

# Short-term forgetting in sentence comprehension: Crosslinguistic evidence from verb-final structures

Shravan Vasishth

University of Potsdam, Potsdam

Katja Suckow

Humboldt University, Berlin

Richard L. Lewis

University of Michigan, Ann Arbor

Sabine Kern

University of Potsdam, Potsdam

Draft of October 13, 2007

## Abstract

Five experiments present the first piece of evidence from online processing (self-paced reading and eyetracking) that omitting the middle verb in a double center embedding yields an illusion of grammaticality in English, but not in German. One commonly accepted explanation for the English illusion – based on data from offline acceptability ratings – is the forgetting hypothesis, which is the claim that working-memory overload leads to the comprehender to forget the prediction of the upcoming verb (Gibson & Thomas 1999). Forgetting results in the reader not noticing the ungrammaticality. We show that the forgetting hypothesis does an excellent job of explaining the English data, but cannot account for the German results. We argue that the German patterns can be explained by the head-final nature of German: since verb-final structures are more frequent in German subordinate clauses than in English, predictions of upcoming verbs are not forgotten. We argue that the present evidence against forgetting effects in online sentence processing leads to the broader implication that processing effects currently attributed to language-independent working memory constraints are conditioned by grammatical properties of the language under study.

The act of comprehending a sentence necessarily involves a working memory system. Lin-

---

We are grateful to Heiner Drenhaus, Reinhold Kliegl, Umesh Patil, Titus von der Malsburg, and Julie Van Dyke for comments and useful discussions.

guistic items must be incrementally accessed from a long-term declarative memory, encoded as a memory trace, and maintained in memory until they are integrated with other linguistic items. The end-product of these processes is a sufficiently complete representation of the syntax and semantics of a sentence.

A perfect memory system would maintain every linguistic unit of processing in its pristine form, with no deterioration in its memorial representation as the sentence unfolds. In reality, however, memory representations degrade over time, resulting in comprehension difficulty and outright comprehension failures. One plausible explanation for such degradation is decay. Chomsky (1965, 13-14) was perhaps the first to propose that the reduced acceptability of sentences containing a “nesting of a long and complex element” arises from “decay of memory.” In related work, Just and Carpenter (1980), (1992) cast decay in terms of dependency resolution difficulties during the moment of integration. They developed a model of integration that involved activation decay (a side-effect of capacity limitations) as a key determinant of processing difficulty. For example, under the rubric of distance effects, they describe the constraints on dependency resolution as follows: “The greater the distance between the two constituents to be related, the larger the probability of error and the longer the duration of the integration process” (1992, 133).

This distance-based measure of difficulty has considerable empirical support from head-final structures in English (Van Dyke & Lewis, 2003), (Grodner & Gibson, 2005), (Warren & Gibson, 2005), and several theories of sentence comprehension include decay as an integral component in one form or another (Gibson, 1998), (Gibson, 2000), (Lewis & Vasishth, 2005). As encapsulated in the Just and Carpenter quote above, the effect of decay is assumed to express itself at the moment of integration.

One unexpected consequence of decay can be illustrated by the contrast in (1), discussed first by Frazier (1985) (the original observation is attributed to Janet Fodor). Example (1a) is a grammatical sentence: the rules of English allow such embedded structures. Although such complex structures occur in natural text corpora with remarkable regularity (Roeck et al., 1982), they are perceived by native English speakers to be less acceptable than their ungrammatical counterparts (1b), which has the middle verb phrase *was cleaning every week* missing.

- (1) a. The apartment that the maid who the service had sent over was cleaning every week was well decorated.
- b. \*The apartment that the maid who the service had sent over was well decorated.

The first published study involving this contrast is an offline questionnaire-based experiment by Gibson and Thomas (1999). They found (inter alia) that ungrammatical sentences such as (1b) were rated no worse than grammatical ones such as (1a).

In related work, Christiansen and MacDonald, whose work was reported in (Christiansen & Chater, 2001) and in (Gibson & Thomas, 1999), found that ungrammatical sentences were rated significantly better than the grammatical ones. An important point is that Christiansen and MacDonald controlled for sentence length in the two sentence types. By doing so, they also demonstrated that the explanation for the illusion cannot lie in the length difference between grammatical and ungrammatical sentences. Christiansen and MacDonald used an online self-paced reading task where participants were incrementally asked to judge whether the sentence was grammatical so far Gibson and Thomas (1999,

239). The grammaticality judgements were thus an offline dependent measure gathered during an online, incremental sentence processing task.

Returning to Gibson and Thomas' original paper, their explanation for Fodor's grammaticality illusion relies on the assumption that the prediction for the middle verb is forgotten when memory cost exceeds an arbitrary threshold. Recasting the original Gibson and Thomas explanation in terms of the more recent Dependency Locality Theory (Gibson, 2000), memory cost can be quantified as the sum of two components: (a) the number of intervening discourse referents between the verb and the dependent to be integrated with it, and (b) the number of predicted (as-yet unseen) heads.<sup>1</sup> As shown in Figure 1, the point of greatest difficulty is at the second (embedded) verb *admitted*: the subject *nurse* is separated by three new discourse referents, *clinic*, *hired*, and *admitted*,<sup>2</sup> the object by four discourse referents, and one head (a matrix verb) is predicted when the verb *admitted* is processed: a total cost of 8. If we assume – following Gibson and Thomas (1999) – that memory is overloaded once it crosses some arbitrary threshold and that the overload results in forgetting, the prediction for the second verb should be forgotten. The consequence is an illusion of grammaticality and its behavioral correlate is increased acceptability of the sentence. We will refer to this explanation as the *forgetting hypothesis*.

The integration-cost locality metric of DLT is apparently intended to be an abstraction over the notion of activation decay in working memory (Gibson, 2000, 103). However, instead of a continuously varying activation value, the DLT posits a simplification based on discretely counting the number of intervening discourse referents. Thus, forgetting a predicted node is equivalent to the activation falling below some arbitrary threshold (this is stated more explicitly in Gibson, 1998, 9). Since the forgetting event is conditional on memory overload, the total activation available must be restricted in the model; assumes this explicitly in (2000, 103): he speaks of “the limited quantity of activation in the system.” The DLT can therefore be seen as an abstraction over capacity-constrained comprehension theories such as the CAPS-based model (Just & Carpenter, 1992).

There are, however, at least two shortcomings in the empirical evidence available for the grammaticality illusion. First, the forgetting hypothesis refers to online processing difficulty; but the published evidence for forgetting comes from acceptability judgements, an offline task. The only exception is the study reported in (Christiansen & Chater, 2001), but here as well, the dependent measure was acceptability rating, elicited during an online processing task. Given that DLT and its precursor, the Syntactic Prediction Locality Theory (Gibson, 1998), are not primarily theories of acceptability judgements but of online, incremental processing difficulty, one cannot conclude much from the offline data.

Second, there is no published research that investigates whether the forgetting hypothesis applies to languages other than English; if forgetting is not a language-specific matter but rather the consequence of a general working memory system, it is critical to establish whether the same pattern is seen across languages. If, on the other hand, language-specific factors can attenuate forgetting, this is a strong argument against an unconditionally language-independent explanation such as the forgetting hypothesis.

---

<sup>1</sup>Gibson's metric relies on the notions Energy Units and Memory Units; however, the DLT's predictions can be derived by simply counting the number of intervening discourse referents (integration cost) and the number of predicted heads at each point (storage cost).

<sup>2</sup>Finite verbs are assumed to introduce discourse referents (Gibson, 2000).

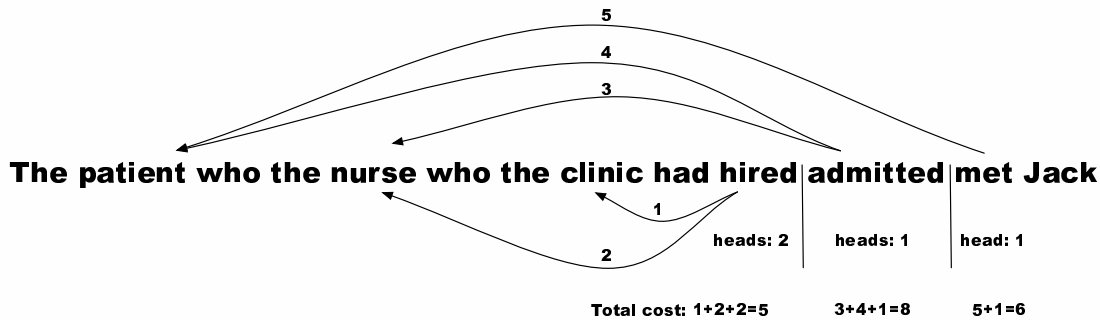


Figure 1. A schematic illustration of DLT’s predictions for multiply embedded structures. Integration costs are labeled along the arcs that define the argument-head dependencies, and storage costs are presented under each verb. The storage costs are the number of heads predicted at each point; only the storage costs at the verb are shown because these are the only critical ones for this paper.

Thus, our first goal was to determine whether the forgetting effect occurs in English in an online comprehension task. We chose two well-known methods, self-paced reading and eyetracking; the motivation for selecting two methods rather than one was to establish robustness of the results independent of the dependent measure. Our second goal was to determine whether head-final structures in head-final languages such as German are subject to the same constraints as head-final structures in English, an SVO language. We chose German (rather than, say, Hindi) as a representative SOV language because it has a critical property that distinguishes it from English – the obligatory presence of commas following embedded verbs. It turns out that this property allows us to evaluate the opposing hypothesis to the forgetting claim.

Regarding the first goal, if forgetting effects are found in online processing as well, this would be a more direct confirmation of the forgetting hypothesis. Regarding the second goal, if the forgetting effect occurs irrespective of the headedness of the language, this is evidence for the cross-linguistic generality of forgetting-based accounts. If, on the other hand, headedness constrains forgetting differently, the explanatory power of locality-based forgetting explanations would be considerably weaker. We consider the broader implications of this issue in the General Discussion.

#### *Predictions of the forgetting hypothesis*

Consider the English grammatical center embedding (2a) and its ungrammatical counterpart (2b), which has the middle verbal phrase *hurt* missing. The corresponding German examples are shown in (3). As mentioned above, an important difference between the two languages is that the embedded verbs require commas.

- (2) a. The carpenter who the craftsman that the peasant carried hurt supervised the apprentice.  
 b. The carpenter who the craftsman that the peasant carried supervised the apprentice.

- (3) a. Der Anwohner, den der Wanderer, den der Pförtner suchte, störte,  
 The resident that the hiker that the doorman searched-for disturbed  
 verarztete den Verletzten.  
 tended-to the injured-person  
 ‘The resident that the hiker that the doorman was looking-for disturbed tended  
 to the injured person.’
- b. Der Anwohner, den der Wanderer, den der Pförtner suchte, verarztete  
 The resident that the hiker that the doorman searched-for tended-to  
 den Verletzten.  
 the injured-person  
 ‘The resident that the hiker that the doorman was looking-for tended to the  
 injured person.’

The forgetting account makes several predictions for online processing of both English and German structures. For ease of exposition, we will ignore below the special property of German – obligatory commas with embedded verbs – that allows us to put the forgetting hypothesis to a more stringent test. We return to the implications of the commas when we discuss the German experiments.

In order to unpack these predictions, we first define the processing steps in both the grammatical and ungrammatical conditions. We will refer the verb sequence in (2a) with indices: V3, V2, V1; V3 is the most embedded verb, V2 the middle verb and V1 the matrix verb. This convention simply reflects the nesting of argument-verb dependencies (as opposed to Dutch (Bach, Brown, & Marslen-Wilson, 1986), where they are crossed; also see Joshi, Becker, & Rambow, 2000; Rambow & Joshi, 1994). We will refer to the verbs in the conditions as *gram-V<sub>n</sub>* and *ungram-V<sub>n</sub>* where *n* is the verb index. In addition, we will use the phrase “second verb” to refer to whichever verb sequentially follows the embedded verb V3; in the grammatical condition it is V2, and in the ungrammatical condition V1.

The parsing events of interest begin as soon as V3 is processed. We separate the events into three steps.

1. **Step 1, process V3:** The prediction for the embedded verb is retrieved, the verb integrated with the prediction, and processing continues to the next word. No reading time difference is predicted at the embedded verb for the grammatical and ungrammatical conditions.

2. **Step 2, process second verb:** The next word is *gram-V2* or *ungram-V1*, depending on the sentence. Since the prediction for the middle verb is by assumption forgotten, the only predicted node available for retrieval in both the ungrammatical and grammatical cases is the matrix verb. That is, in the grammatical condition, the *middle* verb V2 triggers a retrieval of the matrix-verb prediction and attaches to it; by contrast, in the ungrammatical condition the matrix verb V1 retrieves and attaches to the predicted node for the matrix verb. Here too, no difference in reading time should occur at *gram-V2* versus *ungram-V1*.

3. **Step 3a, process gram-V1:** In the grammatical condition, the matrix verb V1 is processed but no predicted verb-node is available for attachment – the middle verb’s prediction has been forgotten and the remaining predicted verb nodes have been consumed. Thus, an error should be signaled in the parser. It is possible that the effects of this error appears in the post-matrix verb region, a spillover effect (Mitchell, 1984).

4. **Step 3b, process post-matrix verb region in ungrammatical condition:** In the ungrammatical condition, after V1 is processed (Step 2), the word following the matrix verb should be processed successfully since all the (unforgotten) predicted verb-nodes have been filled. Thus, the reading time at the post-matrix-verb region in the ungrammatical condition should be shorter compared to the post-matrix verb region in the grammatical condition.

Steps 3a and 3b together lead to the prediction of longer reading time at the matrix verb V1 in the grammatical condition compared to the reading time at the matrix verb V1 in the ungrammatical condition (see Step 2). In self-paced reading, the behavioral correlate of this error – longer reading time – should also occur in the post-matrix verb region. Thus the prediction (for self-paced reading) is that reading time at the post-matrix verb region in the grammatical condition should be longer than at the post-matrix verb region in the ungrammatical condition.

To summarize, the forgetting hypothesis predicts (1) longer reading times at gram-V1 compared to ungram-V1, and (2) longer reading times at the post-matrix verb region in the grammatical condition (gram-V1+1) compared to the post-matrix verb region in the ungrammatical condition (ungram-V1+1). The experiments evaluating these predictions are described next.

### Experiment 1: English self-paced reading

#### *Method*

*Participants.* Forty nine English native-speakers from the University of Michigan participated in the experiment. They received course credit as compensation. All had normal or corrected-to-normal vision.

#### *Procedure and Design.*

The experiment had two factors, grammatical versus ungrammatical sentences. The grammatical sentences had three verbs, while the ungrammatical sentences had the second verb missing.<sup>3</sup>

A self-paced reading comprehension experiment (Just, Carpenter, & Woolley, 1982) was carried out in English at the University of Michigan, USA. Sixteen target sentences were presented in a counterbalanced manner, with 56 filler sentences pseudo-randomly interspersed between the target sentences. The counterbalancing meant that each participant saw each sentence only once, in one of the four conditions. The experiment was run on a Macintosh computer using the Linger software developed by Doug Rohde (<http://tedlab.mit.edu/~dr/Linger/>).

Participants read the introduction to the experiment on the computer screen. In order to read each word of a sentence successively on a moving window display, they had to press the space bar. The word seen previously was masked. Each word was shown separately. At the end of each sentence subjects had to answer yes-no comprehension questions in order to ensure that they would try to comprehend the sentences. In order to prevent them from

---

<sup>3</sup>An orthogonal manipulation in the stimulus items was the animacy of the second NP (*craftsman* versus *pillar* in example (2)). This relates to a different research question – interference in sentence comprehension. This factor did not interact with (was orthogonal to) the grammaticality manipulation; therefore, we do not discuss it in this paper.

developing a question-answering strategy without paying attention to the entire sentence, questions related to different argument-verb relations in the sentences. More specifically, in all the experiments, the questions relating to the target sentences targeted the two embedded verbs (V3 and V2) in the grammatical condition, and both verbs (V3 and V1) in the ungrammatical one; the questions were designed to have “yes” answers in 50% cases. All stimulus items and accompanying questions, along with the expected correct answers, are shown in the Appendix.

#### *Statistical analysis.*

A multilevel (also referred to as hierarchical or mixed-effects) linear model (MLM) was fitted to the data, with random intercepts for participants and items, and with grammaticality as the independent variable. MLMs have several advantages over repeated measures ANOVA, one of them being that they allow us to take by-item and by-participant variance into account simultaneously, making separate analyses or min-F (Clark, 1973), (Raaijmakers, Schrijnemakers, & Gremmen, 1999) estimates redundant; see (Baayen, 2007) for further discussion of this issue. We present coefficient estimates and their standard errors, and t-values. An absolute t-value of 2 or greater indicates significance at the  $\alpha$  level 0.05. Note also that the t-score is not accompanied by degrees of freedom or p-values. This is deliberate; degrees of freedom are trivial to compute in standard t-tests, but only approximations are possible in multilevel models (Baayen, 2007). Analyses using standard repeated measures ANOVA were also carried out; in all cases the analyses led to the same conclusions as presented here.

In all experiments presented, the statistical analyses on reading times were carried out on log-transformed values; the analyses were repeated using raw reading times as well (the outcomes were comparable to the analyses based on transformed values). We report the log-transformed analyses because linear models based on untransformed reading times do not meet the assumptions of additivity and linearity (Gelman & Hill, 2007, 59-65).

#### *Results*

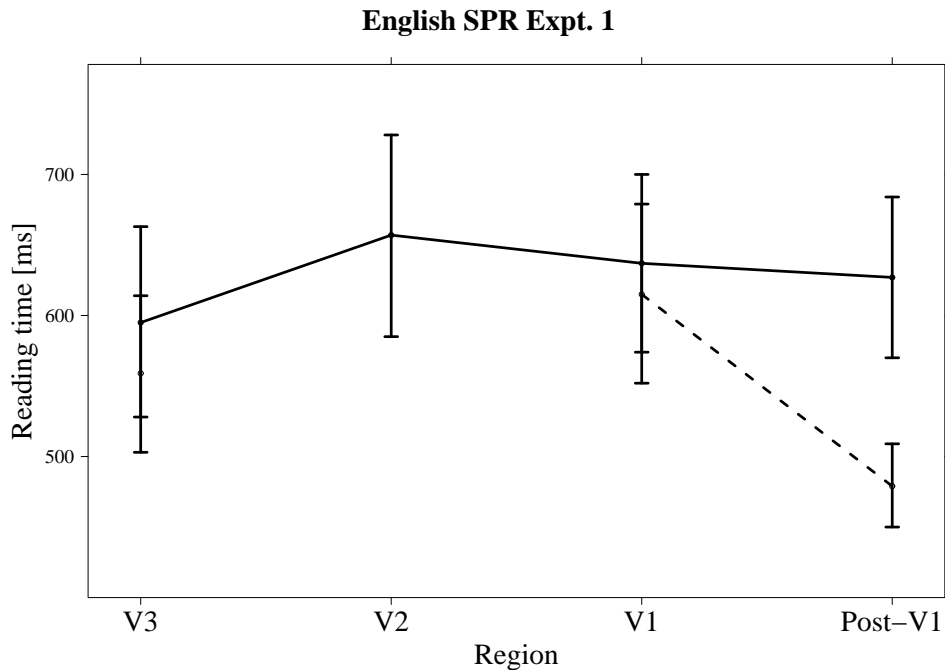
Sentence 11 was not considered in the analysis because it contained a typographic error. The offending sentence was *The clerk who the bureaucrat that the visitor forgotten about. . .*, where it should have been *. . . forgot about. . .*

The mean response accuracy over all items (including distractors) was 80.83%, indicating that the participants were attending to the comprehension task.

The mean reading times and confidence intervals are summarized in Figure 3 and Table 1, and the statistical analyses in Table 2. In the current and subsequent experiments we present reading times for all verb regions and for the post-V1 region; however, the statistical analysis is restricted to the theoretically important regions, V1 and the post-V1 regions; the latter is always the word immediately following V1.

As summarized in Table 2, the only significant effect found was at the post-matrix verb region: as predicted by the forgetting hypothesis, the grammatical condition had longer reading times than the ungrammatical one.

We were concerned that the three verb sequences in succession used in the experiment may have rendered processing unusually difficult, making the results difficult to interpret. We therefore repeated the study with sentences that had more material between the verbs;



*Figure 2.* Mean reading times and 95% confidence intervals for the verb and post-V1 word in the English self-paced reading study (experiment 1).

examples are shown in (4) and the full list of stimuli is available from the authors (these are not included in the paper in order to conserve space). The three verb regions are shown in square brackets. This experiment had 62 fillers interspersed between the target items (the same method was used as described above).

- (4) a. The carpenter who the craftsman that the peasant [ $V_3$  had carried] to the bus-stop [ $V_2$  had hurt] yesterday [ $V_1$  supervised] the apprentice.
- b. The carpenter who the craftsman that the peasant [ $V_3$  had carried] to the bus-stop [ $V_1$  supervised] the apprentice.

The mean response accuracy over all items (including distractors) was 83.5%. As summarized in Figure 3 and Tables 3 and 4, we were able to reproduce the result in experiment 1: the reading time at the post-V1 region was significantly shorter in the ungrammatical condition than the grammatical. In addition, the reading time at V1 also showed the same pattern.

Having confirmed that the predictions of the forgetting hypothesis using self-paced reading, we investigated whether eyetracking during reading could also provide corroborating evidence. This experiment is described next.

Region	Condition	Mean	CI.lower	CI.upper
V3	gram	1041	912	1170
	ungram	983	865	1102
V2	gram	847	753	941
	ungram	-	-	-
V1	gram	886	802	970
	ungram	845	760	930
Post-V1	gram	817	743	890
	ungram	527	480	574

Table 1: Mean reading times and 95% confidence intervals for the verb regions and the post-V1 word in the English self-paced reading study (experiment 1).

Comparison	Coefficient	Standard Error	t-score
gram-V1 vs ungram-V1	-0.03670	0.04264	-0.86
gram-V1+1 vs ungram-V1+1	-0.33963	0.03590	-9.46 *

Table 2: Summary of results of the two comparisons for experiment 1. The asterisk (\*) indicates statistical significance at  $\alpha = 0.05$ .

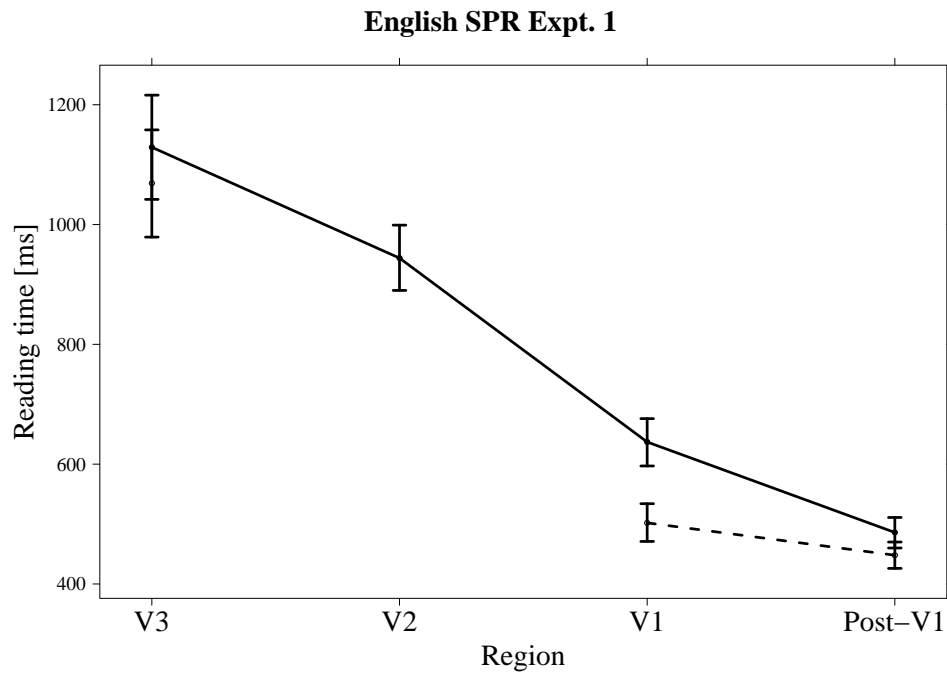


Figure 3. Mean reading times and 95% confidence intervals for the critical regions in the replication of the English self-paced reading study (experiment 1).

Region	Condition	Mean	CI.lower	CI.upper
V3	gram	1129	1042	1216
	ungram	1069	979	1158
V2	gram	944	890	999
	ungram	-	-	-
V1	gram	637	597	676
	ungram	502	471	534
Post-V1	gram	486	460	511
	ungram	448	426	470

Table 3: Mean reading times and 95% confidence intervals for the verb regions and the post-V1 word in the replication of the English self-paced reading study (experiment 1).

Comparison	Coefficient	Standard Error	t-score
gram-V1 vs ungram-V1	-0.20782	0.02848	-7.30 *
gram-V1+1 vs ungram-V1+1	-0.06769	0.02321	-2.92 *

Table 4: Summary of results of the comparisons for the replication of experiment 1. The asterisk (\*) indicates statistical significance at  $\alpha = 0.05$ .

## Experiment 2: English Eyetracking study

### *Method*

#### *Participants.*

Forty seven native English speakers (undergraduates at the University of Michigan) took part in this study, each receiving either course credit of 10 US dollars for completing the task. Participants were tested in individual sessions, and took approximately 30 minutes to complete the experiment. All had normal or corrected-to-normal vision.

#### *Procedure and Design.*

Participants were seated approximately 50 cm from a 19-inch color monitor with  $1024 \times 768$  pixel resolution; twenty-three pixels equaled about one degree of visual angle. Participants wore a SensoMotoric Instruments Eyelink I head-mounted eye-tracker running at 250 Hz sampling rate. Although viewing was binocular, only data from the right eye was used in analyses. Participants were instructed to avoid strong head movements or large shifts in position throughout the experiment. A standard PC keyboard was used to record responses. The presentation of the materials and the recording of the responses was controlled by a PC running software developed (using the Eyelink API) in the University of Michigan eyetracking laboratory

Each participant was randomly assigned one of four stimuli lists which comprised different item-condition combinations according to a Latin Square. The trials per session were randomized individually per participant, subject to the constraints that experimental trials were always separated by at least one filler, and that each session started with at least five fillers. At the start of the experiment, the experimenter performed the standard Eyelink

calibration procedure, which involves participants looking at a grid of nine fixation targets in random succession. Then a validation phase followed to test the accuracy of the calibration against the same targets. If there was a discrepancy between calibration and validation of more than 1.2 degrees of visual angle for any target, or if the average discrepancy across all nine targets was greater than 0.7 degrees of visual angle, then calibration and validation were repeated until the discrepancy was acceptable. Calibration and validation were repeated during the session if the experimenter noticed that measurement accuracy was poor (e.g., after strong head movements or a change in the participant's posture).

Each trial consisted of the following steps. First, a quick single target recalibration ("drift correction") was performed to correct for any degradation of measurement accuracy due to subject movement or slippage of the eyetracker on the head. Then, a fixation target appeared 12 pixels to the left of where the left edge of the text would appear. The stimulus was presented only after the subject fixated on this target for 900 consecutive milliseconds. The participant was instructed to press the spacebar on the keyboard when he/she had finished reading. This triggered the presentation of a simple comprehension question which the participant had to answer either with the 'F' key ('yes') or the 'J' key ('no'). The text stimuli were presented using a Courier New font, printed in white on a black background. The characters (including spaces) were all the same width, approximately 9.1 pixels or 0.39 degrees of visual angle.

The presentation software automatically recorded the coordinates of rectangular interest areas around each word; fixations were then associated with words according to whether their coordinates fell within a word's interest area. The left and right boundaries of each interest area were the midpoints of the spaces between the words (or in the case of the leftmost and rightmost words, the midpoint of where a space would have been had there been another word to the left or right). The upper and lower boundaries were 24 pixels above and below the top and bottom of the line of text. The line of text (i.e., the space taken by a capital letter) was 10 pixels high.

### *Results and Discussion*

The mean response accuracy over all items (including distractors) was 79.2%, indicating that participants were attending to the comprehension task.

In order to map the predictions of the forgetting hypothesis to eyetracking dependent measures, it is necessary to arrive at an understanding of the mapping between eyetracking dependent measures and human parsing processes. The most common dependent measures and their interpretation in terms of reading processes are as follows. *First fixation duration* (FFD) is the first fixation during the first pass, and has been argued to reflect lexical access costs (Inhoff, 1984). *Gaze duration* or *first pass reading time* (FPRT) is the summed duration of all the contiguous fixations in a region before it is exited to a preceding or subsequent word; Inhoff (1984) has suggested that FPRT reflects text integration processes, although Rayner and Pollatsek (1987) argue that FFD and FPRT may reflect similar processes and could depend on the speed of the cognitive process. *Right-bounded reading time* (RBRT) is the summed duration of all the fixations that fall within a region of interest before it is exited to a word downstream; it includes fixations occurring after regressive eye movements from the region, but does not include any regressive fixations on regions outside the region of interest. RBRT may reflect a mix of late and early processes, since subsumes first-fixation

durations. *Re-reading time* (RRT) is the sum of all fixations at a word that occurred after first pass; RRT has been assumed to reflect the costs of late processes (Gordon, Hendrick, Johnson, & Lee, 2006, 1308). Another measure that may be related to late processing is *regression path duration*, which is the sum of all fixations from the first fixation on the region of interest up to, but excluding, the first fixation downstream from the region of interest. Finally, *total reading time* (TRT) is the sum of all fixations on a word.

The early dependent measures (FFD, FPRT) did not show any effect in the eyetracking experiments reported here; the effect showed up in later measures such as re-reading time and total reading time (which subsumes re-reading time and first-pass reading time). We present re-reading times in this paper because recent work has found that effects attributable to retrieval difficulty show up in this dependent measure. For example, (Gordon et al., 2006) and Vasishth, Bruessow, Lewis, and Drenhaus (2008 (to appear)) have found evidence from re-reading time for predicted retrieval difficulty and outright retrieval failures. Since the predictions of the forgetting hypothesis emerge – as discussed earlier – as a consequence of retrieval failure, it is reasonable to focus on these measures.

The mean re-reading times and 95% confidence intervals are summarized in Figure 4 and Tables 5 and 6.

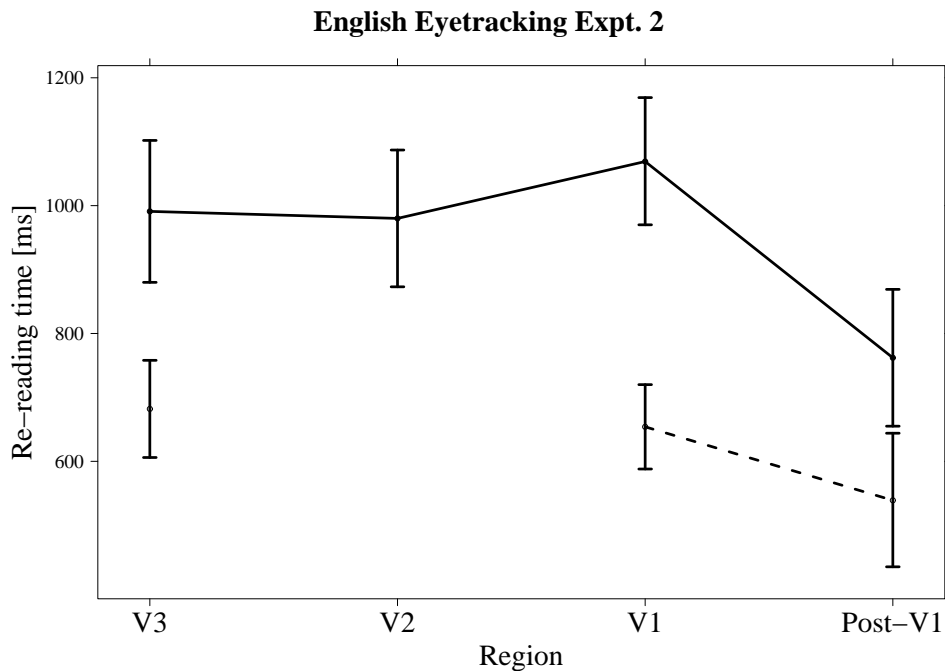


Figure 4. Mean reading times and 95% confidence intervals for the critical regions in the English eyetracking study (experiment 2).

As predicted by the forgetting hypothesis, re-reading time at V1 was significantly shorter in the ungrammatical condition compared to the grammatical one, and the same pattern was seen at the word following V1. In addition, re-reading time at V3 was also shorter in the ungrammatical condition compared to the grammatical one (coefficient -

Region	Condition	Mean	CI.lower	CI.upper
V3	gram	991	880	1102
	ungram	682	606	758
V2	gram	980	873	1087
	ungram	-	-	-
V1	gram	1069	970	1169
	ungram	654	588	720
Post-V1	gram	762	655	869
	ungram	539	435	644

Table 5: Mean reading times and 95% confidence intervals at the verb regions and the post-V1 word in the English eyetracking study (experiment 2).

Comparison	Coefficient	Standard Error	t-score
gram-V1 vs ungram-V1	-0.48972	0.05997	-8.17 *
gram-V1+1 vs ungram-V1+1	-0.31255	0.07874	-3.97 *

Table 6: Summary of results of the comparisons for experiment 2. The asterisk (\*) indicates statistical significance at  $\alpha = 0.05$ .

0.32631, SE 0.06021,  $t=-5.42$ ). This was not predicted by the forgetting hypothesis, but can be explained, as discussed below.

To sum up the English experiments, we found online evidence consistent with the forgetting hypothesis in the predicted regions. The only surprising result was in the eye-tracking data, where we found lower re-reading times at V3 in the ungrammatical condition. However, this effect can easily be explained and is in fact not damaging to the forgetting hypothesis. Re-reading time is a function of regressions to regions that have already been viewed during the first pass. Given that regressions are more frequent in complex sentences (where complexity is defined as increased ambiguity (Clifton Jr, Staub, & Rayner, in press) or any other kind of integration difficulty), and given that the ungrammatical sentences are predicted to be less complex overall, it is not surprising that re-reading time is shorter at V3 in the ungrammatical condition.

Having established that English reading time data is consistent with the forgetting, we turn our attention next to the German experiments.

### Experiment 3: German self-paced reading study

#### *Method*

##### *Participants.*

Thirty nine native German speakers (undergraduates at the University of Potsdam) took part in this study, each receiving 7 Euros for participating. Participants were tested in individual sessions, and took approximately 30 minutes to complete the experiment. All had normal or corrected-to-normal vision.

*Procedure and Design.*

The procedure of the German self-paced reading experiment was identical to the English experiment 1. This experiment had 16 critical items and 60 distractors.

As mentioned earlier, German and English differ in one important respect for the purposes of this paper: the presence of a comma is a clear cue to the reader that a verb occurs in an embedded context; the matrix verb does not require a comma.

It is critical that commas are obligatory for embedded verbs but do not occur with matrix verbs. If – as the forgetting hypothesis predicts – participants forget the middle verb’s prediction, the absence of the comma on V1 should not alert the reader that an embedded verb was expected: in the grammatical condition, they should experience the same difficulty at V1 that we found in the English experiments. Thus, the forgetting hypothesis predicts that German should show the same reading time patterns that English does. The alternative possibility is that the reader does not forget the prediction for the middle verb; this is equivalent to the parsing mechanism retaining the information that it is processing an embedded clause. If the parsing mechanism is expecting an embedded clause context (has not forgotten the middle verb’s prediction), it should trigger an error upon encountering the comma-less matrix verb in the grammatical condition. This predicts – contrary to the forgetting hypothesis – that processing should be more difficult at V1 and the Post-V1 word in the *ungrammatical* condition, compared to the grammatical one. We tested these opposing predictions using self-paced reading.

*Results*

Question-response accuracy (for all items, including distractors) was 79.51%, indicating that participants were attending to the task. The mean reading times and 95% confidence intervals at the verb regions are summarized in Figure 5 and Table 7, and the results in Table 8.

Region	Condition	Mean	CI.lower	CI.upper
V3	gram	1635	1471	1798
	ungram	1645	1484	1806
V2	gram	1517	1310	1724
	ungram	-	-	-
V1	gram	824	771	878
	ungram	1227	1090	1364
Post-V1	gram	1198	1103	1294
	ungram	1894	1673	2114

Table 7: Mean reading times and 95% confidence intervals for the verb regions and the post-V1 word in the German self-paced reading study (experiment 3).

Contrary to the predictions of the forgetting account, at the matrix verb V1 as well as the Post-V1 word, the ungrammatical condition was read *slower* than grammatical one. This is the opposite of the result found for English in experiments 1 and 2. We defer discussion of this surprising result and turn next to the eyetracking version of the German experiment.

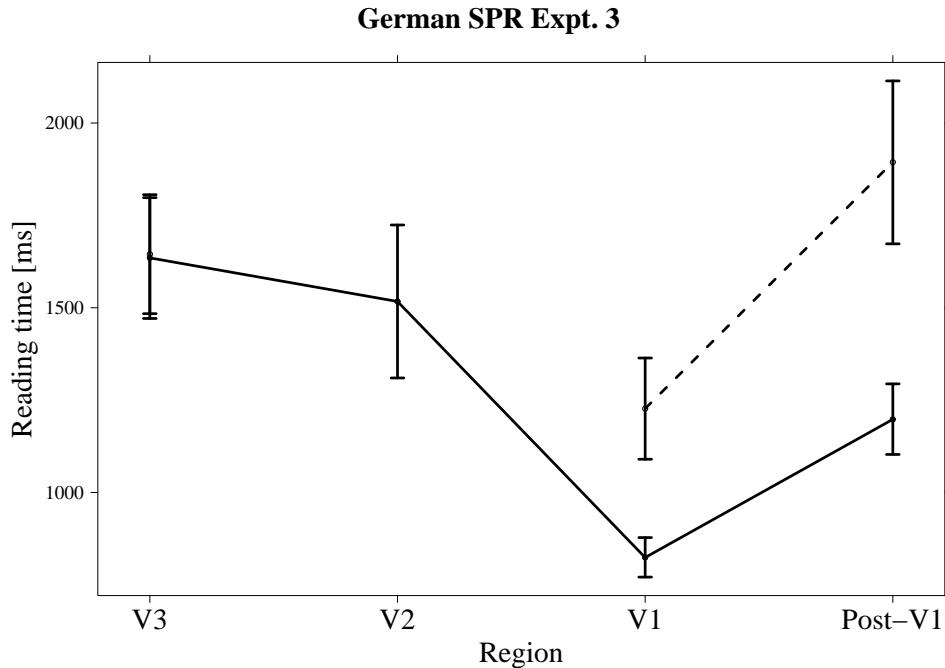


Figure 5. Mean reading times and 95% confidence intervals for the critical regions in the German self-paced reading study (experiment 3).

Comparison	Coefficient	Standard Error	t-score
gram-V1 vs ungram-V1	0.26503	0.03974	6.67 *
gram-V1+1 vs ungram-V1+1	0.35026	0.04166	8.41 *

Table 8: Summary of results of the comparisons for experiment 3. The asterisk (\*) indicates statistical significance at  $\alpha = 0.05$ .

#### Experiment 4: German eyetracking study

##### *Method*

##### *Participants.*

Fifty one native German speakers (undergraduates at the University of Potsdam) took part in this study, each receiving 7 Euros for participating. Participants were tested in individual sessions, and took approximately 30 minutes to complete the experiment. All had normal or corrected-to-normal vision.

##### *Procedure and Design.*

The procedure for this experiment is similar to that of the English experiment 2. There were some minor differences in procedure and apparatus, as discussed next. Participants were seated 55 cm from a 17" color monitor with  $1024 \times 768$  pixel resolution. The eyetracker used was an IView-X eye-tracker (SensoMotoric Instruments) running at 240 Hz

sampling rate, 0.025 degree tracking resolution, < 0.5 degree gaze position accuracy. Participants were asked to place their head in a frame and to position their chin on a chin-rest for stability. The angle per character was 0.26 degrees (3.84 characters per degree of visual angle).

Participants were asked to avoid large head movements throughout the experiment. A standard three-button mouse was used to record button responses. The presentation of the materials and the recording of responses was controlled by two PCs running proprietary software (the software used was Presentation, and SensoMotoric Instruments' own software for eyetracker control).

At the start of the experiment the experimenter performed a standard calibration procedure, which involves participants looking at a grid of thirteen fixation targets in random succession in order to validate their gazes. Calibration and validation were repeated after every 10-15 trials throughout the experiment, or if the experimenter noticed that measurement accuracy was poor (e.g., after large head movements or a change in the participant's posture).

As in the self-paced reading study, there were 60 distractor sentences and 16 stimulus sentences in each list (see appendix for the full list of stimuli), and each list was pseudo-randomly reordered. The trials per session were randomized once for each file, subject to the constraint that each session started with at least one filler.

### *Results and Discussion*

The mean response accuracy for all items in the experiment (distractors and targets) was 77.25%. The mean reading times and 95% confidence intervals are summarized in Figure 6 and Table 9, and the results of the statistical analyses in Table 10.

Region	Condition	Mean	CI.lower	CI.upper
V3	gram	843	731	955
	ungram	731	643	820
V2	gram	856	755	957
	ungram	-	-	-
V1	gram	770	672	867
	ungram	905	799	1010
Post-V1	gram	361	313	409
	ungram	446	384	509

Table 9: Mean reading times and 95% confidence intervals for the verb regions and the post-V1 word in the German eyetracking study (experiment 4).

As in the German SPR study (experiment 3), the matrix verb 1 in the ungrammatical condition had longer reading time than V1 in the grammatical condition. The reading time at the word following V1 showed no significant difference. This result is inconsistent with the forgetting hypothesis and consistent with the hypothesis that the middle verb's prediction is not forgotten.

We turn next to the conclusions we can draw from these experiments.

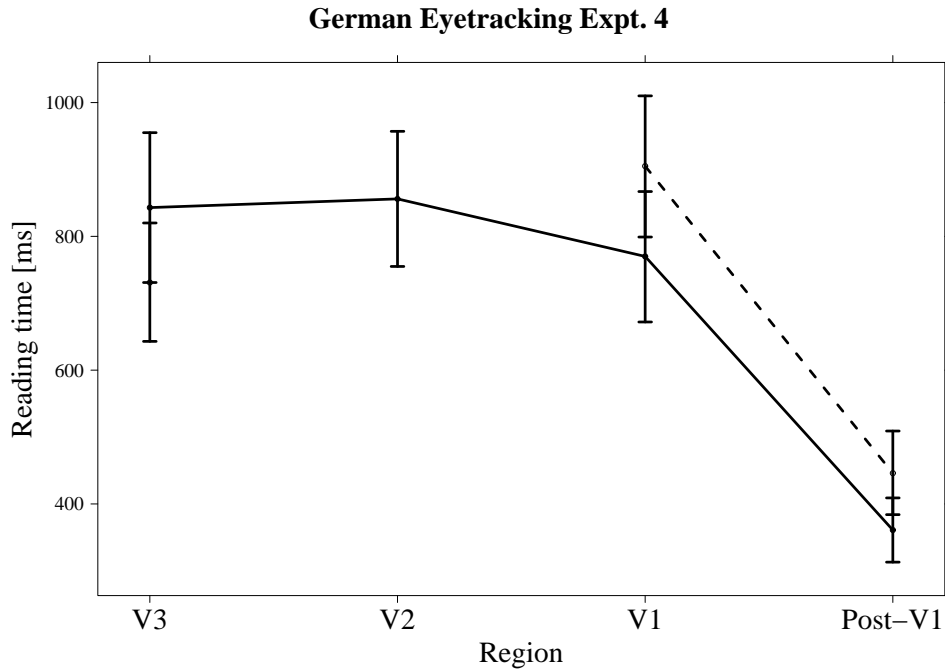


Figure 6. Mean reading times and 95% confidence intervals for the critical regions in the German eyetracking study (experiment 4).

Comparison	Coefficient	Standard Error	t-score
gram-V1 vs ungram-V1	0.20980	0.07497	2.80 *
gram-V1+1 vs ungram-V1+1	0.1746	0.1072	1.63

Table 10: Summary of results of the four comparisons for experiment 4. The asterisk (\*) indicates statistical significance at  $\alpha = 0.05$ .

### General discussion

The English self-paced reading and eyetracking results were consistent with the forgetting explanation: reading times were longer at the matrix verb in the grammatical condition compared to the ungrammatical one. The German data, however, is not: in both SPR and eyetracking, at the matrix verb V1 and/or the post-V1 word the grammatical condition had *shorter* reading times than the ungrammatical one. The German data thus suggest that the reader does not forget the middle verb. Since it is implausible that German speakers have, on average, higher working memory capacity than English speakers, this difference in performance is likely to stem from grammatical differences between the two languages. The most plausible explanation is the fact that English is a non-head final (SVO) language while German, at least in subordinate clauses, is a head-final language (SOV). One consequence of German head-finality is that – due to the frequent occurrence of head-final structures – predictions of upcoming verbs may have more robust memory representations than in

English. This could result in a reduced susceptibility to forgetting the upcoming verb's prediction, even in the face of increased memory load.

Independent evidence supports the above explanation. Classic locality effects disappear – indeed, they are reversed – in head-final languages like German (Konieczny, 2000) and Hindi (Vasishth, 2003) (Vasishth & Lewis, 2006). The central finding for these languages is that increasing argument-head distance results in shorter reading time at the verb, not longer as decay- or locality-based theories would predict. Diverse explanations have been offered for this difference between English and head-final languages. For example, Levy (2007), building on the work of Konieczny (2000) and Hale (2001), argues that the expectation of the upcoming verb may increase as the argument-head distance increases. Our finding is related to this idea: the expectation for a verb occurring head-finally may be stronger in a head-final language than in a non-head final one. The underlying reason for this stronger expectation is the grammatical constraint on German – its head-final nature.

What do these results mean for the forgetting explanation and decay-based theories in general? Whether the prediction is forgotten or not depends on the a priori grammatical constraints in the language, which in turn affects structural frequency patterns of the language. Predictions of verbs are maintained more strongly in head-final languages than non-head-final languages, and this independent factor determines whether or not a prediction is forgotten. In particular, memory overload as defined in the DLT or earlier versions of locality-based theories does not appear to be a language-independent constraint operating on online sentence comprehension.

## References

- Baayen, H. (2007). *Practical data analysis for the language sciences*. Cambridge University Press. (In Press)
- Bach, E., Brown, C., & Marslen-Wilson, W. (1986). Crossed and nested dependencies in German and Dutch: A psycholinguistic study. *Language and Cognitive Processes*, 1(4), 249–262.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Christiansen, M. H., & Chater, N. (2001). Finite models of infinite language: A connectionist approach to recursion. In M. H. Christiansen & N. Chater (Eds.), *Connectionist psycholinguistics* (p. 138-176). Westport, Connecticut: Ablex Publishing.
- Clark, H. (1973). The Language-as-Fixed-Effect Fallacy: A Critique of Language Statistics in Psychological Research. *Journal of Verbal Learning and Verbal Behavior*, 12(4), 335–59.
- Clifton Jr, C., Staub, A., & Rayner, K. (in press). Eye Movements in Reading Words and Sentences. In R. V. Gompel, M. Fisher, W. Murray, & R. L. Hill (Eds.), *Eye movements: A window on mind and brain*. Elsevier.
- Frazier, L. (1985). Syntactic complexity. In D. R. Dowty, L. Karttunen, & A. M. Zwicky (Eds.), *Natural language parsing: Psychological, computational, and theoretical perspectives* (pp. 129–189). Cambridge University Press.
- Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press.
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68, 1–76.

- Gibson, E. (2000). Dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, language, brain: Papers from the first mind articulation project symposium*. Cambridge, MA: MIT Press.
- Gibson, E., & Thomas, J. (1999). Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes*, *14*(3), 225–248.
- Gordon, P. C., Hendrick, R., Johnson, M., & Lee, Y. (2006). Similarity-Based Interference During Language Comprehension: Evidence from Eye Tracking During Reading. *Journal of Experimental Psychology: Learning Memory and Cognition*, *32*(6), 1304–1321.
- Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input. *Cognitive Science*, *29*, 261–290.
- Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics*. Pittsburgh, PA.
- Inhoff, A. W. (1984). Two stages of word processing during eye fixations in the reading of prose. *Journal of verbal learning and verbal behavior*, *23*(5), 612–624.
- Joshi, A. K., Becker, T., & Rambow, O. (2000). Complexity of scrambling: A new twist to the competence-performance distinction. In A. Abeillé & O. Rambow (Eds.), *Tree adjoining grammars* (pp. 167–181). CSLI.
- Just, M., & Carpenter, P. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*(4), 329–354.
- Just, M., & Carpenter, P. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, *99*(1), 122–149.
- Just, M., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, *111*(2), 228–238.
- Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research*, *29*(6), 627–645.
- Levy, R. (2007). *Expectation-based syntactic comprehension*. (In press, Cognition)
- Lewis, R. L., & Vasishth, S. (2005, May). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*, 1–45.
- Mitchell, D. C. (1984). An evaluation of subject-paced reading tasks and other methods of investigating immediate processes in reading. In D. E. Kieras & M. Just (Eds.), *New methods in reading comprehension research*. Hillsdale, N.J.: Erlbaum.
- Raaijmakers, J., Schrijnemakers, J., & Gremmen, F. (1999). How to deal with the “language-as-fixed-effect fallacy”: Common misconceptions and alternative solutions. *Journal of Memory and Language*, *41*(3), 416–426.
- Rambow, O., & Joshi, A. K. (1994). A processing model for free word order languages. In J. C. Clifton, C., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing*. Hillsdale, NJ: L. Erlbaum.
- Rayner, K., & Pollatsek, A. (1987). Eye movements in reading: A tutorial review. *Attention and performance XII: The psychology of reading*, 327–362.
- Roeck, A. D., Johnson, R., King, M., Rosner, M., Sampson, G., & Varile, N. (1982). A myth about center-embedding. *Lingua*, *58*, 327–340.

- Van Dyke, J., & Lewis, R. L. (2003). Distinguishing effects of structure and decay on attachment and repair: A cue-based parsing account of recovery from misanalyzed ambiguities. *Journal of Memory and Language*, 49, 285–316.
- Vasishth, S. (2003). *Working memory in sentence comprehension: Processing Hindi center embeddings*. New York: Garland Press. (Published in the Garland series Outstanding Dissertations in Linguistics, edited by Laurence Horn)
- Vasishth, S., Bruessow, S., Lewis, R. L., & Drenhaus, H. (2008 (to appear)). Processing polarity: How the ungrammatical intrudes on the grammatical. *Cognitive Science*.
- Vasishth, S., & Lewis, R. L. (2006, December). *Argument-head distance and processing complexity: Explaining both locality and antilocality effects*. (To appear in *Language*, Vol. 82.4)
- Warren, T. C., & Gibson, E. (2005). Effects of NP-type on reading English clefts. *Language and Cognitive Processes*, 89–104.

Appendix A  
Experiments 1 and 2 stimuli

- (1) a. The carpenter who the craftsman that the peasant carried hurt supervised the apprentice.  
b. Did the peasant carry the craftsman? Y
- (2) a. The carpenter who the pillar that the peasant carried hurt supervised the apprentice.  
b. Did the pillar hurt the carpenter? Y
- (3) a. The carpenter who the craftsman that the peasant carried supervised the apprentice.  
b. Did the carpenter supervise the apprentice? Y
- (4) a. The carpenter who the pillar that the peasant carried supervised the apprentice.  
b. Did the peasant carry the pillar? Y
- (5) a. The mother who the daughter that the sister found frightened greeted the grandmother.  
b. Did the sister find the daughter? Y
- (6) a. The mother who the gun that the sister found frightened greeted the grandmother.  
b. Did the gun frighten the mother? Y
- (7) a. The mother who the daughter that the sister found greeted the grandmother.  
b. Did the mother greet the grandmother? Y
- (8) a. The mother who the gun that the sister found greeted the grandmother.  
b. Did the sister find the gun? Y
- (9) a. The worker who the tenant that the foreman looked for injured questioned the shepherd.  
b. Did the foreman look for the tenant? Y
- (10) a. The worker who the bucket that the foreman looked for injured questioned the shepherd.  
b. Did the bucket injure the worker? Y
- (11) a. The worker who the tenant that the foreman looked for questioned the shepherd.  
b. Did the worker question the shepherd? Y
- (12) a. The worker who the bucket that the foreman looked for questioned the shepherd.  
b. Did the foreman look for the bucket? Y

- (13) a. The trader who the businessman that the professor hired confused annoyed the investor.  
b. Did the professor hire the businessman? Y
- (14) a. The trader who the computer that the professor hired confused annoyed the investor.  
b. Did the computer confuse the trader? Y
- (15) a. The trader who the businessman that the professor hired annoyed the investor.  
b. Did the trader annoy the investor? Y
- (16) a. The trader who the computer that the professor hired annoyed the investor.  
b. Did the professor hire the computer? Y
- (17) a. The painter who the musician that the father missed sheltered cooked for the artist.  
b. Did the father miss the musician? Y
- (18) a. The painter who the hut that the father missed sheltered cooked for the artist.  
b. Did the hut shelter the painter? Y
- (19) a. The painter who the musician that the father missed cooked for the artist.  
b. Did the painter cook for the artist? Y
- (20) a. The painter who the hut that the father missed cooked for the artist.  
b. Did the father miss the hut? Y
- (21) a. The saxophonist who the trumpeter that the conductor brought along distracted thanked the violinist.  
b. Did the conductor bring along the trumpeter? Y
- (22) a. The saxophonist who the baton that the conductor brought along distracted thanked the violinist.  
b. Did the baton distract the saxophonist? Y
- (23) a. The saxophonist who the trumpeter that the conductor brought along thanked the violinist.  
b. Did the saxophonist thank the violinist? Y
- (24) a. The saxophonist who the baton that the conductor brought along thanked the violinist.  
b. Did the conductor bring along the baton? Y
- (25) a. The pharmacist who the optician that the stranger saw troubled questioned the customer.

- b. Did the stranger see the optician? Y
- (26) a. The pharmacist who the button that the stranger saw troubled questioned the customer.  
b. Did the button trouble the pharmacist? Y
- (27) a. The pharmacist who the optician that the stranger saw questioned the customer.  
b. Did the pharmacist question the customer? Y
- (28) a. The pharmacist who the button that the stranger saw questioned the customer.  
b. Did the stranger see the button? Y
- (29) a. The cleaner who the janitor that the doctor recognized hurt surprised the patient.  
b. Did the doctor recognize the janitor? Y
- (30) a. The cleaner who the ball that the doctor recognized hurt surprised the patient.  
b. Did the ball hurt the cleaner? Y
- (31) a. The cleaner who the janitor that the doctor recognized surprised the patient.  
b. Did the cleaner surprise the patient? Y
- (32) a. The cleaner who the ball that the doctor recognized surprised the patient.  
b. Did the doctor recognize the ball? Y
- (33) a. The dancer who the singer that the bystander admired hurt tipped the doorman.  
b. Did the singer admire the bystander? N
- (34) a. The dancer who the shoe that the bystander admired hurt tipped the doorman.  
b. Did the shoe pinch the bystander? N
- (35) a. The dancer who the singer that the bystander admired tipped the doorman.  
b. Did the singer tip the doorman? N
- (36) a. The dancer who the shoe that the bystander admired tipped the doorman.  
b. Did the bystander admire the dancer? N
- (37) a. The artist who the sportsman that the guard shouted at annoyed instructed the newscaster.  
b. Did the sportsman shout at the guard? N
- (38) a. The artist who the computer that the guard shouted at annoyed instructed the newscaster.  
b. Did the computer annoy the guard? N

- (39) a. The artist who the sportsman that the guard shouted at instructed the newscaster.  
b. Did the sportsman instruct the newscaster? N
- (40) a. The artist who the computer that the guard shouted at instructed the newscaster.  
b. Did the guard shout at the artist? N
- (41) a. The clerk who the bureaucrat that the visitor forgotten about helped annoyed the neighbor.  
b. Did the bureaucrat forget about the visitor? N
- (42) a. The clerk who the walking stick that the visitor forgotten about helped annoyed the neighbor.  
b. Did the walking stick help the visitor? N
- (43) a. The clerk who the bureaucrat that the visitor forgotten about annoyed the neighbor.  
b. Did the bureaucrat annoy the neighbor? N
- (44) a. The clerk who the walking stick that the visitor forgotten about annoyed the neighbor.  
b. Did the visitor forget about the clerk? N
- (45) a. The son who the father that the teacher saw disturbed visited the grandfather.  
b. Did the father see the teacher? N
- (46) a. The son who the loudspeaker that the teacher saw disturbed visited the grandfather.  
b. Did the loudspeaker disturb the teacher? N
- (47) a. The son who the father that the teacher saw visited the grandfather.  
b. Did the father visit the grandfather? N
- (48) a. The son who the loudspeaker that the teacher saw visited the grandfather.  
b. Did the teacher see the son? N
- (49) a. The conductor who the choirmaster that the worker ignored hit berated the musician.  
b. Did the choirmaster ignore the worker? N
- (50) a. The conductor who the sponge that the worker ignored hit berated the musician.  
b. Did the sponge hit the worker? N
- (51) a. The conductor who the choirmaster that the worker ignored berated the musician.

- b. Did the choirmaster berate the musician? N
- (52) a. The conductor who the sponge that the worker ignored berated the musician.  
b. Did the worker ignore the conductor? N
- (53) a. The defence who the prosecutor that the spy looked at surprised convinced the judge.  
b. Did the prosecutor look at the spy? N
- (54) a. The defence who the knife that the spy looked at surprised convinced the judge.  
b. Did the knife surprise the spy? N
- (55) a. The defence who the prosecutor that the spy looked at convinced the judge.  
b. Did the prosecutor convince the judge? N
- (56) a. The defence who the knife that the spy looked at convinced the judge.  
b. Did the spy look at the defence? N
- (57) a. The cousin who the brother that the peasant described pleased hated the uncle.  
b. Did the brother describe the peasant? N
- (58) a. The cousin who the diamond that the peasant described pleased hated the uncle.  
b. Did the diamond please the peasant? N
- (59) a. The cousin who the brother that the peasant described hated the uncle.  
b. Did the brother hate the uncle? N
- (60) a. The cousin who the diamond that the peasant described hated the uncle.  
b. Did the peasant describe the cousin? N
- (61) a. The painter who the musician that the friend liked disturbed admired the poet.  
b. Did the musician like the friend? N
- (62) a. The painter who the film that the friend liked disturbed admired the poet.  
b. Did the film disturb the friend? N
- (63) a. The painter who the musician that the friend liked admired the poet.  
b. Did the musician admire the poet? N
- (64) a. The painter who the film that the friend liked admired the poet.  
b. Did the friend like the painter? N

## Appendix B

## Experiments 3 and 4 stimuli (German SPR and eyetracking)

- (1) a. Der Anwalt, den der Zeuge, den der Spion betrachtete, schnitt, überzeugte den Richter.  
b. Hat der Zeuge den Spion betrachtet? N
- (2) a. Der Anwalt, den der Säbel, den der Spion betrachtete, schnitt, überzeugte den Richter.  
b. Hat der Säbel den Spion betrachtet? N
- (3) a. Der Anwalt, den der Zeuge, den der Spion betrachtete, überzeugte den Richter.  
b. Hat der Zeuge den Spion betrachtet? N
- (4) a. Der Anwalt, den der Säbel, den der Spion betrachtete, überzeugte den Richter.  
b. Hat der Spion den Anwalt betrachtet? N
- (5) a. Der Beamte, den der Bürokrat, den der Besucher vergaß, stützte, verärgerte den Nachbarn.  
b. Hat der Bürokrat den Besucher vergessen? N
- (6) a. Der Beamte, den der Tisch, den der Besucher vergaß, stützte, verärgerte den Nachbarn.  
b. Hat der Tisch den Besucher gestützt? N
- (7) a. Der Beamte, den der Bürokrat, den der Besucher vergaß, verärgerte den Nachbarn.  
b. Hat der Bürokrat den Nachbarn verärgert? N
- (8) a. Der Beamte, den der Tisch, den der Besucher vergaß, verärgerte den Nachbarn.  
b. Hat der Besucher den Beamten vergessen? N
- (9) a. Der Bräutigam, den der Schwiegervater, den der Musiker trug, ablenkte, begrüßte den Pfarrer.  
b. Hat der Schwiegervater den Musiker getragen? N
- (10) a. Der Bräutigam, den der Bilderrahmen, den der Musiker trug, ablenkte, begrüßte den Pfarrer.  
b. Hat der Bilderrahmen den Musiker abgelenkt? N
- (11) a. Der Bräutigam, den der Schwiegervater, den der Musiker trug, begrüßte den Pfarrer.  
b. Hat der Schwiegervater den Pfarrer begrüßt? N
- (12) a. Der Bräutigam, den der Bilderrahmen, den der Musiker trug, begrüßte den Pfarrer.

- b. Hat der Musiker den Bräutigam getragen? N
- (13) a. Der Bruder, den der Cousin, den der Bauer fand, entzückte, hasste den Onkel.  
b. Hat der Cousin den Bauer gefunden? N
- (14) a. Der Bruder, den der Schmuck, den der Bauer fand, entzückte, hasste den Onkel.  
b. Hat der Schmuck den Bauer entzückt? N
- (15) a. Der Bruder, den der Cousin, den der Bauer fand, hasste den Onkel.  
b. Hat der Cousin den Onkel gehasst? N
- (16) a. Der Bruder, den der Schmuck, den der Bauer fand, hasste den Onkel.  
b. Hat der Bauer den Bruder gefunden? N
- (17) a. Der Zauberer, den der Akrobat, den der Zuschauer beschrieb, ärgerte, besuchte den Zirkusdirektor.  
b. Hat der Akrobat den Zuschauer beschrieben? N
- (18) a. Der Zauberer, den der Hut, den der Zuschauer beschrieb, ärgerte, besuchte den Zirkusdirektor.  
b. Hat der Hut den Zuschauer geärgert? N
- (19) a. Der Zauberer, den der Akrobat, den der Zuschauer beschrieb, besuchte den Zirkusdirektor.  
b. Hat der Akrobat den Zirkusdirektor besucht? N
- (20) a. Der Zauberer, den der Hut, den der Zuschauer beschrieb, besuchte den Zirkusdirektor.  
b. Hat der Zuschauer den Zauberer beschrieben? N
- (21) a. Der Einbrecher, den der Dieb, den der Mann beschützte, bezauberte, beschuldigte den Komplizen.  
b. Hat der Dieb den Mann beschützt? N
- (22) a. Der Einbrecher, den der Stein, den der Mann beschützte, bezauberte, beschuldigte den Komplizen.  
b. Hat der Stein den Mann bezaubert? N
- (23) a. Der Einbrecher, den der Dieb, den der Mann beschützte, beschuldigte den Komplizen.  
b. Hat der Dieb den Komplizen beschuldigt? N
- (24) a. Der Einbrecher, den der Stein, den der Mann beschützte, beschuldigte den Komplizen.  
b. Hat der Mann den Einbrecher beschützt? N

- (25) a. Der Neurotiker, den der Exzentriker, den der Psychiater bezahlte, ängstigte, versetzte den Berater.  
b. Hat der Psychiater den Exzentriker bezahlt? Y
- (26) a. Der Neurotiker, den der Dolch, den der Psychiater bezahlte, ängstigte, versetzte den Berater.  
b. Hat der Dolch den Neurotiker geängstigt? Y
- (27) a. Der Neurotiker, den der Exzentriker, den der Psychiater bezahlte, versetzte den Berater.  
b. Hat der Neurotiker den Berater versetzt? Y
- (28) a. Der Neurotiker, den der Dolch, den der Psychiater bezahlte, versetzte den Berater.  
b. Hat der Psychiater den Dolch bezahlt? Y
- (29) a. Der Arbeiter, den der Monteur, den der Vorarbeiter vergaß, verletzte, beschimpfte den Passanten.  
b. Hat der Vorarbeiter den Monteur vergessen? Y
- (30) a. Der Arbeiter, den der Eimer, den der Vorarbeiter vergaß, verletzte, beschimpfte den Passanten.  
b. Hat der Eimer den Arbeiter verletzt? Y
- (31) a. Der Arbeiter, den der Monteur, den der Vorarbeiter vergaß, beschimpfte den Passanten.  
b. Hat der Arbeiter den Passanten beschimpft? Y
- (32) a. Der Arbeiter, den der Eimer, den der Vorarbeiter vergaß, beschimpfte den Passanten.  
b. Hat der Vorarbeiter den Eimer vergessen? Y
- (33) a. Der Banker, den der Kreditgeber, den der Kunde mochte, nervte, bestahl den Vermieter.  
b. Hat der Kunde den Kreditgeber gemocht? Y
- (34) a. Der Banker, den der Geldautomat, den der Kunde mochte, nervte, bestahl den Vermieter.  
b. Hat der Geldautomat den Banker genervt? Y
- (35) a. Der Banker, den der Kreditgeber, den der Kunde mochte, bestahl den Vermieter.  
b. Hat der Banker den Vermieter bestohlen? Y
- (36) a. Der Banker, den der Geldautomat, den der Kunde mochte, bestahl den Vermieter.

- b. Hat der Kunde den Geldautomat gemocht? Y
- (37) a. Der Pianist, den der Cellist, den der Hausmeister sah, traf, ersetzte den Violinisten.  
b. Hat der Hausmeister den Cellist gesehen? Y
- (38) a. Der Pianist, den der Ball, den der Hausmeister sah, traf, ersetzte den Violinisten.  
b. Hat der Ball den Pianist getroffen? Y
- (39) a. Der Pianist, den der Cellist, den der Hausmeister sah, ersetzte den Violinisten.  
b. Hat der Pianist den Violinisten ersetzt? Y
- (40) a. Der Pianist, den der Ball, den der Hausmeister sah, ersetzte den Violinisten.  
b. Hat der Hausmeister den Ball gesehen? Y
- (41) a. Der Anwohner, den der Wanderer, den der Pförtner suchte, störte, verarztete den Verletzten.  
b. Hat der Pförtner den Wanderer gesucht? Y
- (42) a. Der Anwohner, den der Stuhl, den der Pförtner suchte, störte, verarztete den Verletzten.  
b. Hat der Stuhl den Anwohner gestört? Y
- (43) a. Der Anwohner, den der Wanderer, den der Pförtner suchte, verarztete den Verletzten.  
b. Hat der Anwohner den Verletzten verarztet? Y
- (44) a. Der Anwohner, den der Stuhl, den der Pförtner suchte, verarztete den Verletzten.  
b. Hat der Pförtner den Stuhl gesucht? Y
- (45) a. Der Tänzer, den der Artist, den der Zuschauer bewunderte, drückte, beobachtete den Einlasser.  
b. Hat der Zuschauer den Artist bewundert? Y
- (46) a. Der Tänzer, den der Schuh, den der Zuschauer bewunderte, drückte, beobachtete den Einlasser.  
b. Hat der Schuh den Tänzer gedrückt? Y
- (47) a. Der Tänzer, den der Artist, den der Zuschauer bewunderte, beobachtete den Einlasser.  
b. Hat der Tänzer den Einlasser beobachtet? Y
- (48) a. Der Tänzer, den der Schuh, den der Zuschauer bewunderte, beobachtete den Einlasser.  
b. Hat der Zuschauer den Schuh bewundert? Y