Procedure

Sentences were presented using the Linger software package (Rohde 2003) on a PC in a masked, word-by-word fashion. Each sentence was followed by a yes/no comprehension question. Yes/no answers were counterbalanced across conditions. Participants were instructed that their goal was to answer as many questions correctly as possible, and that to do so it was important to read at a careful, comfortable pace that matched their natural reading rate as closely as possible.

3.1.6 Analysis

Self-paced reading times for experimental sentences were examined region-by-region. Sentences were aligned word-for-word up to the second verb, such that each ordinal word position corresponded to a separate region. Data were systematically treated against outliers. Exceptionally long reaction times, above 3500 ms, were removed. One participant was replaced in the design due to exceptionally slow mean reading times. All trials on which the comprehension question was answered incorrectly were excluded. Finally, subsequent to these exclusions, reading times beyond 3 standard deviations of the mean at each region and in each condition were excluded (affecting fewer than 1% of trials). Evaluation of statistical reliability was carried out

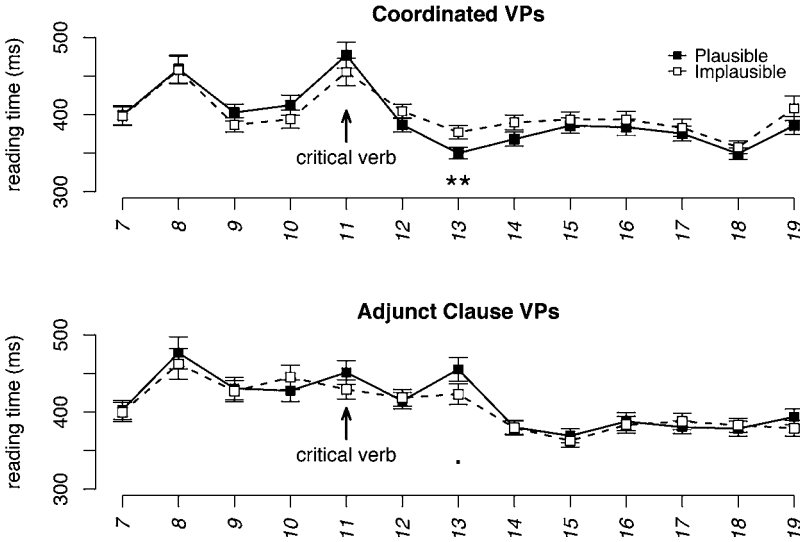


Figure 1

Experiment 1. Region-by-region reading times.

(The wines/cheese which the gourmets were)₁₋₆ energetically₇ discussing₈ ...

Coordinate: or₉ slowly₁₀ sipping₁₁ during₁₂ the₁₃ banquet₁₄ were₁₅ rare₁₆ imports₁₇ from₁₈ ...

Adjunct: before₉ slowly₁₀ sipping₁₁ (the samples)/(some wine)₁₂₋₁₃ during₁₄ the₁₅ banquet₁₆ ...

Region-by-region reading times from the onset of the second VP to 8 regions beyond the critical verb. The symbols that indicate the result of a pairwise by-participants RMANOVA are as follows: ** < 0.01 * < 0.05 <• < 0.10.

by repeated measures analysis of variance. Both participants and items analyses are presented in appendix tables. In the text, however, only the participants analysis is given (which has been argued to be the correct, sufficient test statistic for counterbalanced designs such as our own: Raaijmakers, Schrijnemakers & Gremmen 1999).

3.2 Results

Comprehension question accuracy for the target sentences was high (average: 88.8%) and did not differ reliably across conditions.

Figure 1 presents the region-by-region condition means for regions 7–19. The omnibus repeated measures ANOVA report is given in table A1 for Regions 7–15 (in the appendix). Region-by-region condition means and test results for Regions 1–6 are not reported, as materials did not differ across the structural conditions. Materials did differ across the plausibility conditions, as the filler was introduced in Region 2. However, no effect of filler type was observed in that region, or in any region before Region 8.

PRECEDING THE SECOND VERB. Unremarkably, there were no main effects or interactions at Regions 7–8, the adverb and verb of the first VP. Materials were constructed to be structurally identical in these regions; and the plausibility norming survey reported in the Materials section (section 3.1.4) confirmed that both types of fillers were equally plausible as the direct object of the verb in Region 8. Materials diverged structurally in Region 9, which consisted in a coordinator for Coordinated VPs and a preposition for Adjunct Clauses, and there was a reliable increase in reading times for Adjunct Clauses in this region and in Region 10, the second VP adverb.

THE SECOND VERB. No reliable effect of VP STRUCTURE, PLAUSIBILITY, or their interaction was observed at the critical second-VP verb (Region 11) in the participants analysis. A reliable main effect of PLAUSIBILITY was observed in the items analysis, due to plausible sentences being read slightly more slowly, although this was not reliable in pairwise comparisons.

SECOND VP POST-VERBAL REGION. In the Coordinated VP conditions, implausible sentences were read more slowly in all regions (12 and higher) subsequent to the critical verb, reaching significance two words downstream of the critical verb (Region 13). In Adjunct Clause sentences, no consistent effect of plausibility was observed. In the region two words beyond the critical verb, there was a reliable interaction of plausibility and VP structure, due to slower reading times for implausible-filler sentences in coordinate VP conditions, and an opposite tendency in adjunct VP conditions. Planned pairwise comparisons in this region revealed a highly reliable slowdown due to implausibility for Coordinated VPs ($F_{1(1,35)}: 7.87$; $MSE: 6329$; $p < 0.01$). The opposite reading time pattern was observed in Adjunct Clauses, but it was only marginally significant ($F_{1(1,35)}: 2.92$; $MSE: 36583$; $p < 0.10$). This latter tendency may reflect the fact that lexical materials differed at this position ('the samples' vs. 'some wine'; see (9a–b)).

3.3 *Discussion*

Two conclusions follow from the reading time results in Experiment 1. First, in Adjunct Clause VPs the lack of a slowdown due to implausibility suggests that dependency completion does not proceed actively in those environments. As no gap occurs in these conditions there is no bottom–up evidence to prompt dependency completion. Therefore, the plausibility comparison within this condition constitutes a fair test of purely active dependency completion both at the verb and in the following regions.⁷ Second, for

[7] An anonymous reviewer raises the concern that implausibility effects have been shown to ease later reanalysis (Pickering & Traxler 1998, 2003), leading to a 'cross-over' effect in which initial processing difficulty is followed by facilitation. If the difficulty and facilitation were measured in the same region in our experiment, and were of the same magnitude, then

Coordinated VPs, the presence of a reliable reading time slowdown for implausible fillers in the immediate post-verbal regions shows that filler–gap dependencies are constructed in the second VP. However, this finding does not provide definitive evidence for active dependency formation, because of the timing of the effect. By the time the effect becomes reliable, there is bottom–up evidence for a gap in the form of missing constituents, and therefore dependency completion could have been cued from the input. Nonetheless, since spill-over effects are commonly observed in self-paced reading, this slowdown could be an effect of active dependency completion at the verb.

If the effect in Coordinate VPs were indicative of active dependency completion, then we could conclude that the parser is sensitive to the Coordinate Structure Constraint, such that it recognizes when a filler that has already been successfully integrated with the first verb must participate in subsequent dependencies. This conclusion would be consistent with a grammatical licensing parser, in which active dependency formation is driven not by the need to discharge thematic roles as soon as possible, but rather by the need to satisfy grammatical requirements. On the other hand, if the observed plausibility effect in Coordinate VPs reflects non-active bottom–up processes, then the difference between the plausibility contrasts in Coordinated VPs and in Adjunct Clauses may have a more mundane explanation: there is a gap in one structure, but not the other. In order to determine whether active dependency completion in fact persists in coordinate VP environments, and thus tease apart the two possible interpretations of Experiment 1, Experiment 2 was designed in such a way that effects of spill-over and of bottom–up gap-detection would be well separated over the time course of reading.

4. EXPERIMENT 2

In Experiment 1, the closeness of the critical verb and bottom–up evidence for a gap led to an ambiguous result. In order to separate the verb from the gap, ditransitive verbs were used in Experiment 2, as illustrated in (II):

- (II) The adhesive coating that the engineer **sprayed** the special test surfaces with ___ in his new laboratory ...

In this example, the verb *spray* subcategorizes for two internal arguments. When the second argument is relativized, the regions immediately following the verb (the direct object regions) do not provide evidence for a gap site.

the net effect might be no difference between conditions. It is unlikely that this explanation applies to the p-gap comparisons, since we find robust plausibility effects elsewhere (Experiment 1 coordinate conditions, Experiments 2–3), and the size of the cross-over facilitation effect in previous studies on filler–gap comprehension is consistently smaller than the implausibility effect (Pickering & Traxler 2003).

In a semantic fit manipulation, a slowdown due to implausibility could be observed at or beyond the verb but before bottom-up evidence for a missing constituent. It is thus possible to avoid the confound seen in Experiment 1. If the slowdown occurs at the verb or in the direct object regions, we can conclude that the parser actively completed the dependency, since it had to project the gap site before the input unambiguously signaled its location.

In this experiment we used coordinate VPs in which the second verb participates in *spray/load*-type locative constructions. As there was no implausibility effect in the Experiment 1 Adjunct Clause conditions, there were no such conditions in this experiment. Instead, the coordinate VPs were compared with length-matched conditions with a single filler-gap dependency, in order to compare patterns of dependency formation in a second coordinate VP, on the other hand, with dependency completion in a single dependency, on the other; the latter of these was expected to be uncontroversially active.

4.1 *Methods*

4.1.1 *Participants*

Thirty-two native speakers of American English from the University of Maryland community were paid \$10 to participate in an experimental session lasting 50 minutes. All were naive to the purpose of the experiment and gave informed consent.

4.1.2 *Materials and procedure*

Experimental materials consisted of 24 sets of 4 conditions organized in a 2×2 factorial design that independently manipulated the factors VP STRUCTURE and PLAUSIBILITY. The sentences followed the scheme in (12) and (13), locating the filler-gap dependency within an object relative clause as in Experiment 1.

(12) *Coordinate VP, Plausible*

- (a) The adhesive coating that the talented engineer designed __ for his boss and methodically **sprayed** the special test surfaces with __ in his new laboratory could make the company lots of money.

Coordinate VP, Implausible

- (b) The computer program that the talented engineer designed __ for his boss and methodically **sprayed** the special test surfaces with __ in his new laboratory could make the company lots of money.

(13) *Single VP, Plausible*

- (a) The adhesive coating that the talented engineer from the high-tech aerospace firm methodically **sprayed** the special test surfaces with __ in his new laboratory could make the company lots of money.

Single VP, Implausible

- (b) The computer program that the talented engineer from the high-tech aerospace firm methodically **sprayed** the special test surfaces with ___ in his new laboratory could make the company lots of money.

The VP STRUCTURE factor manipulated the VP structure containing the critical verb. *Coordinate VP* sentences contained two coordinated VPs, as in Experiment 1. The critical verb was the second verb in the coordinate. *Single VP* sentences contained only a single verb. A five-word PP modifier was attached to the relative clause subject in this condition, so that the ordinal position of the critical verb would match the position of the second verb in the Coordinate VP conditions.

The PLAUSIBILITY factor manipulated the semantic fit of the filler with respect to the critical verb by creating two classes of fillers. *Plausible* fillers were plausible as the direct object of both the first and second verb ((12a)–(13a)). *Implausible* fillers were plausible as the direct object of the first verb, but implausible as the direct object of the second verb ((12b)–(13b)). As in Experiment 1, the plausibility manipulation provided a measure of dependency formation.

Syntactically alternating locative verbs in the *spray/load* class were selected for the critical verb in each item set (Anderson 1971, Fraser 1971, Pinker 1989, Rappaport & Levin 1986, inter alia). Verbs from this class tend to impose greater semantic restrictions on both of their arguments than do simple datives, like *give*, or benefactives, like *buy*. This feature made it feasible to design a large number of items with a semantic fit manipulation that would apply to the filler regardless of the internal position it occupied. It is important that the filler be implausible both as direct and as oblique object. Previous research has suggested that if a verb has multiple syntactic frames or argument positions, then fillers that are implausible solely as a direct object do not elicit a slowdown in filler-gap constructions (Boland et al. 1995). Representative verbs are given in (14), with arguments in the specified configuration:

- (14) splash/spray/sprinkle/spread NP₁ with NP₂

For these verbs, NP₁, referred to as the *ground* argument, must typically be a concrete entity, while NP₂, referred to as the *figure* argument, must typically be a liquid, plastic or particulate substance. There is nothing crucial about what the particular semantic selectional restrictions are, just that restrictions exist for both arguments in *spray/load* verbs. Twelve *spray/load* verbs were chosen from Levin (1993) as the critical verbs. Each verb was used in two item sets. The first argument of the critical verb was always four words long, providing a large region between the verb position and the first direct evidence for a gap position. Any slow-down due to implausibility observed within this spill-over region could be attributed to active dependency

formation, since direct evidence for the gap does not occur until after the subcategorized preposition *with*. The bottom-up cue for the gap was very strong, as two prepositions occurred in sequence (*sprayed the special test surfaces with in his new ...*).

Three further design constraints applied. As in Experiment 1, pre-verbal adverbs were used in all VPs. Unlike Experiment 1, all verbs appeared in the simple past tense form; the progressive verb forms used in Experiment 1 were required by the Adjunct Clause conditions included in that study, which were not present in Experiment 2. Finally, the complementizer *that* was used to signal the onset of the relative clause, instead of the pronouns *who/which*. Both the complementizer and the relative pronoun serve as effective signals to the parser for a relative clause, and plausibility effects are obtained in both environments (first author, unpublished pilot results). In the context of these materials, the complementizer *that* was judged to be more natural.

Seventy-two fillers were adapted from the fillers in Experiment 1, so that the distribution of sentence lengths in fillers matched the distribution of target items.

Participants read sentences in a self-paced reading task, with the same procedure as in Experiment 1.

4.1.3 *Plausibility rating study*

At the conclusion of the on-line study, 16 of the 32 participants were presented with a questionnaire containing 24 sentences, and were asked to rate each on a five-point scale for plausibility. The sentences were simple SVO matrix-clause versions of the relative clauses from the on-line study. The design was the same as in Experiment 1, with verb position (first or second verb) crossed with filler class (Plausible or Implausible) in a 2×2 factorial design.

Sentences containing first-position verbs were rated equally highly, regardless of whether the object corresponded to a *Plausible* or an *Implausible* filler (mean for both 4.2; difference n.s.). The average rating for *Plausible*-class objects as objects of the second-position verb remained high (mean: 3.7), but the average rating for *Implausible*-class direct objects was considerably lower (mean: 1.5, $p < 0.0001$).

We conclude that the ‘filler class by verb position manipulation’ met the desired specifications.

4.1.4 *Analysis*

Self-paced reading times were examined in the same region-by-region fashion as in Experiment 1. We followed the same exclusion criteria as above. One participant who failed to perform the task as instructed was removed

from further analysis. Unless otherwise noted, all analyses involved repeated measures ANOVAs.

4.2 Results

4.2.1 Question-answering accuracy

Question-answering accuracy was uniformly high. For the 24 experimental targets, accuracy was 92.3% overall. There were modest and reliable differences due to the VP STRUCTURE and PLAUSIBILITY manipulations. For coordinate structures, accuracy was 95.7% for plausible conditions and 90.3% for implausible conditions; for single VP controls, accuracy was 89.2% for plausible conditions and 94.1% for implausible conditions. A logistic mixed-effect model estimated that all factors had an odds ratio significantly different from zero. In contrast to Experiment 1, this model was significantly different than a null model that attributed all variation to participants, with no effects for each condition ($\chi^2(12)$: 2765.0; $p \sim 0$).

4.2.2 Reading times

Figure 2 presents the region-by-region condition means, segregated into two pair-wise comparisons for regions 6–20. Regions 6–20 extend from the lexical offset of the embedded subject head noun (here, *coating* or *program*) and continuing up to three words beyond the preposition heading the critical verb's figure argument. The omnibus repeated measures ANOVA report is given in table A2 for Regions 6–20. Materials did not differ in Regions 1–5, up to the lexical offset of the subject head noun, apart from the filler manipulation. No reading time differences were observed in those regions.

PRECEDING THE SECOND VERB. Materials diverged at the offset of the subject head noun: for Coordinate VP sentences, a verb followed the subject noun, while for Single VP sentences, a PP followed the subject noun. Accordingly, reading times differed across these two conditions as a function of VP STRUCTURE, beginning three words downstream from the subject head noun, in Region 8, and persisting until nine words downstream, in Region 14.

THE SECOND VERB. At the second verb (Region 12), in addition to the effect of VP STRUCTURE described above, there was a clear effect of PLAUSIBILITY, due to slower reading times for Implausible filler conditions. However, Coordinate VP sentences showed this contrast most strongly, with a variance-normalized mean difference of 0.18, compared to 0.04 for Single VP sentences. Moreover, as figure 2 shows, the plausibility contrast was robust in pairwise comparisons for Coordinate VPs ($F_1(1,30)$: 5.63; MSE: 12782; $p < 0.05$) but not for Single VPs ($F_1(1,30) < 1$). However, the interaction of PLAUSIBILITY and VP STRUCTURE was not significant in this region ($F_1(1,30) < 1$).

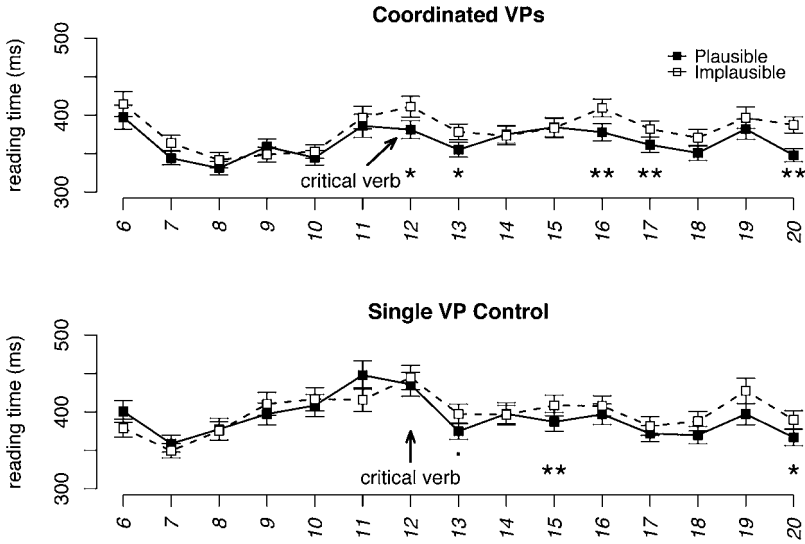


Figure 2

Experiment 2. Region-by-region reading times.

(The adhesive coating/computer program that the talented engineer)₁₋₅ ...

Coordinate: designed₆ for₇ his₈ boss₉ and₁₀ methodically₁₁ sprayed₁₂

Single VP: from₆ the₇ high-tech₈ aerospace₉ firm₁₀ methodically₁₁ sprayed₁₂
the₁₃ special₁₄ test₁₅ surfaces₁₆ with₁₇ in₁₈ his₁₉ laboratory₂₀ ...

Region-by-region reading times from the offset of the subject head noun to 8 regions beyond the critical verb, with example text for each region. The symbols that indicate the result of a pairwise by-participants RMANOVA are as follows:

p: ** < 0.01 < * < 0.05 < • < 0.10.

THE GROUND ARGUMENT REGION. In the ground argument regions following the critical verb (Regions 13–16; ‘the special test surfaces’ in figure 2), we observed persistent effects of plausibility, especially at the determiner in Region 13 and at the noun in Region 16. The slow-down due to implausibility was of comparable size for both coordinate VPs and long single VPs in Region 13 (in raw milliseconds). The effect was reliable in pairwise comparisons for Coordinate VPs ($F_{1(1,30)}: 7.19$; MSE: 6367; $p < 0.05$), but only marginally so for Single VPs ($F_{1(1,30)}: 3.68$; MSE: 15461; $p < 0.10$). Region 14 showed no effect of plausibility. Region 15 showed an effect only for Single VPs ($F_{1(1,30)}: 7.59$; MSE: 7872; $p < 0.01$). Region 16 showed a strong effect of implausibility for Coordinate VPs ($F_{1(1,30)}: 10.36$; MSE: 7240; $p < 0.01$), and a much weaker, and unreliable, effect for Single VPs ($F_{1(1,30)}: 2.50$; MSE: 11844; $p < 0.15$).

THE GAP REGION. An effect of plausibility was present in Regions 17–20, which corresponded to the regions containing the preposition that selects the gap site and subsequent regions. Since Regions 17–18 consisted of two

prepositions in sequence (here *with in*) they provided clear evidence of a missing constituent. Once again, the size and location of specific effects differed slightly across structural conditions, with Coordinate VP sentences showing contrasts in more regions, and displaying the largest contrast in Region 20.

4.3 Discussion

The reading time results from Experiment 2 showed that an implausible filler led to a slowdown at the second verb in a Coordinate VP structure and in subsequent regions in its argument field. The timing of this slowdown provides evidence that the second gap in coordinate structures is constructed actively, because the slowdown occurs unambiguously before the direct evidence of a gap position in the second coordinate. The use of *spray/load* verbs made it possible to put sufficient distance between the verb and the gap position, so that we can confidently interpret the slowdown as an effect of active dependency formation.

In Experiment 2, in contrast to Experiment 1, the effect of filler-verb plausibility appeared on the verb itself, and not one or two words downstream. There were differences in the experimental materials that may explain why the effect emerged earlier in Experiment 2. One important difference is that in Experiment 1 the second verb occurred only two words after the first verb, whereas in the Experiment 2 materials an additional three-word PP occurred in the first VP that further separate the two verbs. To see why this could make a difference, consider that in order to detect implausibility comprehenders must not only posit a gap location, but must also retrieve and integrate the filler syntactically and semantically. Even when there is no disruption in dependency formation many comprehension processes must presumably take place at a verb. It is more likely that these processes could have persisted to the second verb in Experiment 1 than in Experiment 2, and thus could have delayed the emergence of an implausibility effect by one or two words.

An unexpected pattern of results in this experiment was that implausible fillers appeared to have a more disruptive effect on processing in a second coordinate than in the single dependency control conditions. The single dependency controls showed word-by-word disruptions that were fewer, smaller, and later. In particular, while we observed three reliable disruptions in the coordinate conditions (Regions 12, 13, Region 16), we observed only one reliable disruption in the single VP condition, in a later Region (15), accompanied by two marginal ones (Regions 13, 16). One might have expected the single VP condition to show an equivalent or stronger response to the implausibility in light of the general robustness of active dependency formation effects. On these grounds, then, perhaps these results constitute evidence that active dependency formation is very robustly re-engaged by coordinate structures.

However, one factor that must be considered is the presence of the subject-adjoined PP, which was inserted in order to control the serial distance between the overt filler at the relative clause boundary and the verb that hosts the gap site. The concern for matching dependency lengths across conditions was originally motivated by evidence that the effectiveness of filler-gap resolution can be highly length-sensitive, though this is usually seen most robustly across clauses (McElree, Foraker & Dyer 2003, Phillips, Kazanina & Abada 2005). However, it may be that the relevant controlling factor is not the distance between the lexical occurrence of the filler and its associated gap site, but the time elapsed since the filler was last processed. In the case of a second-coordinate dependency, the filler has been processed very recently, in resolving the dependency in the first coordinate. Conversely, in the single dependency control conditions the filler was probably last activated in the processing system no later than the onset of the subject phrase. This account predicts that if the subject-adjoined PP is removed, then a robust active dependency formation effect should once again emerge. Since it was unexpected that active dependency formation should truly be more reliable in coordinate structures than in a single dependency, in Experiment 3 we tested the hypothesis that it was the presence of the subject-adjoined PP that led to the weaker response in the Experiment 2 control conditions.

5. EXPERIMENT 3

In Experiment 2, the second VP in a coordinate structure gave rise to a more robust implausibility effect than did a length-matched single dependency. In the single dependency condition a PP modifier was included to match the ordinal position of the critical verb. Experiment 3 tested the idea that it was the added length of the PP modifier that dampened the implausibility response. Therefore, in a simple manipulation, the subject-adjoined PP was removed, and the strength of the implausibility effect was compared to the same materials with the PP included.

While the outcome of this experiment was not expected to directly affect the logic of our conclusions regarding the role of grammatical constraints tested in Experiments 1–2, it was necessary to clarify the unexpected failure to observe robust active dependency in the Single VP condition. In particular, we wanted to establish an independent reason why the control condition should show a relatively weaker response to the plausibility manipulation than the coordinate condition, so that we could firmly conclude that the coordinate condition was exhibiting sustained active dependency formation. However, this experiment could also provide insight into the nature of the information used to complete dependencies, by testing the novel hypothesis that, in multiple dependencies, the linking of the filler to the first gap site increases its reliability of processing at subsequent sites.

5.1 *Method*5.1.1 *Participants*

Twenty-four native speakers of American English from the University of Maryland community were paid \$10 to participate in an experimental session lasting 50 minutes. All were naive to the purpose of the experiment and gave informed consent.

5.1.2 *Materials, procedure and analysis*

Experimental materials consisted of 24 sets of 4 conditions organized in a 2×2 factorial design that independently manipulated the factors DEPENDENCY LENGTH and PLAUSIBILITY. Experimental materials followed the scheme in (15) and (16).

(15) *Short, Plausible*

- (a) The adhesive coating that the talented engineer methodically **sprayed** the special test surfaces with __ in his new laboratory could make the company lots of money.

Short, Implausible

- (b) The computer program that the talented engineer methodically **sprayed** the special test surfaces with __ in his new laboratory could make the company lots of money.

(16) *Long, Plausible*

- (a) The adhesive coating that the talented engineer from the high-tech aerospace firm methodically **sprayed** the special test surfaces with __ in his new laboratory could make the company lots of money.

Long, Implausible

- (b) The computer program that the talented engineer from the high-tech aerospace firm methodically **sprayed** the special test surfaces with __ in his new laboratory could make the company lots of money.

The factor DEPENDENCY LENGTH manipulated the serial distance between filler and gap. The Long dependency sentences were identical to the corresponding conditions in Experiment 2. The Short dependency sentences were derived from the Long dependency sentences by removing the subject-adjoined PP. Dependency Length here is operationally defined as the number of words between the introduction of the filler and the verb. The PLAUSIBILITY factor involved exactly the same manipulation as in Experiment 2. Seventy-two fillers were adapted from the fillers in Experiment 2 so that the distribution of lengths of fillers matched the distribution of lengths of target items.

The experimental procedure was identical to Experiment 2. Data treatment and analysis were the same as in Experiments 1 and 2. No participants were excluded.

5.2 Results

Question-answering accuracy was uniformly high. For the 24 experimental targets, accuracy was 89.4%. There were no reliable differences among conditions (logistic mixed-effects estimate, comparison to null model: $\chi^2(7): 4.40; p = 0.49$).

In both Long and Short dependencies, a slow-down due to implausibility appeared in the first word of the ground argument. As this effect occurred well before direct evidence for the gap, we interpret it as an effect of active dependency formation. In short dependencies, slower reading times persisted for Implausible filler sentences throughout the ground argument region and into the gap region, whereas for long dependencies the effect of implausibility was observed only once in the ground argument regions and then not again until the gap region. The results in these conditions were thus similar to the results observed in the Experiment 2 Single VP condition.

Figure 3 presents the region-by-region condition means from the beginning of the sentence until region 20, segregated into two pair-wise comparisons. The omnibus repeated measures ANOVA report is given in table A3 (in the appendix). Region-by-region condition means and test results for preceding regions did not differ (when the unmatched PP regions of the Long condition were excluded).

At the adverb preceding the verb (Region 6) there were no reliable main effects or interactions, although there was a non-significant tendency for long dependency conditions to be read more slowly. On the critical verb itself (Region 7), there were also no reliable effects or interactions.

In the ground argument regions (8–11) following the critical verb, plausibility effects were found in short and long conditions alike. However, the effect of plausibility was more long-lasting in the short dependencies. In the ground argument determiner region (Region 8), Implausible sentences were read more slowly in both Long and Short dependency conditions. The size of the effect in long and short dependency conditions alike was consistent, but it was reliable in pair-wise comparisons only for the Short dependency conditions. No implausibility effect in the second and third words of the ground argument (Regions 9–10), although Short dependency conditions consistently displayed slower reading times for Implausible sentences. In the final word of the ground argument region (Region 11), corresponding to the head noun, Short dependencies reliably showed a large slowdown due to implausibility, whereas Long dependencies did not.

Short dependency conditions displayed sensitivity to the plausibility of the filler one word downstream from the figure-argument preposition (Region 13), whereas Long dependency conditions did not. At the immediately following region, however, there was a reliable slowdown due to implausibility for both Long and Short dependencies. In the third subsequent region

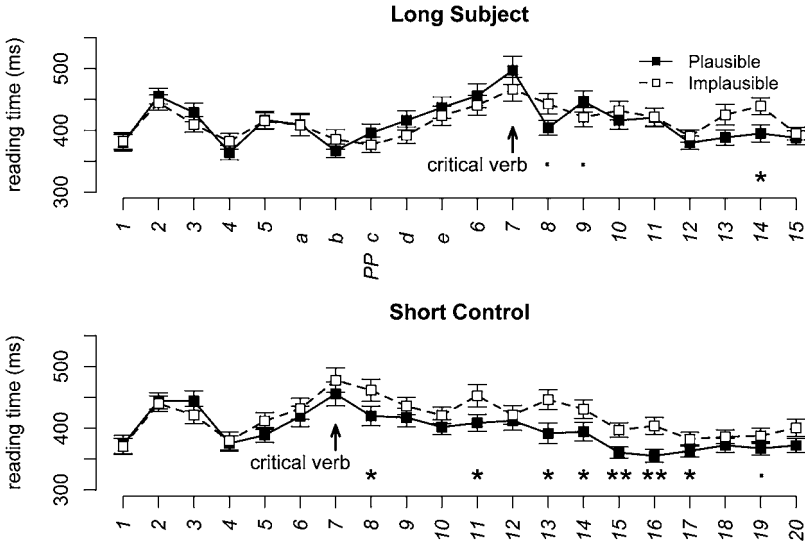


Figure 3
Experiment 3: Region-by-region reading times.
(The coating/program that the engineer)₁₋₅ ...

Long Subject: [pp from_a the_b high-tech_c aerospace_d firm_e]
methodically₆ sprayed₇ the₈ special₉ test₁₀ surfaces₁₁ with₁₂ in₁₃ his₁₄
laboratory₁₅

Short Control: methodically₆ sprayed₇ the₈ special₉ test₁₀ surfaces₁₁ with₁₂ in₁₃ his₁₄
laboratory₁₅ (...)₁₆₋₂₀ ...

Region-by-region reading times. Arrows indicate the critical verb. The symbols that indicate the result of a pairwise by-participants RMANOVA are as follows:
p: ** < 0.01 < * < 0.05 < • < 0.10.

(Region 15), Short dependency conditions again showed a reliable slowdown whereas Long dependencies did not.

5.3 Discussion

The results of Experiment 3 provided a useful check against those obtained in Experiment 2. The slowdown due to implausibility was attenuated when the distance between filler and gap was lengthened. This can be seen in the comparison of Short and Long conditions in Experiment 3, and can also be seen in the comparison of Coordinate and Single VP conditions in Experiment 2. The largest and most sustained responses to an implausible filler were obtained when the filler-gap distance was short, or in a second coordinated VP.

To further test the strength of this result, we conducted a pooled analysis of Experiments 2-3, combining Long conditions together, and Short/Coordinate conditions together. We estimated the difference in plausibility

effects from a linear mixed-effects model. In the verb region, the plausibility effect was 55 milliseconds greater for Short/Coordinate conditions than for Long conditions (95% C.I.: [-4, 120] ms; $p < 0.10$); in the combined four-word argument region, it was also greater by 25 ms (95% C.I.: [2, 49]; $p < 0.05$). No significant differences between plausibility effects as a function of length were detected at the gap-hosting preposition or in the combined three-word region that followed. The common plausibility effect in this latter region was 30 ms (95% C.I.: [1, 58]; $p < 0.05$).

The common property underlying the variation in the strength of the active dependency formation effect in Experiments 2 and 3 may be the time elapsed between two important processing events: when the filler was most recently processed, and when it was again required for processing.

These results do not affect any conclusions about the Adjunct Clause condition in Experiment 1, since the second verb in that experiment was as close to the first verb as in the Coordinate VP condition. Below we discuss how our finding of a length effect might help to clarify why active dependency formation occurs as soon as possible, regardless of what processing goals motivate it being active.

6. GENERAL DISCUSSION

This study has tested whether the parser persists in actively and incrementally constructing gaps in multiple-gap dependencies, even after the first filler-gap relationship has been successfully constructed. The goal of the study was to assess whether grammatical constraints actively direct the formation of *wh*-dependencies. Two kinds of multiple-gap dependencies were compared: across-the-board extraction from coordinate VP structures, and parasitic gaps inside post-verbal adjunct clauses. Crucially, multiple gaps are obligatory in true coordinate structures, while parasitic gaps are always optional. In Experiment 1 these generalizations were confirmed in an off-line rating study. The self-paced reading studies in Experiments 1 and 2 showed a strong effect of the semantic fit between the *wh*-phrase and the verb in the second coordinate of a coordinated VP. The use of the ditransitive *spray/load*-type verbs in Experiment 2 confirmed that this effect emerges before direct evidence for the gap position. These results suggest that comprehenders engage in active dependency completion when they detect a coordinate structure containing a gap in the first coordinate. In contrast, encountering a potential parasitic gap environment after completing a filler-gap dependency does not re-engage active dependency completion mechanisms, despite the fact that an additional gap is fully acceptable. Therefore, we conclude that comprehenders are sensitive in real time to the grammatical implications of building multiple-gap constructions. The grammatical generalizations about across-the-board extraction and parasitic gaps do actively guide parsing decisions.

6.1 *Positive evidence for restrictions on wh-dependency formation*

In Experiments 1 and 2, the requirement that there be a gap in each member of a gapped coordinate VP was reflected in a plausibility effect at or beyond the second verb. This effect shows a positive behavioral effect of constraints on unbounded dependencies. This contrasts with the null effects offered as real-time consequences of island constraints in previous studies (e.g. Stowe 1986, Bourdages 1992, Phillips 2006, Pickering et al. 1994, Traxler & Pickering 1996; cf. footnote 5). We should emphasize that the null effect logic of these studies is not in itself faulty. However, the results can be subjected to alternative explanations. Particularly, there is a general concern that many of the environments tested are independently complex to process and, for that reason, dependency construction is difficult or seldom observed in those domains (Deane 1991, Kluender 2004). Our results resist an account of this type. The contrast between dependency formation in coordinate and parasitic gap structures cannot be attributed to a complexity difference between the two, because the results of the acceptability study accompanying Experiment 1 show that constructions with either across-the-board extraction from coordinates or parasitic gaps were rated equally highly. Participants were thus never challenged to entertain ungrammatical or difficult analyses. It thus seems hard to explain away a lack of active dependency formation in parasitic gap environments on complexity grounds.

However, we note that the parser's close adherence to grammatical principles and constraints in the present study, and in most studies on *wh*-dependencies, does not automatically generalize to other kinds of grammatical dependencies. A number of instances have been documented in which the comprehender appears to entertain relationships excluded by the grammar: for example, in the licensing of subject-verb agreement (Pearlmutter, Garnsey & Bock 1999), case dependencies (Meng & Bader 2000), and negative polarity items (Vasishth, Brüssow, Lewis & Drenhaus 2008). This grammatical fallibility does appear selective, however, affecting only certain category configurations and feature values. A constellation of closely-related and overlapping factors seem to play a role in accurately licensing a category, such as its predictability, morpho-syntactic markedness, relative distinctiveness among nearby categories, and the size and structure of its grammatical domain (see Wagers 2008 for discussion). It is an important topic for future investigation exactly what conditions matter for how grammatically faithful the comprehender is in real time.

6.2 *Multiple dependencies and the grammatical motivation for active dependency formation*

The goal of the comparison between coordinate VPs and parasitic gap environments was to test several alternative accounts of the mechanisms that underlie the active construction of unbounded dependencies. Several lines of

experimental research have shown that filler–gap dependencies are resolved as soon as possible, before the input directly signals the position of the gap site. Previous studies of active dependency formation have been consistent with several distinct accounts of what motivates the active construction of dependencies: a drive to license structure and satisfy grammatical constraints, a drive to provide an interpretive role for the displaced filler, or to assign an interpretation to the partial string, all as soon as possible. The comparison of multiple dependency formation in coordinate VPs and parasitic gap environments allows us to distinguish these possibilities. The existence of persistent active dependency formation effects in coordinated phrases (Experiments 1 and 2) precludes a parser in which dependency formation is driven solely by the need to license unsatisfied features of the filler. Once the first filler has been integrated in the first coordinate, its thematic and other syntactic properties are, by hypothesis, already satisfied. Something else must be prompting the active integration of the filler at the second potential gap site.

Pritchett's Thematic Attachment hypothesis (Pritchett 1992), or its non-grammaticalized analogue (Altmann & Kamide 1999), which we have termed Maximal Interpretation, provides a candidate explanation of past results: the parser attempts to satisfy the Theta Criterion at all moments and thus should seek to combine a filler with an unsaturated verb at any opportunity. However, this view cannot explain the pattern of filler–gap resolution observed in multiple-dependency constructions. Both Thematic Attachment and Maximal Interpretation predict that we should find active dependency formation at the second coordinate, but they also predict active dependency formation in parasitic gap environments. Yet we found no evidence that comprehenders were sensitive to the semantic fit between the filler and the adjunct clause verbs. If comprehenders attempted to interpret as much as possible, as soon as possible, then we should have expected verbs within legal p-gap environments to exhibit a sensitivity to semantic fit, contrary to fact.

The most salient difference between the two environments is that multiple gaps are obligatory in gapped coordinates, but are not obligatory in p-gap configurations. The present findings indicate that the mechanisms that build structure in real time must also be capable of evaluating and conforming to the constraints on across-the-board extraction. The parser must be able to recognize a CSC environment in real time, taking note that a gap-containing constituent has been constructed and that that same constituent is to be coordinated. Once the well-formedness requirement has been recognized, that requirement exerts pressure on the processing system to satisfy it.

Several potential mechanisms could subserve the real-time recognition of the coordinate structure constraints on extraction. For syntactic theories with complex categories whose features or type encodes information about

missing constituents (e.g. Generalized Phrase Structure Grammar, Gazdar, Klein, Pullum & Sag 1985; or Categorical Grammar, Ades & Steedman 1982), the real-time application of the coordination rule would extend the phrase marker with a gap-containing VP. Such syntactic representations are not required, however, to capture our results. It has also been proposed that the parser provisionally uses the first coordinate as a template for further parsing (Frazier & Clifton 2001, Steiner 2003). This proposal is intended to capture extensive findings that second coordinates are processed more rapidly the greater their structural parallelism to the initial constituent (Frazier, Munn & Clifton 2000, *inter alia*).

6.3 *Pre-verbal parasitic gaps*

Our results contrast interestingly with those reported by Phillips (2006), and provide a more nuanced perspective on how uninterpreted material can motivate the parser's decisions. Phillips was interested in the fact that the grammar allows pre-verbal parasitic gaps in certain configurations. While gaps are generally prohibited inside subjects, they are permitted inside subjects with an infinitival complement if there is a licensing gap in the matrix clause, as (17) illustrates.

- (17) (a) What did the attempt to repair the car ultimately damage __?
 (b) *What did the attempt to repair __ ultimately damage the car?
 (c) What did the attempt to repair ___{PG} ultimately damage __?

A gap is permitted in the direct object position alone (17a), but not for an object NP inside the subject projection (17b). Only when a gap is present in the direct object position can a gap also appear inside the subject (17c). Thus, the subject-contained gap is standardly understood to be a p-gap. Phillips found evidence that the parser actively postulates gaps inside the infinitival subject complements, using the same experimental method and plausibility manipulation technique we have employed here. No actual p-gaps appeared in his design (as was the case in the current study), so this effect must have been mediated top-down, by the prediction for a gap.

The present study and Phillips (2006) both consider positions that optionally host a p-gap, but the results differ in that Phillips (2006) found evidence for active dependency formation, whereas the present study did not. What differs from the environments examined in the present study is the status of the filler. In Phillips (2006), when the parser encounters the potential p-gap in a subject, the filler has not yet been thematically interpreted. By contrast, in the present study, when the parser encounters the potential p-gap environment in a post-verbal adjunct clause, the filler has already been thematically interpreted. Therefore, in both cases the parser is still motivated to satisfy grammatical constraints as soon as possible. It is interesting, however, that in Phillips (2006) the parser seems willing to assign a thematic role or

interpretation to the filler inside a subject, but at the cost of undertaking another licensing requirement, namely that between a p-gap and the matrix clause gap.⁸

6.4 *Length effects*

We found two pieces of evidence that the plausibility effect which indexes active dependency completion is attenuated by the serial length of the dependency. In Experiment 2 the response to implausibility was stronger in the coordinated VP condition than in the single VP control, as reflected in earlier and more persistent slowdown effects. Similarly, in Experiment 3 we were able to elicit a stronger response in a single VP by removing a subject-adjoined PP. This pattern of results likely reflects the nature of the memory representations that underlie long-distance dependency formation. We hypothesize that the representation of the filler is, at least in part, subject to processes of interference or decay (Gordon, Hendrick & Johnson 2001, McElree et al. 2003, Lewis & Vasishth 2005). Consequently, the strength of the filler or the likelihood of successful retrieval should decline as more material intervenes before filler-gap resolution is attempted. This view is somewhat at odds with previous claims that the filler occupies a distinguished representational state in real-time processing (Wanner & Maratsos 1978, Frazier & Flores d'Arcais 1989). However, it also provides another perspective on why active dependency completion is active. As the strength or retrievability of the filler representation declines, so might the likelihood of successfully constructing a grammatical representation or deriving a sound interpretation. In this way, the nature of memory provides further insight into why any incentive for the parser to construct filler-gap dependencies involves an 'as soon as possible' pressure.

7. CONCLUSION

Multiple *wh*-dependencies are obligatory in some environments, as in the case of displacement from coordinate structures, but optional in others, as in the case of parasitic gaps. In three on-line studies we showed that comprehenders are sensitive to this distinction in real time and use it to motivate parsing decisions. In the processing of coordinated VPs, comprehenders

[8] Consistent with the conclusions of this paper, this willingness to postulate a gap inside the subject projection is nonetheless constrained by the grammar. Phillips (2006) also tested finite complements of definite subjects, which, unlike those in (1), cannot host p-gaps (but cf. Kayne 1983). There he found no evidence that a gap was postulated inside the subject. An anonymous reviewer suggests an ingenious follow-up, which would be to test for the category sensitivity of optional p-gap parsing. English p-gaps may only be DPs (see Postal 1993). Other displaced categories (AP, PP, AdvP, etc.) should therefore not show signs of active gap creation in the infinitival environments.

persist in actively constructing a dependency between the *wh*-filler and subsequent potential gap sites. However, they refrain from doing so in the post-verbal adjunct clauses that can support parasitic gaps. The construction of filler–gap dependencies in coordinate VPs appears to be active, that is, it precedes a bottom–up signal for the location of a missing constituent. By investigating a constraint on long-distance dependencies that requires the construction of subsequent dependencies, the evidence that a grammatical constraint plays a real-time role was reflected in a positive contrast, instead of the null effect found in previous studies of island-constraint effects on dependency construction. Consequently, our results resist an explanation in terms of complexity. Finally, the present study clarifies the motivation for active dependency construction. The real-time distinction between optional and obligatory dependencies suggests that it is insufficient to appeal to general principles of maximal incremental interpretation to motivate active dependency formation. Instead, this distinction provides evidence that parsing decisions strongly rely on constraints found in the grammar.

APPENDIX

Analysis of variance tables

	VP STRUCTURE	FILLER PLAUSIBILITY	STRUCTURE × PLAUSIBILITY
<i>Region 7</i>	F ₁ : 0.4; MSE: 19058	F ₁ : 0.0; MSE: 18370	F ₁ : 0.0; MSE: 18834
VP₁ adverb	F ₂ : 0.0; MSE: 52169	F ₂ : 0.1; MSE: 25930	F ₂ : 0.0; MSE: 53606
<i>Region 8</i>	F ₁ : 0.7; MSE: 58561	F ₁ : 0.3; MSE: 38954	F ₁ : 0.0; MSE: 42244
VP₁ verb	F ₂ : 0.4; MSE: 64638	F ₂ : 0.5; MSE: 43092	F ₂ : 0.2; MSE: 66466
<i>Region 9</i>	**F ₁ : 11.1; MSE: 26723	F ₁ : 1.4; MSE: 25773	F ₁ : 0.3; MSE: 19862
Coordinator/ Preposition	*F ₂ : 5.2; MSE: 44981	F ₂ : 1.3; MSE: 20997	F ₂ : 0.3; MSE: 22969
<i>Region 10</i>	**F ₁ : 8.7; MSE: 28942	F ₁ : 0.0; MSE: 26914	F ₁ : 2.8; MSE: 29280
VP₂ adverb	*F ₂ : 5.0; MSE: 36689	F ₂ : 0.0; MSE: 21492	F ₂ : 1.3; MSE: 50524
<i>Region 11</i>	F ₁ : 1.6; MSE: 52966	F ₁ : 2.4; MSE: 31544	F ₁ : 0.0; MSE: 37185
VP₂ verb	F ₂ : 2.0; MSE: 77546	*F ₂ : 5.5; MSE: 14062	F ₂ : 0.0; MSE: 42741
<i>Region 12</i>	•F ₁ : 3.0; MSE: 33518	F ₁ : 0.9; MSE: 16257	F ₁ : 0.6; MSE: 13524
VP₂ verb + 1	•F ₂ : 3.0; MSE: 31019	F ₂ : 1.1; MSE: 14415	F ₂ : 0.4; MSE: 21094
<i>Region 13</i>	***F ₁ : 31.6; MSE: 37801	F ₁ : 0.20; MSE: 24965	**F ₁ : 7.9; MSE: 18598
VP₂ verb + 2	***F ₂ : 19.9; MSE: 54156	F ₂ : 0.07; MSE: 31296	*F ₂ : 4.5; MSE: 38588
<i>Region 14</i>	F ₁ : 0.3; MSE: 10790	F ₁ : 1.7; MSE: 10777	F ₁ : 0.8; MSE: 20436
VP₂ verb + 3	F ₂ : 0.0; MSE: 33018	F ₂ : 0.9; MSE: 22256	F ₂ : 0.8; MSE: 25454
<i>Region 15</i>	•F ₁ : 3.9; MSE: 20308	F ₁ : 0.1; MSE: 10602	F ₁ : 0.4; MSE: 12182
VP₂ verb + 4	F ₂ : 2.7; MSE: 40005	F ₂ : 0.0; MSE: 12055	F ₂ : 0.3; MSE: 22781

Numerator *df* in each manipulated factor: 1. Subject *n*: 36. Item *n*: 24.
 Significance tests: p: 0 < *** < 0.001 < ** < 0.01 < * < 0.05 < • < 0.10.

Table A1

Experiment 1. Omnibus Repeated Measures ANOVA: Regions 7–15.

MULTIPLE DEPENDENCIES AND REAL-TIME COMPREHENSION

	VP STRUCTURE	FILLER PLAUSIBILITY	STRUCTURE × PLAUSIBILITY
<i>Region 6</i> VP₁ verb or PP prep.	F ₁ : 2.0; MSE: 23615 F ₂ : 0.8; MSE: 58759	F ₁ : ~0; MSE: 20043 F ₂ : 0.1; MSE: 50008	F ₁ : 2.5; MSE: 19593 F ₂ : 1.0; MSE: 60626
<i>Region 7</i> VP₁ verb + 1 PP prep. + 1	F ₁ : 0.03; MSE: 14573 F ₂ : 0.01; MSE: 8288	F ₁ : 1.0; MSE: 6047 F ₂ : 0.2; MSE: 23731	*F ₁ : 4.8; MSE: 6997 F ₂ : 1.8; MSE: 18003
<i>Region 8</i> VP₁ verb + 2 PP prep + 2	**F ₁ : 12.6; MSE: 20801 **F ₂ : 8.5; MSE: 32931	F ₁ : 0.3; MSE: 8192 F ₂ : 0.0; MSE: 25296	F ₁ : 0.3; MSE: 11147 F ₂ : 0.4; MSE: 32869
<i>Region 9</i> VP₁ verb + 3 PP prep + 3	***F ₁ : 21.9; MSE: 22732 **F ₂ : 9.5; MSE: 44493	F ₁ : 0.1; MSE: 19089 F ₂ : 0.0; MSE: 26167	F ₁ : 1.7; MSE: 12237 F ₂ : 0.5; MSE: 18997
<i>Region 10</i> Coordinator or PP prep + 4	***F ₁ : 27.2; MSE: 24100 ***F ₂ : 16.7; MSE: 43081	F ₁ : 0.6; MSE: 15286 F ₂ : 0.2; MSE: 24689	F ₁ : 0.1; MSE: 20195 F ₂ : 0.1; MSE: 29751
<i>Region 11</i> Adverb	***F ₁ : 15.7; MSE: 18461 F ₂ : 2.7; MSE: 103985	F ₁ : 0.5; MSE: 36748 F ₂ : 0.4; MSE: 59255	F ₁ : 2.4; MSE: 28162 F ₂ : 2.3; MSE: 38392
<i>Region 12</i> VP₂ verb	***F ₁ : 17.7; MSE: 18669 *F ₂ : 5.9; MSE: 63170	*F ₁ : 4.9; MSE: 15335 F ₂ : 1.5; MSE: 28529	F ₁ : 0.4; MSE: 23148 F ₂ : 0.3; MSE: 36476
<i>Region 13</i> Ground Arg 1	*F ₁ : 4.8; MSE: 13487 *F ₂ : 6.4; MSE: 11921	*F ₁ : 6.8; MSE: 15099 •F ₂ : 3.0; MSE: 29803	F ₁ : 0.1; MSE: 7205 F ₂ : 0.0; MSE: 17689
<i>Region 14</i> Ground Arg 2	**F ₁ : 8.9; MSE: 8707 •F ₂ : 4.2; MSE: 21185	F ₁ : ~0; MSE: 21050 F ₂ : 0.0; MSE: 31929	F ₁ : 0.3; MSE: 6480 F ₂ : 0.1; MSE: 21227
<i>Region 15</i> Ground Arg 3	F ₁ : 3.8; MSE: 12779 F ₂ : 2.3; MSE: 19477	•F ₁ : 3.5; MSE: 7515 F ₂ : 0.6; MSE: 23215	*F ₁ : 4.6; MSE: 7257 F ₂ : 0.6; MSE: 29131
<i>Region 16</i> Ground Arg 4	F ₁ : 0.9; MSE: 16370 F ₂ : 1.3; MSE: 16370	**F ₁ : 9.1; MSE: 10959 F ₂ : 2.6; MSE: 28742	F ₁ : 0.5; MSE: 8342 F ₂ : 1.3; MSE: 14212
<i>Region 17</i> Figure prep.	F ₁ : 0.9; MSE: 7891 F ₂ : 1.0; MSE: 7350	**F ₁ : 7.8; MSE: 4747 F ₂ : 1.5; MSE: 28479	F ₁ : 0.3; MSE: 5515 F ₂ : 0.3; MSE: 36237
<i>Region 18</i> AdvP Word 1	*F ₁ : 6.7; MSE: 7935 F ₂ : 3.5; MSE: 16831	*F ₁ : 8.0; MSE: 8884 F ₂ : 2.2; MSE: 27038	F ₁ : 0.0; MSE: 11602 F ₂ : ~0; MSE: 50
<i>Region 19</i> AdvP Word 2	*F ₁ : 4.8; MSE: 15702 •F ₂ : 4.1; MSE: 29531	*F ₁ : 5.1; MSE: 27154 •F ₂ : 3.0; MSE: 32501	F ₁ : 0.1; MSE: 12559 F ₂ : 0.4; MSE: 37746
<i>Region 20</i> AdvP Word 3	F ₁ : 1.9; MSE: 13354 F ₂ : 1.0; MSE: 23418	***F ₁ : 22.3; MSE: 8016 **F ₂ : 10.4; MSE: 18352	F ₁ : 1.1; MSE: 6538 F ₂ : 0.4; MSE: 20754

Numerator *df* in each manipulated factor: 1. Subject *n*: 31. Item *n*: 24.
Significance tests: p: 0 < *** < 0.001 < ** < 0.01 < * < 0.05 < • < 0.10.

Table A2

Experiment 2. Omnibus Repeated Measures ANOVA: Regions 6–20.

	DEPENDENCY LENGTH	FILLER PLAUSIBILITY	LENGTH × PLAUSIBILITY
<i>Region 6</i>	F ₁ : 2.4; MSE: 35453	F ₁ : 0.5; MSE: 20365	F ₁ : 0.7; MSE: 26578
Adverb	F ₂ : 2.6; MSE: 25998	F ₂ : 0.0; MSE: 39828	F ₂ : 1.1; MSE: 32647
<i>Region 7</i>	F ₁ : 0.8; MSE: 37684	F ₁ : 0.2; MSE: 33576	F ₁ : 2.2; MSE: 43787
Verb	F ₂ : 0.6; MSE: 41450	F ₂ : 0.0; MSE: 57704	F ₂ : 1.8; MSE: 37796
<i>Region 8</i>	F ₁ : 1.2; MSE: 32378	**F ₁ : 13.3; MSE: 15166	F ₁ : 0.2; MSE: 26131
Ground Arg 1	F ₂ : 1.7; MSE: 26677	**F ₂ : 9.1; MSE: 30646	F ₂ : 0.1; MSE: 28781
<i>Region 9</i>	F ₁ : 0.3; MSE: 28541	F ₁ : 0.3; MSE: 17132	•F ₁ : 3.5; MSE: 19030
Ground Arg 2	F ₂ : 0.3; MSE: 24414	F ₂ : 0.1; MSE: 24949	F ₂ : 1.9; MSE: 24593
<i>Region 10</i>	F ₁ : 0.7; MSE: 28607	F ₁ : 1.6; MSE: 17003	F ₁ : 0.3; MSE: 16607
Ground Arg 3	F ₂ : 1.1; MSE: 23550	F ₂ : 1.3; MSE: 32913	F ₂ : 0.0; MSE: 18780
<i>Region 11</i>	F ₁ : 1.4; MSE: 9124	•F ₁ : 3.4; MSE: 14640	F ₁ : 1.8; MSE: 37582
Ground Arg 4	F ₂ : 0.5; MSE: 21895	F ₂ : 2.3; MSE: 31237	F ₂ : 1.5; MSE: 22613
<i>Region 12</i>	*F ₁ : 6.3; MSE: 17362	F ₁ : 0.6; MSE: 14182	F ₁ : 0.0; MSE: 20995
Figure Prep	**F ₂ : 9.4; MSE: 13939	F ₂ : 1.1; MSE: 10475	F ₂ : 0.1; MSE: 15739
<i>Region 13</i>	F ₁ : 0.7; MSE: 20411	*F ₁ : 7.4; MSE: 30608	F ₁ : 0.5; MSE: 33919
AdvP Word 1	F ₂ : 0.7; MSE: 30299	*F ₂ : 7.2; MSE: 28486	F ₂ : 0.1; MSE: 37792
<i>Region 14</i>	F ₁ : 0.3; MSE: 13985	**F ₁ : 13.1; MSE: 15064	F ₁ : 0.0; MSE: 26306
AdvP Word 2	F ₂ : 0.1; MSE: 24550	**F ₂ : 8.9; MSE: 27296	F ₂ : 0.0; MSE: 30208
<i>Region 15</i>	F ₁ : 1.5; MSE: 11902	*F ₁ : 5.4; MSE: 9714	•F ₁ : 4.1; MSE: 8551
AdvP Word 3	F ₂ : 1.9; MSE: 8331	**F ₂ : 9.4; MSE: 8361	F ₂ : 1.1; MSE: 20133

Numerator *df* in each manipulated factor: 1. Subject *n*: 24. Item *n*: 24.
Significance tests: p: 0 < *** < 0.001 < ** < 0.01 < * < 0.05 < • < 0.10.

Table A3

Experiment 3. Omnibus Repeated Measures ANOVA: Regions 6–15.

REFERENCES

- Ades, Anthony & Mark Steedman. 1982. On the order of words. *Linguistics and Philosophy* 4, 517–558.
- Altmann, Gerry T. M. & Yuki Kamide. 1999. Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition* 73, 247–264.
- Anderson, Stephen R. 1971. On the role of deep structure in semantic interpretation. *Foundations of Language* 7, 387–396.
- Aoshima, Sachiko, Colin Phillips & Amy S. Weinberg. 2004. Processing filler–gap dependencies in a head-final language. *Journal of Memory and Language* 51, 23–54.
- Boland, Julie E., Michael K. Tanenhaus, Susan M. Garnsey & Greg N. Carlson. 1995. Verb argument structure in parsing and interpretation: Evidence from *wh*-questions. *Journal of Memory and Language* 34, 774–806.
- Bourdages, Johanne S. 1992. Parsing complex NPs in French. In Helen Goodluck & Michael S. Rochemond (eds.), *Island constraints: Theory, acquisition and processing*, 61–87. Dordrecht: Kluwer.
- Crain, Stephen & Janet D. Fodor. 1985. How can grammars help parsers? In David Dowty, Lauri Karttunen & Arnold Zwicky (eds.), *Natural language parsing*, 94–128. Cambridge: Cambridge University Press.

- Deane, Paul 1991. Limits to attention: A cognitive theory of island constraints. *Cognitive Linguistics* 2, 1–63.
- Engdahl, Elisabet. 1983. Parasitic gaps. *Linguistics and Philosophy* 6, 5–34.
- Fodor, Janet D. 1978. Parsing strategies and constraints on transformations. *Linguistic Inquiry* 9, 427–473.
- Fraser, Bruce. 1971. A note on the spray-paint cases. *Linguistic Inquiry* 2, 603–607.
- Frazier, Lyn. 1987. Syntactic processing: Evidence from Dutch. *Natural Language & Linguistic Theory* 5, 519–560.
- Frazier, Lyn & Charles Clifton. 2001. Parsing coordinates and ellipsis: Copy *a*. *Syntax* 4, 1–22.
- Frazier, Lyn & Giovanni B. Flores D'Arcais. 1989. Filler-driven parsing: A study of gap filling in Dutch. *Journal of Memory and Language* 28, 331–44.
- Frazier, Lyn, Alan Munn & Charles Clifton. 2000. Processing coordinate structures. *Journal of Psycholinguistic Research* 29, 343–370.
- Freedman, Sandra E. & Kenneth I. Forster. 1985. The psychological status of overgenerated sentences. *Cognition* 19, 101–131.
- Garnsey, Susan M., Michael K. Tanenhaus & Robert M. Chapman. 1989. Evoked potentials and the study of sentence comprehension. *Journal of Psycholinguistic Research* 18, 51–60.
- Gazdar, Gerald, Ewan Klein, Geoffrey K. Pullum & Ivan A. Sag. 1985. *Generalized Phrase Structure Grammar*. Oxford: Blackwell.
- Gibson, Edward. 1998. Syntactic complexity: Locality of syntactic dependencies. *Cognition* 68, 1–76.
- Gibson, Edward & Gregory Hickok. 1993. Sentence processing with empty categories. *Language and Cognitive Processes* 8, 147–161.
- Goldsmith, John. 1985. A principled exception to the coordinate structure constraint. *Twenty-First Annual Regional Meeting of the Chicago Linguistic Society (CLS 21)*, 133–143. Chicago: Chicago Linguistic Society.
- Gordon, Peter C., Randall Hendrick & Marcus Johnson. 2001. Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory and Cognition* 27, 1411–1423.
- Gorrell, Paul. 1993. Evaluating the direct association hypothesis: A reply to Pickering and Barry 1991. *Language and Cognitive Processes* 8, 129–146.
- Kaan, Edith. 1997. *Processing subject–object ambiguities in Dutch*. Ph.D. dissertation, University of Groningen.
- 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.
- Kayne, Richard S. 1983. Connectedness. *Linguistic Inquiry* 14, 223–249.
- Kluender, Robert. 2004. Are subject islands subject to a processing account? In Vineeta Chand, Ann Kelleher, Angelo J. Rodriguez & Benjamin Schmeiser (eds.), *23rd West Coast Conference on Formal Linguistics (WCCFL 23)*, 514–527. Somerville, MA: Cascadilla Press.
- Kluender, Robert & Marta Kutas. 1993. Subjacency as a processing phenomenon. *Language and Cognitive Processes* 8, 573–633.
- Kurtzman, Howard S. & Loren E. Crawford. 1991. Processing parasitic gaps. In Tim Sherer (ed.), *21st Annual Meeting of the North Eastern Linguistics Society (NELS 21)*, 217–231. Amherst, MA: LSA Publications.
- Lakoff, George. 1986. Frame semantic control of the coordinate structure constraint. *Twenty-Second Annual Regional Meeting of the Chicago Linguistic Society (CLS 22)*, 152–167. Chicago: Chicago Linguistic Society.
- Lau, Ellen F., Henny H. Yeung, Ryu Hashimoto, Allen Braun & Colin Phillips. 2006. Time-course and localization of syntactic and semantic anomaly responses in sentence processing: A within-subjects fMRI/MEG design. Presented at the Society for Neuroscience Annual Meeting, October.
- Lee, Ming-Wei. 2004. Another look at the role of empty categories in sentence processing (and grammar). *Journal of Psycholinguistic Research* 33, 51–73.
- Levin, Beth. 1993. *English verb classes and alternations: A preliminary investigation*. Chicago: University of Chicago Press.
- Lewis, Richard L. & Shrawan Vasishth. 2005. An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science* 29, 375–419.

- McElree, Brian & Thomas Bever. 1989. The psychological reality of linguistically defined gaps. *Journal of Psycholinguistic Research* 18, 21–35.
- McElree, Brian, Stephanie Foraker & Lisbeth Dyer. 2003. Memory structures that subserve sentence comprehension. *Journal of Memory and Language* 48, 67–91.
- McKinnon, Richard & Lee Osterhout. 1996. Constraints on movement phenomena in sentence processing: Evidence from event-related potentials. *Language and Cognition* 11, 495–523.
- Meng, Michael & Markus Bader. 2000. Ungrammaticality detection and garden path strength: Evidence for serial parsing. *Language and Cognitive Processes* 16, 615–666.
- Neville, Helen J., Janet L. Nicol, Andrew Barss, Kenneth Forster & Merrill Garrett. 1991. Syntactically-based sentence processing classes: Evidence from event-related potentials. *Journal of Cognitive Neuroscience* 3, 151–165.
- Nicol, Janet L., Janet D. Fodor & David Swinney. 1994. Using cross-modal lexical decision tasks to investigate sentence processing. *Journal of Experimental Psychology: Learning, Memory & Cognition* 20, 1229–1238.
- Nicol, Janet L. & David Swinney. 1989. The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research* 18, 5–19.
- Pearlmutter, Neal J., Susan M. Garnsey & Kathryn Bock. 1999. Agreement processes in sentence comprehension. *Journal of Memory and Language* 41, 427–456.
- Phillips, Colin. 2006. The real-time status of island phenomena. *Language* 82, 795–823.
- Phillips, Colin, Nina Kazanina & Shani H. Abada. 2005. ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research* 22, 407–428.
- Phillips, Colin & Matthew W. Wagers. 2007. Relating structure and time in linguistics and psycholinguistics. In Gareth Gaskell (ed.), *Oxford handbook of psycholinguistics*, 739–756. Oxford: Oxford University Press.
- Pickering, Martin J. & Guy D. Barry. 1991. Sentence processing without empty categories. *Language and Cognitive Processes* 6, 229–259.
- Pickering, Martin J., Stephen Barton & Richard Shillcock. 1994. Unbounded dependencies, island constraints and processing complexity. In Charles Clifton, Lyn Frazier & Keith Rayner (eds.), *Perspectives on sentence processing*, 199–224. London: Lawrence Erlbaum.
- Pickering, Martin J. & Matthew J. Traxler. 1998. Plausibility and recovery from garden paths: An eye-tracking study. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24, 940–961.
- Pickering, Martin J. & Matthew J. Traxler. 2003. Evidence against the use of subcategorisation frequency in the processing of unbounded dependencies. *Language and Cognitive Processes* 18, 469–503.
- Pinker, Steven. 1989. *Learnability and cognition*. Cambridge, MA: MIT Press.
- Postal, Paul M. 1993. Parasitic gaps and the across-the-board phenomenon. *Linguistic Inquiry* 24, 735–754.
- Postal, Paul. 1998. *Three investigations of extraction*. Cambridge, MA: MIT Press.
- Pritchett, Bradley L. 1992. *Grammatical competence and parsing performance*. Chicago: University of Chicago Press.
- Radó, Janina. 1999. Discourse effects in gap-filling. Presented at AMLaP-99, University of Edinburgh.
- Raaijmakers, Jeroen G. W., Joseph M. C. Schrijnemakers & Frans Gremmen. 1999. How to deal with ‘The language-as-fixed-effect fallacy’: Common misconceptions and alternative solutions. *Journal of Memory and Language* 41, 416–426.
- Rappaport, Malka & Beth Levin. 1986. What to do with theta roles. *Lexicon Project Working Papers* 11. Cambridge, MA: Center for Cognitive Science, MIT.
- Rohde, Douglas. 2003. Linger, version 2.94. <http://tedlab.mit.edu/~dr/linger/>.
- Ross, John R. 1967. *Constraints on variables in syntax*. Ph.D. dissertation, MIT.
- Schlesewsky, Matthias, Gisbert Fanselow, Reinhold Kliegl & Josef Krems. 2000. The subject preference in the processing of locally ambiguous *wh*-questions in German. In Barbara Hemforth & Lars Konieczny (eds.), *German sentence processing*, 65–93. Dordrecht: Kluwer.
- Sedivy, Julie C., Michael K. Tanenhaus, Craig G. Chambers & Greg N. Carlson. 1999. Achieving incremental semantic interpretation through contextual representation. *Cognition* 71, 109–147.
- Sekerina, Irina A. 2003. Scrambling and processing: Dependencies, complexity, and constraints. In Simin Karimi (ed.), *Word order and scrambling*, 301–324. Malden, MA: Blackwell.

- Steiner, Ilona. 2003. Parsing syntactic redundancies in coordinate structures. Presented at the European Cognitive Science Conference (EuroCogScio3), Osnabrück, 10–13 September.
- Stowe, Laurie. 1986. Parsing *wh*-constructions: Evidence for on-line gap location. *Language and Cognitive Processes* 1, 227–246.
- Sturt, Patrick. 2003. The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language* 48, 542–562.
- Sussman, Rachel & Julie C. Sedivy. 2003. The time course of processing syntactic dependencies. *Language and Cognitive Processes* 18, 143–163.
- Szabolcsi, Anna & Marcel den Dikken. (2002). Islands. In Lisa Cheng & Rint Sybesma (eds.), *The second state-of-the-article book*. Berlin: Mouton de Gruyter.
- Tanenhaus, Michael K., Laurie A. Stowe & Greg N. Carlson. 1985. The interaction of lexical expectation and pragmatics in parsing filler–gap constructions. *Seventh Annual Cognitive Science Society Conference*, 361–365. Irvine, CA.
- Traxler, Matthew J. & Martin J. Pickering. 1996. Plausibility and the processing of unbounded dependencies. *Journal of Memory and Language* 35, 454–475.
- Vasishth, Shravan, Sven Brüssow, Richard L. Lewis & Heiner Drenhaus. 2008. Processing polarity: How the ungrammatical intrudes on the grammatical. *Cognitive Science* 32, 685–712.
- de Vincenzi, Marica. 1991. *Syntactic parsing strategies in Italian*. Dordrecht: Kluwer.
- Wagers, Matthew W. 2008. *The structure for memory meets memory for structure in linguistic cognition*. Ph.D. dissertation, University of Maryland, College Park.
- Wanner, Eric & Michael Maratsos. 1978. An ATN approach to comprehension. In Morris Halle, Joan Bresnan & George A. Miller (eds.), *Linguistic theory and psychological reality*, 119–161. Cambridge, MA: MIT Press.
- Weinberg, Amy S. 1992. Parameters in the theory of sentence processing: Minimal commitment theory goes east. *Journal of Psycholinguistic Research* 3, 339–364.

Authors' addresses : (Wagers)

*Department of Linguistics, University of California, Santa Cruz,
1156 High Street, Santa Cruz, CA 95064, USA
mwagers@ucsc.edu*

(Phillips)

*Department of Linguistics, University of Maryland,
1401 Marie Mount Hall, College Park, MD 20742, USA
colin@umd.edu*