Anti-locality in English: Consequences for Theories of Sentence Comprehension

T. Florian Jaeger

Brain and Cognitive Sciences, University of Rochester

Evelina Fedorenko

Brain and Cognitive Sciences, MIT

Edward Gibson

Brain and Cognitive Sciences, MIT

Contact address:

T. Florian Jaeger
Assistant Professor
Brain and Cognitive Sciences,
University of Rochester,
Meliora Hall, Box 270268,
Rochester, NY 14627-0268

P: (585) 276 3611
E: fjaeger@bcs.rochester.edu
U: http://www.bcs.rochester.edu/people/fjaeger/

Abstract

In English, integration of an argument with its head generally leads to slower comprehension at the head (e.g., verb), the more material intervenes between the two. Such "locality" effects have been interpreted as evidence for memory-based comprehension accounts. Most memory-based comprehension accounts are, however, challenged by cases where more intervening material facilitates processing, which have been observed for head-final languages (e.g. German and Japanese). Such "anti-locality" effects are often interpreted as evidence for expectation-based processing. An alternative memory-based account of anti-locality is proposed in Vasishth, S., & Lewis, R. (2006, Argument-head distance and processing complexity: Explaining both locality and anti-locality effects. *Language*, 82(4), 767-794): intervening material can lead to repeated retrieval and hence reactivation of dependents. This reactivation can counteract activation-decay and lead to faster retrieval of the dependents at the head. We present new evidence of anti-locality from English relative clauses that existing memory-based accounts, including Lewis et al., 2006, cannot explain. We argue that memory-based and expectation-based models need to be integrated in order to account for the existence of locality and anti-locality effects. We also introduce a two-step mixed effect analysis of reading time data that incorporates controls for several potential confounds common in, but not restricted to, experiments that investigate locality and anti-locality effects.

Keywords: expectation-based comprehension; memory-based comprehension; syntactic processing; locality; anti-locality

Anti-locality: Consequences for Theories of Sentence Comprehension

When we use language, we relate entities and events to one another, creating dependencies between linguistic expressions such as nouns and verbs. As a consequence of the hierarchical structure of language and the sequential nature of language output, not all dependencies can be resolved locally. Comprehenders therefore often have to maintain elements in working memory until they can integrate these elements with their dependent(s). For example, in an English object-extracted relative clause (RC), as in (1), the noun phrase ("the reporter") is integrated with the RC verb ("attacked") only after the noun phrase "the senator" is processed.

(1) The reporter [who the senator attacked] admitted the error.

The resolution of non-local dependencies results in comprehension difficulty compared to the resolution of local dependencies (Gibson, 1998, 2000; Gibson & Wu, submitted; Gordon, Hendrick, & Johnson, 2001; Grodner & Gibson, 2005; Hsiao & Gibson, 2003; Levy, Fedorenko, & Gibson, 2007; Van Dyke & Lewis, 2003). These "locality" effects have been attributed to the difficulty of retrieving previously encountered elements from some memory resource involved in sentence processing (Gibson, 1998, 2000; Lewis & Vasishth, 2005; Lewis, Vasishth, & Van Dyke, 2006; Van Dyke & Lewis, 2003; Vasishth, 2003). In such memory-based models, locality effects are accounted for by activation-decay over time and / or interference of the intervening material (e.g. "the senator" in (1) above) with the representation of the to-be-retrieved dependent at the retrieval stage.

Memory-based accounts are, however, challenged by the existence of *anti-locality* effects. For example, German clause-final verbs are read faster if more words intervene between

the subject and the verb, even when the intervening material does not provide information about lexical properties of the upcoming verb (Konieczny, 2000; Konieczny & Döring, 2003). In Konieczny's (2000) Experiment 1, subject-modifying RCs either intervened between the subject and the verb or were extraposed following the verb, as in (2):

(2) Er hat die Rose (die wunderschön war) hingelegt (die wunderschoen war), und …
He has the rose (that was beautiful) laid_down (that was beautiful) and …
'He has laid down the rose that was beautiful, and …'

The matrix verb ("hingelegt") was read faster if the relative clause preceded it, despite the fact that the relative clause does not bias the comprehenders towards a specific verb (c.f. Grodner & Gibson, 2005:285 for discussion). A similar anti-locality effect has been observed in Hindi, where clause-final verbs are read faster if an additional adverbial phrase (e.g., "as soon as possible" in (3) below) intervenes between the subject and the verb (Vasishth, 2003):

(3) Sita-ne Hari-ko Ravi-ko kitaab-ko (jitnii.jaldii.ho.sake) khariid-neko bol-neko kahaa.
Sita-ERG Hari-DAT Ravi-DAT book-ACC (as.soon.as.possible) buy-INF tell-INF told
'*Sita told Hari to tell Ravi to buy the book (as soon as possible).*'

Vasishth and Lewis (2006) provide additional evidence for the facilitating effect of a variety of intervening phrases (subject- and object-extracted RCs, PPs, and adverbs), all of which were set up so as not to predict a specific matrix verb. Nakatani & Gibson (2008) found similar results for Japanese structures.

Anti-locality effects have been attributed to expectation-based processing (Hale, 2001; Konieczny, 2000; Konieczny & Döring, 2003; Levy, 2008; Narayanan & Jurafsky, 2002): in examples like (2), the probability that the next word is the matrix verb is higher after an intervening relative clause (*Er hat die Rose, die wunderschön war* … "He has the rose that was beautiful") than immediately after the matrix subject (*Er hat die Rose*, "He has the rose"), simply because additional interveners become less and less likely. According to expectation-based comprehension accounts, it is this increase in probability of a matrix verb that causes the faster processing of the matrix verb following the relative clause. One appealing feature of expectation-based accounts is that they attribute anti-locality effects to the same source as lexical expectation effects (Altmann & Kamide, 1999; Kamide, Altmann, & Haywood, 2003; Staub & Clifton, 2006).

An alternative account of anti-locality effects within the memory-based framework of Lewis and Vasishth (2005) is proposed by Vasishth and Lewis (2006). Lewis and Vasishth (2005) present an ACT-R model that combines activation-decay and interference during retrieval (to account for locality effects) with the possibility of repeated *re*-activation of dependents (to account for anti-locality effects). Intervening material can lead to reactivation of previously encountered elements, thereby increasing their activation levels and facilitating their retrieval. For example, in (2) the intervening relative clause (*die wunderschön war* "that was beautiful") reactivates the object (*die Rose* "the rose"), making retrieval of this noun phrase (NP) faster when the matrix verb (*hingelegt* "laid down") is processed. If the activation boost of an element due to reactivation is higher than the decay since its most recent reactivation, the account predicts faster retrieval of that element at the integration point. If the combined advantage of all

elements that are retrieved (e.g., the subject *Er* ("he") and the object *die Rose* ("the rose") in (2) above) outweighs their total activation-decay, the proposed account predicts faster reading times at the verb and hence an anti-locality effect. Similarly, for materials like (3), the adverb retrieves and hence reactivates the predicted verb phrase (VP) node that it attaches to. This reactivation boosts the activation of the upcoming VP node, which leads to faster processing of the head ("buy"). If, on the other hand, activation-decay due to intervening material is higher than the associated gain from repeated activation (Christianson, 2002), or if the nature of the intervening material is such that it does not cause reactivation (Van Dyke & Lewis, 2003), the account predicts a locality effect. Vasishth and Lewis (2006) show that their model successfully accounts for a variety of locality and anti-locality results in the literature.

In this paper, we present new data from two experiments investigating English relative clauses that demonstrate that existing memory-based accounts, including the one proposed by Lewis and colleagues, are insufficient to account for all anti-locality effects. Taken together with previous results, our findings argue that an adequate theory of comprehension will need to refer to both memory retrieval and expectation-based processing.

Experiment 1

In the materials for Experiment 1, the distance between the subject and the verb was varied by including subject-modifying RCs of different lengths: with one, two, or three prepositional phrases (PPs) after the RC verb.

(4) The understudy [$_{RC}$ that telephoned the agent ((about the job) in Los Angeles) an hour ago] <u>shared the story</u> and felt relieved immediately.

Additional PPs in the RC increase the distance between the matrix subject ("the understudy" in (4)) and the verb ("shared" in (4)). Memory-based accounts (Gibson, 1998, 2000; Lewis, 1998; Lewis & Vasishth, 2005; Vasishth, 2003) therefore predict that the retrieval of the subject should be more difficult at the verb in these cases. Even the memory-based account proposed by Lewis and Vasishth (2005, henceforth LV05) predicts a locality effect on the matrix verb for examples like (4). Consider the incremental parse of (4) given in Figure 1. The crucial nodes are the matrix nodes, the matrix subject ("the understudy") and the matrix IP, which is predicted by the matrix subject (for details, see Lewis & Vasishth, 2005, pp. 390-391). Only if processing of the PPs was to involve retrieval and hence reactivation of any of these matrix nodes would LV05 predict an anti-locality effect.

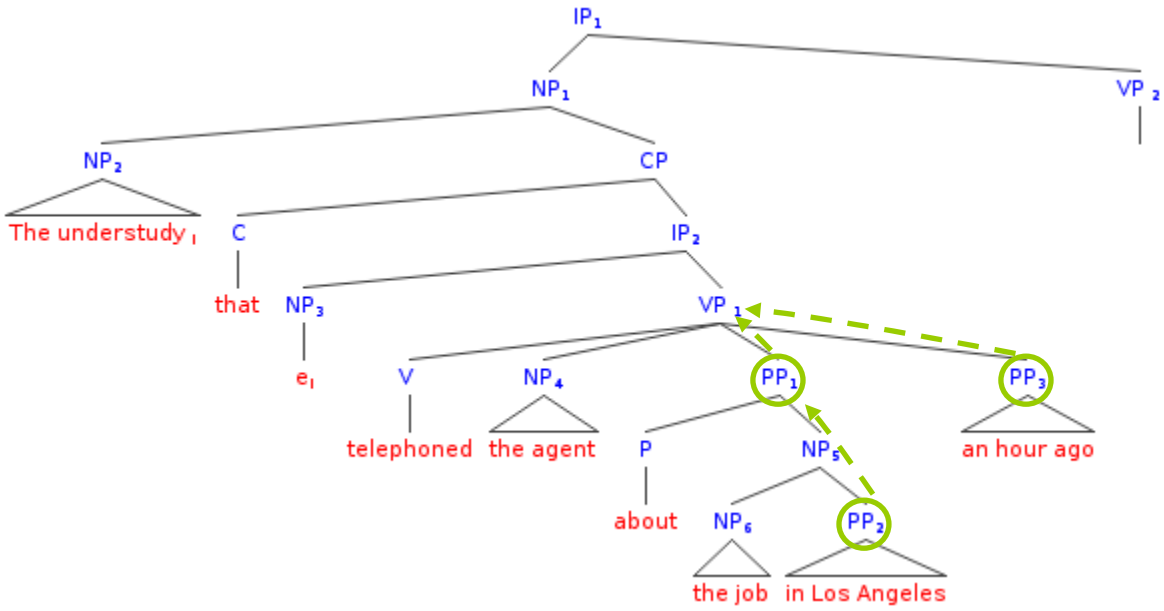

Figure 1 **Incremental parse of (4) as assumed in Lewis and Vasishth (2005). Dashed arrows indicate retrieval cued by the circled phrase.**

However, as can be seen in Figure 1, the PPs only involve retrieval of the RC's VP or a preceding PP in the RC (e.g. "in Los Angeles" is likely to attach to the preceding PP, but could

also attach to the RC's VP, leading to a different interpretation; for the current argument, it does not matter which parse comprehenders choose). This is indicated by the circles and dashed arrows in Figure 1. Neither the previously processed matrix subject NP nor the predicted matrix IP or VP nodes are retrieved during the processing of the PPs. This means that processing the PPs in the RC should not facilitate processing of the matrix verb. To the contrary – since time passes for each PP that needs to be processed – the activation of all previously processed and predicted matrix nodes decays further, the more PPs the RC contains. Processing of the matrix verb requires retrieval of the matrix IP (containing the matrix subject and the matrix VP). Hence, LV05 makes the same prediction as other memory-based accounts: a locality effect on the matrix verb for stimuli like (4).

Expectation-based accounts make the opposite prediction, because the matrix verb becomes increasingly more probable after each additional word following the RC verb. The increase in the probability of a matrix verb for increasingly longer relative clauses is shown in Figure 2 (probability estimates taken from Levy (2008). The probability estimates were derived from the Penn Treebank portion (release 3, Marcus et al., 1999) of the Brown corpus (a corpus consisting of texts from a variety of written genres, Francis & Kucera, 1961) using an unlexicalized probabilistic context-free grammar (Levy, 2008). Generally, the probability of a matrix verb increases with increasing post-verbal relative clause length, independent of the structural properties of the post-verbal material. It is also worth noting that the probability of a matrix verb increases more from 1 PP to 2 PPs, compared to from 2 PPs to 3 PPs. We will return to this issue in the discussion of Experiment 1.

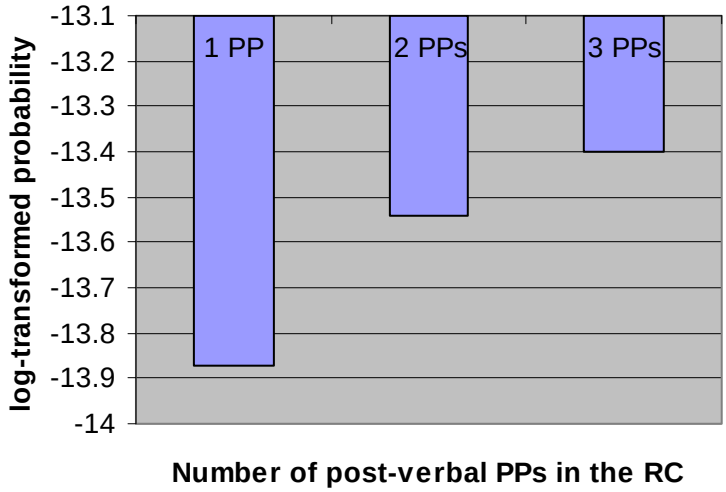

Figure 2 Log-transformed probability that the next word will be the matrix verb after 1 PP, 2PPs, or 3 PPs in an RC (probability estimates taken from Levy, 2008).

To ensure that the hypothesized effects would be due to syntactic expectation (increased probability of *any* matrix verb), the PPs in the RC were chosen in such a way that they did not bias towards the *specific* matrix verb (in contrast to stimuli like "*The fisherman who was cutting onions cried"*; Grodner and Gibson, 2005:285). Expectation-based comprehension accounts (Hale, 2001; Levy, 2008; Narayanan & Jurafsky, 2002) then predict that the matrix verb should be read faster, the more PPs the intervening RC contains.

Design

Experiment 1 used a 3 x 3 design crossing RC length and RC type in stimuli like (4) above. The RC intervening between the matrix subject and the verb contained one, two, or three PPs after the RC verb. The intervening RC was a subject-extracted relative clause (SRC) or an object-extracted relative clause (ORC) either with or without the relativizer *that* (which is only optional for ORCs). This second manipulation (RC type) was included as a control manipulation. After

non-pronoun subjects, ORC verbs are processed more slowly than SRC verbs (Gibson, 1998; Gordon et al., 2001; Gordon, Hendrick, & Johnson, 2004; Grodner & Gibson, 2005). There also exists evidence that verbs in ORCs with *that* are processed faster than verbs in ORCs without *that* (Hakes & Cairns, 1970; Hakes, Evans, & Brannon, 1976; Jaeger, 2007; Race & MacDonald, 2003).

Participants

Thirty-nine participants from MIT and the surrounding community were paid for their participation. All were native speakers of English and were naive as to the purposes of the study.

Stimuli

We created thirty-six items in nine (3 x 3) conditions. To avoid potential confounds due to spill-over from the pre-verbal region onto the matrix verb (cf. discussion in Vasishth & Lewis, 2006, p. 773ff.), the local context around the matrix verb was kept identical for all conditions. The overall length of the stimuli was kept constant by attaching the PPs that were not attached to the RC verb to the matrix verb (note that in masked self-paced moving-window reading, stimuli are displayed with a dash for each letter and spaces for word boundaries, so that participants have an estimate of the stimulus complexity from the stimulus onset). The PPs were also chosen in such a way that they were plausible modifiers of either the embedded verb or the matrix verb. An example item in all 9 conditions is given below:

(5) **1 PP, 2 PPs, or 3 PPs in ORC with *that*:** The understudy [$_{RC}$ that the agent telephoned (about the job / about the job in Los Angeles) an hour ago <u>shared the story</u> (about the job / about the job in Los Angeles) and felt relieved immediately.

(6) **1 PP, 2 PPs, or 3 PPs in ORC without *that*:** The understudy [$_{RC}$ the agent telephoned (about the job / about the job in Los Angeles) an hour ago shared the story (about the job / about the job in Los Angeles) and felt relieved immediately.

(7) **1 PP, 2 PPs, or 3 PPs in SRC:** The understudy [$_{RC}$ that telephoned the agent (about the job / about the job in Los Angeles) an hour ago shared the story (about the job / about the job in Los Angeles) and felt relieved immediately.

Experimental stimuli were distributed over nine lists following a Latin Square design, so that each participant saw only one condition of every item and so that each condition appeared equally often (four times) in each list. In addition to the target sentences, 49 sentences from two unrelated experiments and 20 filler sentences with a variety of structures were included. The length and syntactic complexity of the sentences in the two unrelated experiments and of the fillers were similar to that of the target sentences. The stimuli were pseudo-randomized separately for each participant with at least one filler or one experimental item from an unrelated experiment separating the target sentences.

Procedure

The task was self-paced moving-window presentation (Just, Carpenter, & Woolley, 1982). The moving-window advanced word by word except in a few cases when several words that formed one unit (e.g. *city hall*) were presented together. The experiment was run using the Linger 2.85 software by Doug Rohde. The amount of time the participant spent reading each word was recorded as the time between key-presses.

To assure that the participants read the sentences for meaning, a comprehension question was presented at the end of each trial, asking about the propositional content of the sentence.

Participants pressed one of two keys to respond "yes" or "no". After an incorrect response, the word "INCORRECT" flashed briefly.

Before the experiment started, a short list of practice items and questions was presented in order to familiarize the participants with the task. Participants took approximately 40 minutes to complete the experiment.

Comprehension question performance Across conditions, participants answered the comprehension question correctly 65.9% of the time. Table 1 presents the mean accuracies across the nine conditions of Experiment 1. A two-way ANOVA crossing RC type (SRCs, ORC with *that*, and ORC without *that*) and RC length (1 PP, 2 PPs, 3PPs) on the responses to the comprehension questions revealed a main effect of RC type, but no other effects ($F_s<1$). The two ORC conditions resulted in lower accuracies than the SRC condition ($F1(2,76)=4.14$; $MSe=2545$; $p < .05$; $F2(2,70)=7.12$; $MSe=2409$; $p<.005$).

The low overall percentage of correct answers is likely to be due in part to the complexity of the stimuli. In addition, some of the comprehension questions may have been misleading. In particular, half of the questions were designed to test whether participants parsed the PP attachments in the RC correctly. In a small subset of these questions one or more of the PPs could be interpreted as either modifying the RC verb or modifying the matrix verb. Therefore, some of the responses that were coded as incorrect (because the interpretation was inconsistent with the intended PP attachment) actually reflect a possible alternative interpretation. (Excluding the items asking about the information in the prepositional phrases increases the overall accuracy level to 72.7%.)

RC type	RC length		
	1 PP	2 PPs	3PPs
Subject-extracted RC (SRC)	71.2 (4.0)	71.8 (3.6)	69.9 (3.5)
Object-extracted RC (ORC) with *that*	63.5 (3.0)	59.6 (3.6)	62.2 (3.8)
Object-extracted RC (ORC) without *that*	64.1 (3.4)	64.7 (4.1)	66.0 (3.5)

Table 1 Comprehension accuracies in percent correct, as a function of RC type and RC length in Experiment 1 (standard errors in parentheses).

Reading times Because the comprehension question accuracies were not very high, we analyzed all the trials, regardless of whether the comprehension question was answered correctly (excluding the trials with incorrect answers leads to qualitatively similar results; furthermore, in the section "Meta-analysis of Experiments 1 and 2" we present a mixed effect analysis that includes answer performance in the model). Reading times more than three standard deviations away from the mean for a position within condition were removed from the analysis, excluding 1.6% of the data.

For the purposes of the analysis, we divided each sentence into several regions, as shown for the subject-extracted and the object-extracted versions below: (1) the matrix NP (e.g., *The understudy*), (2) the complementizer (absent in ORCs without *that*), (3) RC verb and RC object/subject (e.g., *telephoned the agent* or *the agent telephoned*), (4) PP1 (e.g., *about the job*), (5) PP2 (e.g., *in Los Angeles*), (6) PP3 (e.g., *an hour ago*), (7) the matrix verb (e.g., *shared*), (8) two words after the matrix verb (e.g., *the story*), and (9) the sentence ending (e.g., *and felt relieved immediately).*

(8) **SRC:** The understudy / that / telephoned the agent / about the job / in Los Angeles / an hour ago / shared / the story / and felt relieved immediately.

(9) **ORC:** The understudy / (that) / the agent telephoned / about the job / in Los Angeles / an hour ago / shared / the story / and felt relieved immediately.

The critical regions for the RC length manipulation were defined as the matrix verb and the two following words (underlined in the example above). The critical region for the RC type manipulation was defined as the RC region, which included the RC verb and the RC subject/object (depending on the extraction type). We will first present the results for the critical RC length manipulation, followed by the presentation of the results for the RC type manipulation.

RC length manipulation

At the matrix verb, we observed two main effects. First, there was a main effect of RC length ($F1(2,76)=9.06$; $MSe=304955$; $p<.001$; $F2(2,70)=9.76$; $MSe=248795$; $p<.001$). The matrix verb was processed the fastest in the conditions with 3PPs modifying the RC verb (381 ms), followed by the conditions with 2PPs (392 ms), with the conditions with 1PP leading to the slowest reading times (474 ms). Second, there was a main effect of RC type ($F1(2,76)=12.1$; $MSe=204168$; $p<.001$; $F2(2,70)=12.5$; $MSe=170845$; $p<.001$). The matrix verb was processed faster in SRCs (371 ms), compared to ORCs with *that* (453 ms) and to ORCs without *that* (424 ms). Figure 3 shows the reading times on the matrix verb for the nine conditions.

At the following region (the two words after the matrix verb), we again observed two main effects. First, there was a main effect of RC length ($F1(2,76)=3.69$; $MSe=17190$; $p<.05$; $F2(2,70)=3.66$; $MSe=19556$; $p<.05$). The matrix verb was processed faster in the conditions

with 3PPs modifying the RC verb (362 ms) and in the conditions with 2PPs modifying the RC verb (361 ms), compared to the conditions with 1PP (382 ms). Second, there was a main effect of RC type (F1(2,76)=9.27; MSe=40794; p<.001; F2(2,70)=10.4; MSe=38633; p<.001). The matrix verb was processed faster in SRCs (347 ms), compared to ORCs with and without *that* (382 ms and 376 ms, respectively). We also observed a marginal interaction in the by-participants analysis, which was not significant in the by-items analysis.

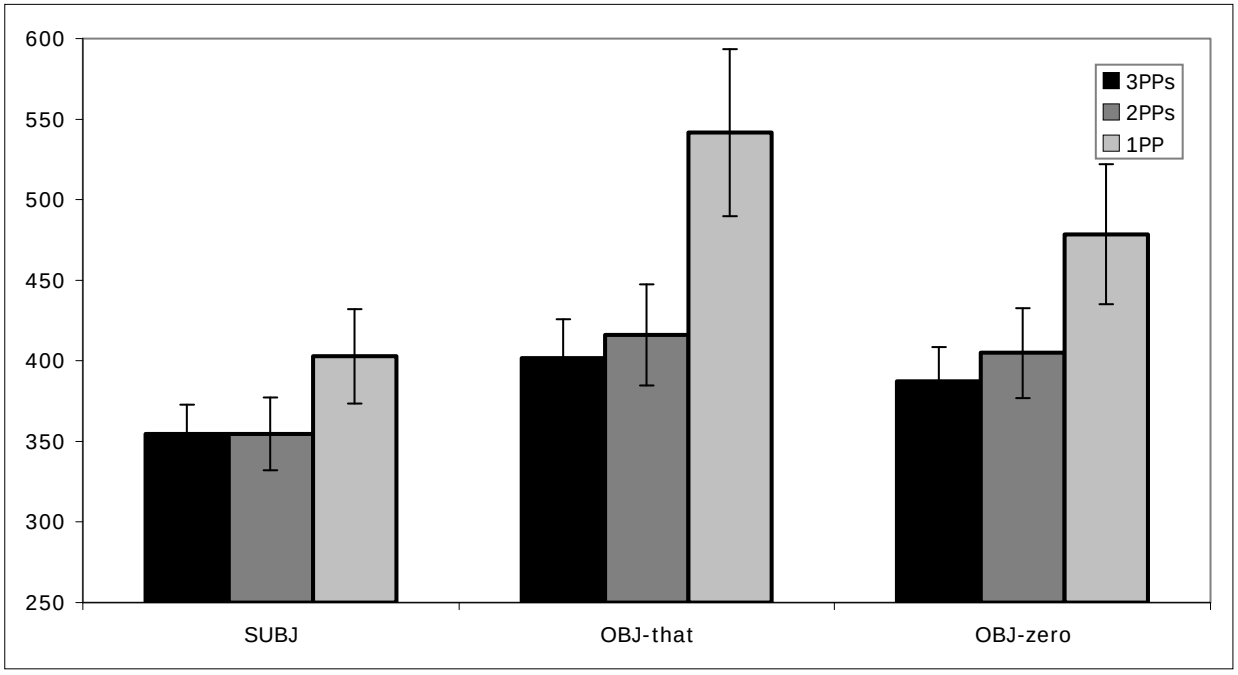

Figure 3 Reading times at the matrix verb in Experiment 1. Error bars represent +/- 1 standard error.

RC type manipulation

At the RC region we observed a main effect of RC type (F1(2,76)=24.9; MSe=360337; p<.001; F2(2,70)=28.7; MSe=314047; p<.001). The RC region was processed the fastest in the subject-extracted conditions (371 ms), slower in the object-extracted conditions with complementizer (406 ms), and the slowest in the object-extracted conditions without complementizer (480 ms).

All pair-wise comparisons were significant (Fs >8; ps<.01). Figure 4 shows the reading times on the RC region, collapsing across the RC length manipulation. We also observed a marginal interaction in the by-participants analysis, which was not significant in the by-items analysis. There is no reason to expect an effect of the RC length factor at this point in the sentence, or an interaction between the two factors, because in this region the materials are identical in the different levels of RC length.

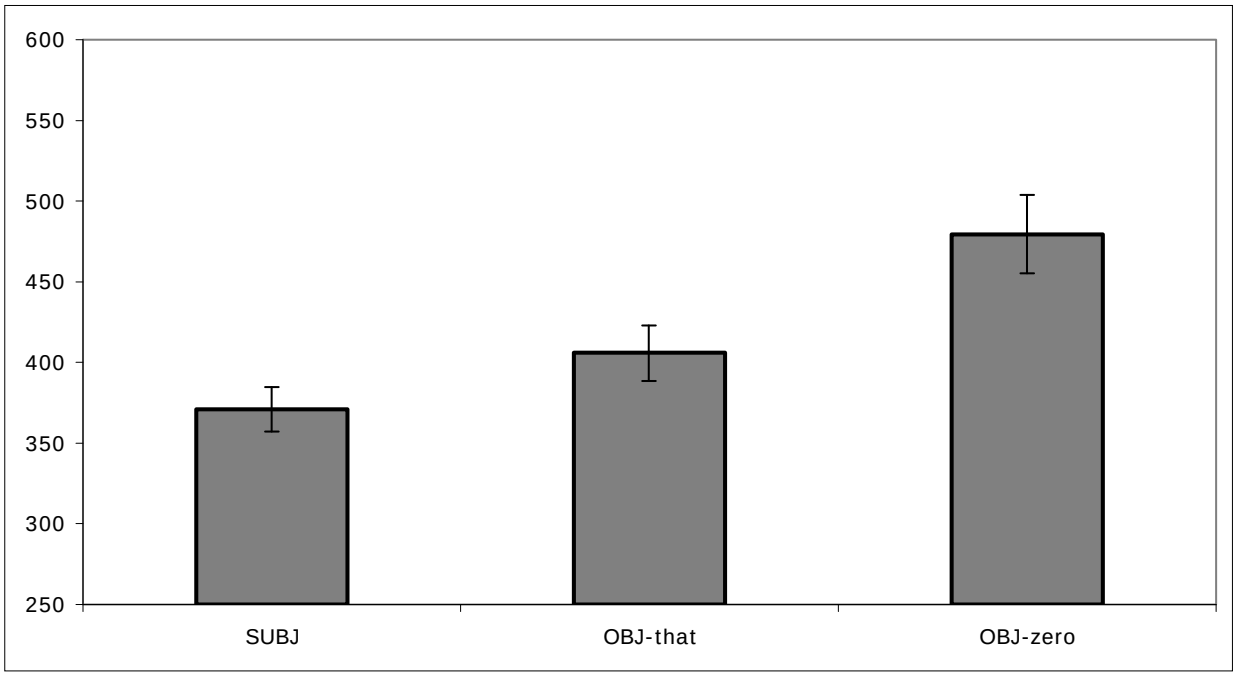

Figure 4 Reading times at the RC region in Experiment 1 (collapsing across the RC length manipulation). Error bars represent +/- 1 standard error.

Discussion

The results for the matrix verb region match the predictions of expectation-based theories, and are not predicted by memory-based theories. The longer the RC and hence the more expected the matrix verb (cf. Figure 2), the faster the matrix verb is read. Furthermore, it appears that the

RC length effect is mostly driven by the contrast between the 1 PP condition and conditions with more PPs in the RC (see Figure 3), consistent with the estimates of the matrix verb probability given in Figure 2. In addition to the anti-locality effect on the matrix verb, we also observed a locality effect during the RC region for our control manipulation of the RC type, as predicted by memory-based theories. RCs are read faster in the SRC conditions than in the ORC conditions, and RCs in the ORC conditions with *that* are read faster than in ORC conditions without *that* (for stimuli of the type used here with complex ORC subjects, Race & MacDonald, 2003 observe the same). Experiment 1 therefore provides evidence for both memory-based accounts (at the RC region) and expectation-based accounts (at the matrix verb). We postpone further discussion of the consequences of these results for theories of sentence comprehension until the general discussion. Some readers may be struck by the fact that our results seem to be in conflict with earlier results from very similar stimuli (Bartek, Smith, Lewis, & Vasishth, in prep; Grodner & Gibson, 2005). We return to this issue in the general discussion, too, after we have addressed two potential confounds of Experiment 1.

The distance of the matrix verb to the beginning of the stimulus differs among the levels of RC length (cf. (5) - (7) above). Given that the position of a word in a sentence influences the speed with which it is read, with later words generally being read faster (Demberg & Keller, to appear; Ferreira & Henderson, 1993, p. 261; Just et al., 1982, p. 232), it could be that the observed results are due to a positional confound. This potential position confound is addressed in Experiment 2.

It is also possible that the effect of RC length is driven by 'spill-over' from the preceding region(s). The immediately preceding words before the matrix verb are held constant across the three levels of RC length (cf. (5) - (7)), preventing a spill-over confound due to random

differences in the difficulty associated with the lexical processing of the preceding words (unlike in Vasishth & Lewis, 2006, Experiment 1). The RC length conditions differ, however, in terms of the distance between the matrix verb and the RC verb (3, 6, and 9 words for the 1PP, 2PPs, and 3PPs conditions, respectively). Given that the RC verb is a point of high processing load, higher matrix verb reading times after shorter RCs could possibly be due to spill-over of that difficulty, rather than being due to expectation-based processing.

We first present Experiment 2, which addresses the potential position confound, and then we address the potential spill-over confound, as well as other potential confounds, using mixed effect regression analysis for the combined data from Experiments 1 and 2.

Experiment 2

Experiment 2 addresses the potential positional confound from Experiment 1 by manipulating RC length while keeping the distance of the matrix verb from the beginning of the stimulus constant.

Participants

Forty-eight participants from MIT and the surrounding community were paid for their participation. All were native speakers of English and were naive as to the purposes of the study. None participated in Experiment 1.

Design and Stimuli

Experiment 2 used a 3 x 2 design crossing RC length and RC type. First, as in Experiment 1, the RC intervening between the matrix subject and the verb contained one, two, or three PPs

modifying the RC verb. Second, the intervening RC (always object-extracted in this experiment) either did or did not contain a complementizer. Unlike in Experiment 1, the distance (in words) of the matrix verb to the sentence onset was kept constant, which was achieved by topicalizing PPs that were not attached to the RC verb (rather than attaching them to the matrix verb, as in Experiment 1).

The materials consisted of 30 items (a subset of the 36 sets of items used in Experiment 1) in six conditions. An example item is given below:

(10) **1 PP, 2 PPs, or 3 PPs in ORC with or without *that*:** (*In the spring* / *In the spring before the takeover*) the client (that) the retailer contacted (*in the spring / in the spring before the takeover*) about hazardous chemicals broke the contract and […]

Experimental stimuli were distributed over six lists following a Latin Square design, so that each participant saw only one condition of every item and so that each condition appeared equally often (five times) in each list. In addition to the target sentences, 49 sentences from two unrelated experiments and 20 filler sentences with a variety of structures were included. The length and syntactic complexity of the sentences in the two unrelated experiments and of the fillers was similar to that of the target sentences. The stimuli were pseudo-randomized separately for each participant with at least one filler separating the target sentences.

Results

Comprehension question performance Across conditions, participants answered the comprehension question correctly 75.7% of the time. Table 2 presents the mean accuracies across the six conditions of Experiment 2. A two-factor ANOVA crossing RC type (ORCs with

that and ORCs without *that*) and RC length (1 PP, 2 PPs, 3PPs) on the responses to the comprehension questions revealed no effects (Fs<1.5).

RC type	RC length		
	1 PP	2 PPs	3PPs
ORC with *that*	74.2 (2.7)	79.2 (3.0)	75.4 (2.8)
ORC without *that*	72.9 (3.2)	76.3 (3.4)	76.3 (3.1)

Table 2 Comprehension accuracies in percent correct, as a function of RC type and RC length in Experiment 2 (standard errors in parentheses).

Reading times As with Experiment 1, we analyzed all the trials, regardless of whether the comprehension question was answered correctly. Reading times more than three standard deviations away from the mean for a position within condition were removed from the analysis, excluding 1.8% of the data.

For the purposes of the analysis, we divided each sentence into several regions, as shown below (the division into regions and the critical regions for the two manipulations are the same as in Experiment 1):

(11) The understudy / (that) / the agent telephoned / about the job / in Los Angeles / an hour ago / shared / the story / and felt relieved immediately.

We first present the results for the RC length manipulation, followed by the presentation of the results for the RC type manipulation.

RC length manipulation

At the matrix verb, we observed a main effect or RC length and an interaction. First, there was a main effect of RC length ($F1(2,94)=6.25$; $MSe=68123$; $p<.005$; $F2(2,58)=6.57$; $MSe=41302$; $p<.005$). The matrix verb was processed the fastest in the conditions with 3PPs modifying the RC verb (403 ms), followed by the conditions with 2PPs (423 ms), with the conditions with 1PP leading to the slowest reading times (456 ms). Second, there was an interaction between the two factors, which did not reach significance in the by-items analysis ($F1(2,94)=3.50$; $MSe=37210$; $p<.05$; $F2(2,58)=2.72$; $MSe=23443$ $p=.075$). Figure 5 shows the reading times on the matrix verb for the nine conditions.

At the following region (the two words after the matrix verb), we again observed a main effect of RC length ($F1(2,94)=5.73$; $MSe=20417$; $p<.005$; $F2(2,58)=6.76$; $MSe=12869$; $p<.005$). There was also a marginal effect of RC type in the by-participants analysis, which was not significant in the by-items analysis ($F1(1,47)=2.87$; $MSe=11607$; $p=.097$; $F2(1,29)=1.49$; $MSe=8715$; n.s.). The interaction between the two factors was not significant ($Fs<1$, $ps>.5$).

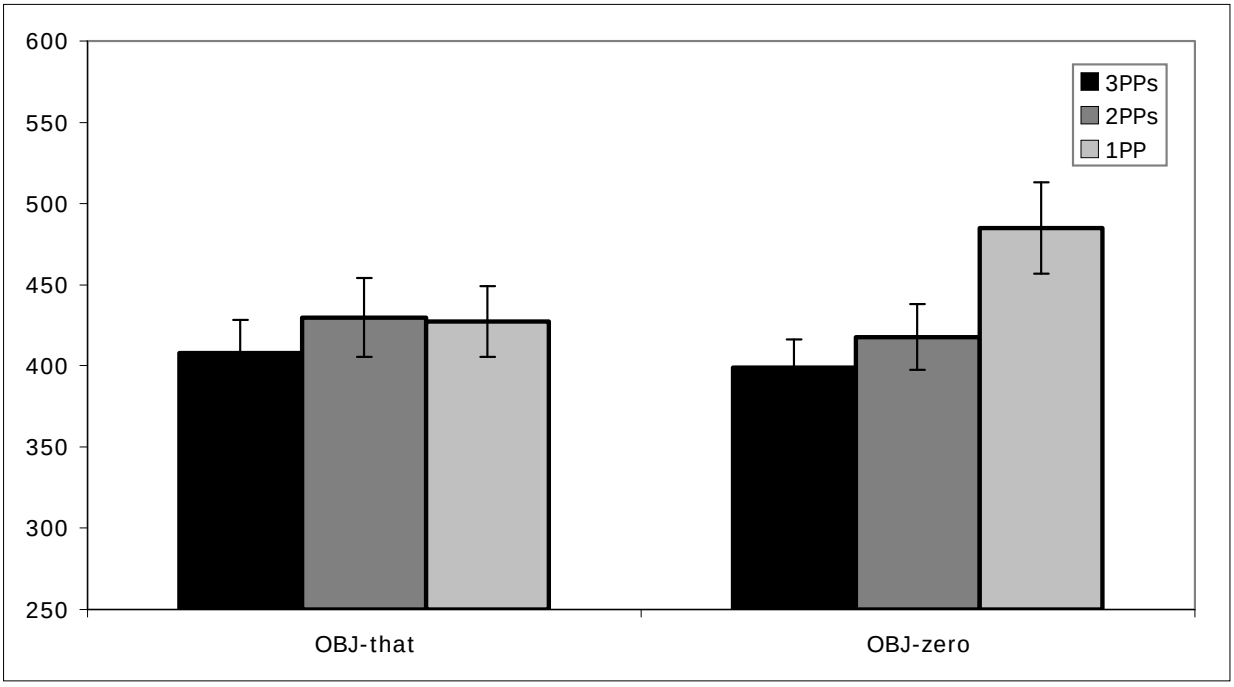

Figure 5 Reading times at the matrix verb in Experiment 1. Error bars represent +/- 1 standard error.

RC type manipulation

At the RC region we observed a main effect of RC type ($F1(1,47)=54.8$; $MSe=248422$; $p<.001$; $F2(1,29)=15.1$; $MSe=152993$; $p<.002$) and an interaction between the two factors ($F1(2,94)=7.52$; $MSe=48735$; $p<.002$; $F2(2,58)=9.84$; $MSe=29058$; $p<.001$). First, the RC region was processed faster in the ORCs with *that* (396 ms), compared to the ORCs without *that* (454 ms). And second, the two factors appeared to interact in a way such that RCs in the ORC conditioned with *that* are read the slowest in the 2PP condition (413 ms), but RCs in the ORC conditions without *that* are read slowest in the 1PP condition (492 ms). This interaction was not predicted.

Discussion

Experiment 2 replicates the RC length effect while avoiding the potential positional confound from Experiment 1. The control manipulation of RC type also came out as expected with RCs in the ORC conditions with *that* being processed faster than in the ORC conditions without *that*. However, we also observed an unexpected and non-linear interaction between RC length and RC type during the RC region. It is possible that this pattern is due to the lack of spillover modeling in the ANOVA (see discussion of Experiment 1). Next we present a mixed effect analysis that addresses the potential spill-over confound and introduces additional control variables that have been previously shown to affect reading times.

Meta-Analysis of Experiments 1 and 2

We use mixed effect modeling to include additional control variables addressing the potential confounds raised in the discussion of Experiments 1 and 2 while combining the by-participants and the by-items analyses. For this analysis, the data from Experiments 1 and 2 were pooled together. This step was taken to maximize the power for our analysis given that the approach we take is extremely conservative, including several additional covariates that are usually not modeled in experiments of the type presented above (but perhaps should be, as we argue below). As mentioned above, the potential position and spill-over confounds we control for are, by no means, unique to the experiments reported here, but rather are typical for experiments investigating locality and anti-locality effects, and self-paced reading experiments, more generally.

Data

Word reading times for all 6,003 filler and all 2,844 critical item trials of Experiments 1 and 2 were pooled, yielding 185,763 word reading times. All trials with reading times that were likely to be due to accidental button presses, attention lapses, or machine errors (raw RTs < 100 or > 5000 ms, 0.2% of all word RTs) were excluded. Next, within-subject z-scores of reaction times were calculated for all remaining trials and trials more than 2.5 standard deviations from the subject's mean reading time were excluded (an additional 2.8% of all word RTs). All data from two subjects were excluded because reading times for fewer than 20 target trials were left after the initial exclusion criteria had been applied. Finally, one further subject was excluded because the average comprehension accuracy was lower than 70% (including filler items), resulting in a total exclusion of 5.3% of all word RTs (2.3% of all item trials). After all exclusions, the database for the meta-analysis contained data from 66 items and 85 participants.

Model

Our analysis consists of two steps. First we calculated residual log reading times using a linear mixed model fitted to all data (fillers and experimental trials) from Experiments 1 and 2. In the second step, we analyzed the effect of RC length on the residual log reading times (created in step 1) for the critical matrix verb region, while simultaneously controlling for RC type and spill-over effects.

In step 1, a linear mixed model (Baayen, Davidson, & Bates, 2008; Pinheiro & Bates, 2004) is used to regress logarithm transformed reading times against the following predictors (all of which were centered before including them in the analysis):

- **Word length (in letters):** to model that longer words are usually read more slowly. We also tested models with log transformed word length as a predictor, but those models were not significantly better than the simple model we present here.
- **Logarithm transformed position of the stimulus in the list:** to model speed-up during the course of the experiment.
- **Restricted cubic spline of the position of the word in the stimulus:** to model the increase and decrease in reading times depending on the word's distance from the stimulus onset. Restricted cubic splines (Harrell, 2001) provide a convenient way to test whether an effect contains a non-linear component.
- **Stimulus type:** to model overall complexity differences among the different filler types and the experimental items.
- **Individual differences among readers:** a random intercept parameter was included in the model to account for differences in participants' reading rates.

All of the above factors contribute significantly to word reading times. For example, word reading times increase about 1.3% for each additional letter ($\beta = 0.013$, $p < 0.0001$). Readers also speed up throughout the experiment with increasing log-transformed stimulus position within the list ($\beta = -0.082$, $p < 0.0001$), so that a word in the 50^{th} stimulus is read in less than 73% of the time it takes to read a comparable word in the first stimulus (additional test with restricted cubic splines over stimulus position confirmed that the log-transform adequately models the effect of stimulus position). Finally, the word position within a stimulus had a highly non-linear, but relatively weak effect on reading times, which is illustrated in Figure 6. A separate analysis found that word position effects differed for different stimulus types,

suggesting that word position effects on reading times depend on clausal or other phrasal boundaries rather than distance from the stimulus onset (cf. Just & Carpenter, 1980 for evidence that readers read clause by clause). A full investigation of this possibility is beyond the scope of this article. Similarly, we did not have enough data to investigate potential interactions between word length, stimulus position, and item type without running into issues of overfitting. We note that the analysis contains more control parameters than any previous experiment on locality or anti-locality effects that we are aware of.

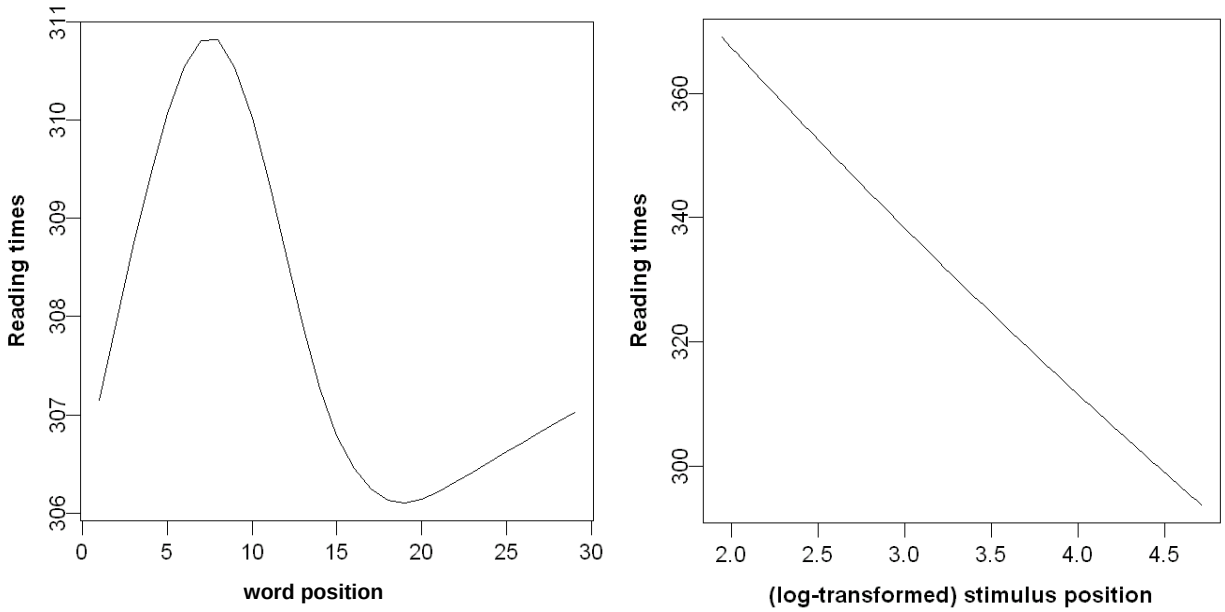

Figure 6 Effect of a word's position within a stimulus (left) and stimulus position within the experimental list (right) on words' log-transformed reading times.

The estimated residual log-transformed reading times on the matrix verb (the residuals returned by the first model, i.e. the fraction of the log reading times that is not accounted for by the above variables) constitute the input to the second step in the analysis. We use a linear mixed model to regress the residual log-transformed reading times on the target region (the matrix verb)

against RC length and RC type, four spillover covariates, as well as comprehension accuracy, and the interaction between comprehension accuracy with RC length and RC type.

The three levels of RC length were coded using Helmert contrast, comparing RTs for 2PPs against 3PPs and the mean RTs of those two conditions against RTs for 1PP. The three levels of RC type were treatment-coded, with one predictor indicating whether the RC was object-extracted (vs. subject-extracted), and one predictor indicating whether, for object-extracted RCs, the presence of *that* affected RTs. (The specific coding does not affect the significance of any of the manipulations.)

The four spillover covariates are the centered residual log-transformed reading times from the four words immediately preceding the matrix verb (i.e. the final PP in the RC and the word preceding it, which is either the final word of the penultimate PP or the RC verb, depending on the RC length condition). Spillover to the matrix verb should be higher, the higher the processing times on the preceding words. So, if processing from the preceding four words spills over onto the matrix verb region, we should see positive correlations between the spillover variables with the residual log-transformed reading times on the matrix verb. Furthermore, if spillover is the sole reason for the effects of RC length observed in Experiment 1 and 2, the four spillover variables should entirely subsume the effect of RC length. Thus, the inclusion of the spillover variables addresses the potential confound raised above: it is possible that having more PPs in the RC simply gives comprehenders more time to finish the processing of the dependencies in the RC (between the RC verb and the RC subject and object) and this may cause the effect of RC length observed in Experiment 1 and 2.

Since the four spillover predictors unsurprisingly showed signs of collinearity (a word is likely to be read slower if the preceding word is read slowly), residuals of the spillover times

were used. The centered spillover RT for the second closest preceding word was regressed against the centered spillover RT of the closest preceding word; the centered spillover RT of the third closest preceding word was regressed against the spillover from the closest and second closest word; etc. The centered spillover RT from the closest and the residual spillover RTs from the second to fourth closest preceding word were entered into the model. After this step was taken, the model contained no signs of collinearity (all fixed effect correlations < 0.2; all fixed effect correlations with RC length < 0.1). Removal of collinearity is crucial for two reasons: (1) collinearity would inflate the estimates of coefficients' standard errors leading to unreliable results, and (2) collinearity would lead to uninterpretable coefficient values.

Finally, comprehension accuracy was treatment-coded and centered (1 for correct answers and 0 for incorrect answers). The centering removed collinearity between the main effect of comprehension accuracy and its interactions with RC length and RC type.

In addition to the fixed effect predictors, the linear mixed model also contained random intercepts for participant and item effects (a random effect for Experiment 1 vs. 2 was originally added, but did not account for any variance and thus was removed). The random effects serve the same purpose that F1 and F2 ANOVAs are meant for – to account for the fact that all data points from the same subject or the same item might have shared properties influencing the dependent variable (to avoid violations of the assumptions of independence, Baayen et al., 2008).

The mixed model was fit using in the statistical software R (R Development Core Team, 2008) using the function *lmer* from the package *lme4* (version 0.99875-9, Bates, 2007; for an introduction, see Baayen et al., 2008). To avoid anti-conservative p-values, the reported p-values are based on MCMC sampling (with B=25,000 iterations; see Baayen, 2007). The sampled

parameters are clearly normally distributed, and model validation indicates that the model is not overfitted to the data or that further outlier analysis is necessary.

Results

Table 3 contains the summary of the results. Below we give the coefficient and the MCMC-based p-value estimate for significant effects. Positive coefficients indicate significantly slower residual log RTs, negative coefficients indicate significantly faster residual log RTs. For non-significant effects, we only give the p-value.

ORCs lead to marginally slower reading times on the matrix verb ($\beta = 0.019$, $p < 0.054$). This effect is at least partly due to a trend for an interaction with comprehension accuracy ($\beta = -0.029$, $p < 0.16$). A simple effect post-hoc analysis confirmed that for correctly answered trials alone, the slowdown for ORCs vanishes, while it persists for incorrectly answered trials alone. The presence of *that* had no effect ($p > 0.3$). There was no main effect of comprehension accuracy and all other interactions were non-significant (all $ps > 0.6$).

All four spillover predictors had independent significant effects in the expected direction. The target region was read more slowly, the higher the residual log RT on the closest preceding word ($\beta = 0.311$, $p < 0.0001$), the second closest word ($\beta = 0.165$, $p < 0.0001$), the third closest word ($\beta = 0.108$, $p < 0.0001$), and the fourth closest word ($\beta = 0.052$, $p < 0.001$). As can be seen from the coefficients, the spillover effects decrease with increasing distance between the spillover source and the target region.

Crucially, RC length had a significant effect beyond the spillover effects. The matrix verb was read faster when the RC contained two or three PPs compared to one PP ($\beta = -0.011$, $p < 0.05$). There was no significant difference between two compared to three PPs ($p > 0.3$).

	Coefficient est.		95% HPD		P	
	Original	**MCMC**	**lower**	**upper**	***t*-based**	**MCMC**
Main effects						
Correct answer	-0.0021	-0.0026	-0.0289	0.0223	0.8723	0.8434
Object relative	0.0193	0.0195	-0.0012	0.0389	0.0593	0.0536
Object relative w/ *that*	-0.0056	-0.0058	-0.0166	0.0054	0.3198	0.3010
1 PP vs. 2 & 3 PPs	-0.011	-0.0112	-0.022	-0.0005	0.0490	0.0408
2 PPs vs. 3 PPs.	0.0059	0.0058	-0.0067	0.0177	0.3502	0.3544
Spillover (closest)	0.3114	0.3039	0.2661	0.3413	0.0001	0.0001
Spillover (2^{nd} closest)	0.1649	0.1594	0.1187	0.1984	0.0001	0.0001
Spillover (3^{rd} closest)	0.1079	0.1043	0.0662	0.1431	0.0001	0.0001
Spillover (4^{th} closest)	0.0517	0.0483	0.0119	0.0833	0.0046	0.005
Interactions w/ Correct answer						
Object relative	-0.0293	-0.0295	-0.0707	0.0116	0.1622	0.1600
Object relative w/ *that*	0.0037	0.0039	-0.0213	0.0290	0.7708	0.7606
1 PP vs. 2 & 3 PPs	-0.006	-0.0064	-0.0301	0.0185	0.6269	0.6142
2 PPs vs. 3 PPs.	0.0032	0.0029	-0.0253	0.0305	0.8232	0.8308

Table 3 Summary of linear mixed model results for pooled analysis of Experiments 1 and 2. The first two columns give the coefficient estimates based on the original model fit, and the MCMC-sampling, respectively. The third and fourth column give the 95% Highest Posterior Density interval derived by the MCMC-sampling. The penultimate column reports the anti-conservative p-values based on the t-test of the original standard error estimates for the coefficients. The final column reports the corrected p-value based on MCMC-sampling. For our data, the two p-values do not differ much.

General Discussion

The results of the mixed effect analysis confirm the results found in the ANOVAs for Experiments 1 and 2. RC length has the effect predicted by expectation-based theories (Hale, 2001; Konieczny, 2000; Konieczny & Döring, 2003; Levy, 2008; Narayanan & Jurafsky, 2002): the matrix verb is read faster the more expected it is given the preceding RC (cf. Figure 2). This effect is observed after the position of the matrix verb in the stimulus, stimulus position in the list, stimulus type, spill-over from the preceding four words, and individual differences among participants are accounted for. Interestingly, only the difference between 1 PP vs. 2 PPs in the RC has a significant effect on the reading times on the matrix verb. The difference between 2 PPs and 3 PPs is not significant. This is consistent with the fact that the difference in the probability of the matrix verb between the 1 PP and the 2 PPs condition is much bigger than the difference in probability between the 2 PPs and 3 PPs condition (cf. Figure 2).

To the best of our knowledge, Experiments 1 and 2 provide the first evidence of anti-locality effects (faster reading times for more intervening material) outside of head-final languages (such as German and Hindi). This suggests that anti-locality effects are not the result of different processing strategies for head-final languages. This may seem like a trivial observation, but as a matter of fact previous studies on English stimuli similar to ours have found locality rather than anti-locality effects (Bartek et al., in prep; Grodner & Gibson, 2005). For example, Grodner and Gibson (2005, Experiment 1) compared comprehension times on a matrix verb that is either adjacent to its subject, or is separated from its subject by an intervening preposition phrase or relative clause:

(12) The nurse Ø/[from the clinic]/[who was from the clinic] <u>supervised</u> the administrator ...

As predicted by memory-based accounts, and in apparent conflict with out results, Grodner and Gibson (2005) found slower processing of the matrix verb for more intervening material. This was replicated by Bartek et al. (in prep), raising the question why our results differ from previous results. There are two major differences between our Experiments and previous work. First, unlike previous work, the mixed effect analysis of Experiment 1 and 2 contains explicit modeling of spill-over effects. Given the lack of such controls, it is possible that the locality effects found on the matrix verb in previous studies are due to spill-over (for the same point, see Vasishth & Lewis, 2006:781). Second, our locality manipulation (RC length) differs from those in previous work. Both Grodner and Gibson (2005) and Bartek and colleagues (Bartek et al., in prep) manipulate locality by either having no intervener between the matrix subject and verb, one PP, or a relative clause. They also manipulate locality by having the same three conditions in the intervening relative clause (center-embedding). In our experiments, either 1, 2, or 3 PPs intervene between the matrix subject and verb. The type of PPs in our experiments is not very different from those used in the experiments by Grodner & Gibson's and Bartek and colleagues (all experiments use mostly locative PPs). But it is possible that comprehenders process multiple PPs differently than they do with a single PP. We plan future studies to test whether the anti-locality effect observed in Experiment 1 and 2 extends to RC length manipulations that do not employ chains of PPs.

We close by discussing a potential extension of the model introduced in Lewis and Vasishth (2005). While the current version of the model outlined in Lewis and Vasishth (2005), here dubbed LV05, cannot account for the results of Experiment 1 and 2, the architecture of that model is very much compatible with expectation-based processing. Here we focus on LV05 because of the precise specification of the retrieval mechanism and its success is accounting for

previous locality and anti-locality results, but other memory-based accounts (e.g. Gibson, 1998, 2000) could be extended in a similar fashion. We first describe the LV05 model in more detail and then turn to possible extensions of it that incorporate expectation-based processing.

LV05 is a memory-based model that links processing (reading) times of words to memory retrieval. Processing of a word involves retrieval of a word into a lexical buffer (lexical access) and retrieval of a constituent from memory into the so-called retrieval buffer, where the word is attached to the constituent. There also is a control buffer setting the current goal for the parser (a type of syntactic expectation). Active processing is limited to the contents of these buffers. In LV05 lexical access and setting a syntactic goal are assumed to always take 50msecs. Only retrieval into the retrieval buffer is assumed to vary depending on the activation of the to-be-retrieved chunk. The base activation of a chunk is a function of all previous retrievals and the time that has passed since then (Equation 1, Lewis & Vasishth, 2005:381; *d* is an ACT-R constant determining the decay rate; during a retrieval call, chunks receive additional activation from the cues used for retrieval; here this is not important):

(13) $B_i = \ln \sum_{j=1}^{n} t_j^{-d}$

In addition to the base activation, a chunks activation level is also determined by the cues that are used during a retrieval call (Equation 2, Lewis & Vasishth, 2005: 381, where the strength of a particular cue S_{ik} depends on how many other chunks in memory match that cue, too):

(14) $A_i = B_i + \sum_k W_k S_{ik}$

We see three ways to integrate probability-sensitive effects into LV05 without changing its general architecture: as part of lexical access, as part of the base activation of chunks in

working memory, and/or as part of the formulation of syntactic goals (the goal buffer). We discuss these three possibilities in turn.

There is much evidence that comprehenders use linguistic and non-linguistic context to predict upcoming words (e.g. visual cues; pragmatics, plausibility, context, Garnsey, Perlmutter, Meyers, & Lotocky, 1997; Ni, 1996; Spivey-Knowlton, Trueswell, & Tanenhaus, 1993; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002). For example, more frequent words are processed faster (e.g. Just et al., 1982), and so are words that are more predictable given immediately preceding words (Demberg & Keller, to appear; McDonald & Shillcock, 2003). Lewis and Vasishth (2005:389) acknowledge that LV05 does not account for this fact, but also note that the architectural assumptions of ACT-R provide a straightforward way to account for frequency and contextual cues to lexical access. A lexical retrieval buffer that (unlike in LV05) is subject to the same retrieval mechanism as the retrieval buffer would be subject to activation-decay and cue-based retrieval. This would predict recency effects (due to recent reactivation of a lexical item) and effects of neighborhood density (due to interference during cue-based retrieval). Frequency effects may fall out of recency effects (frequent words being more likely to have been mentioned more recently) or there may be a direct effect of frequency on base activation. Contextual cues would also follow from the cue-based retrieval assumption, although –unlike for the retrieval buffer in LV05- cues would include non-linguistic cues in addition to bottom-up linguistic (acoustic) cues. Such a context-sensitive lexical retrieval process may be sufficient to explain the effects observed in Experiment 1 and 2. When the probability that the next word will be the matrix verb increases after each additional PP in the RC, so does the probability of any specific matrix verb. In other words, even if (as we have tried to ensure) the PPs do not bias towards a specific matrix verb, they increase the probability of any matrix verb. While it may seem

unlikely that the tiny shift in probability mass distributed over all possible matrix verb may cause the observed effect, it is important to keep in mind that it should be the *relative* change in probability that matters. If the PPs do indeed not bias towards any specific verb, the relative change in the probability of each possible matrix verb is the same as the overall change in probability for any matrix verb. This may be enough to cause the observed effect.

Alternatively or in addition, it is possible that contextual cues affect the activation (and hence retrieval times) of chunks in memory thereby affecting retrieval into the standard retrieval buffer of LV05. In other words, the base activation of chunks may be affected directly by their predictability in context. Regardless of whether retrieval from the retrieval buffer, lexical access, or both mechanisms are sensitive to contextual cues, a model that incorporates them can be seen as incorporating one of the central ideas of activation-based models (such as any type of connectionist model, e.g. Christiansen & Chater, 1999; MacDonald, Pearlmutter, & Seidenberg, 1994; Rohde, 2002; Spivey & Tanenhaus, 1998; Tabor, Juliano, & Tanenhaus, 1997) – that the time needed to process an item depends on its activation level and that the activation level of an item depends on its base activation and the activation received through other connected items.

It is also possible that merely *predicting* a constituent (i.e. setting the goal buffer to contain that constituent) affects later retrieval of that component. In LV05 setting a new goal state does not per se facilitate later retrieval of that chunk. This could be changed, allowing repeated prediction of a node to lead to higher activation of that node and hence facilitation of later retrieval of that node. Unlike the first two possible extensions of the LV05 model, this approach does not include any direct reference to contextual cues and is as such much more constrained.

The three possible extensions of the LV05 model are not mutually exclusive and it could be that contextual probabilities affect sentence comprehension at several levels of processing (e.g. lexical access and retrieval of syntactic chunks). Further work is necessary to compare implementations of the ideas we have outlined against human behavior.

Conclusions

We observe anti-locality effects in English. These effects constitute evidence for expectation-based syntactic comprehension (Hale, 2001; Konieczny, 2000; Konieczny & Döring, 2003; Levy, 2008; Narayanan & Jurafsky, 2002): At least in the context we investigated, comprehenders seem to build expectations not just about a specific upcoming word, but also upcoming syntactic material. The results are problematic for memory-based accounts of sentence comprehension (Gibson, 1998, 2000; Lewis & Vasishth, 2005; Lewis et al., 2006; Van Dyke & Lewis, 2003; 2003). Unlike anti-locality effects previously observed for head-final languages (Konieczny, 2000; Konieczny & Döring, 2003; Vasishth, 2003; Vasishth & Lewis, 2006), the anti-locality effects observed in our experiments are also unexpected given the memory-based model presented in Lewis and Vasishth (2005). Since there is, on the other hand, strong evidence for locality effects and since most of these locality effects are not predicted by existing expectation-based accounts, we have argued that sentence comprehension must involve both a memory component and an expectation component.

Acknowledgments

We are grateful for countless discussions with Shravan Vasishth, Roger Levy, and Julie Van Dyke, and for feedback from the 2005 and 2008 CUNY audiences, where parts of this work were presented.

Appendices

Stimuli of Experiment 1 (in the SRC with 1 PP condition; for an example item in all 9 conditions, see (5)-(7))

1. The janitor that saw the plumber near the stadium lost the key on the street outside the bar but he later found it.
2. The pharmacist that helped the assistant with the patient placed the order with the prescription after the conversation but it has not yet arrived.
3. The waitress that hugged the bartender on Saturday night dropped the tray without any hesitation during the dinner but nobody saw any of that.
4. The client that contacted the retailer about hazardous chemicals offered a deal for more information more than once but the timing was wrong.
5. The celebrity that admired the athlete during the games won the award at the ceremony for fair play but many thought he didn't deserve it.
6. The detective that recognized the spy around seven o'clock attended a meeting in Eastern Europe early yesterday night so he missed an important phone call.
7. The journalist that complimented the editor in the morning revised the article on the story after the meeting but he was too tired to give it a final read.
8. The employee that praised the executive for a customer finished the project right on time for the analysis but her effort was not recognized by the manager.
9. The legislator that visited the senator in his home state falsified the documents despite unforeseen problems during the trip but the fraud was discovered last week.
10. The soldier that wounded the rebel on European grounds received a medal for no reason after the war but never was able to forget.
11. The officer that described the criminal at the city hall told a lie on multiple occasions during press conferences and eventually he lost his job.
12. The hairdresser that hired the beautician in New York transformed the salon for the fair late in July but he soon regretted it.
13. The reporter that followed the cameraman through Saudi Arabia damaged the equipment several months ago during the journey but he still was able to get some footage.

14. The understudy that telephoned the agent an hour ago shared the story about the job in Los Angeles and felt relieved immediately.
15. The consultant that confronted the programmer late yesterday night broke the computer in a rage at two o'clock and sent the parts back to the manufacturer.
16. The supervisor that deceived the owner before her confession kept the money without any regret for several years and nobody ever suspected her.
17. The entrepreneur that congratulated the stockbroker three years ago sold the shares after the merger in the spring and was looking forward to a long holiday.
18. The traitor that revealed the defector in New Mexico rejected the offer on the spot after recent events and contacted his lawyer immediately.
19. The singer that blamed the producer at the last moment cancelled the recording for the movie in Las Vegas and so, his reputation was damaged.
20. The acrobat that mocked the clown at the circus performed the trick at a show in front of the children and the audience was amazed.
21. The customer that upset the seller not just once forgot the book with bad jokes at the convention and so, he called the lost-and-found office.
22. The partner that introduced the businessman not without pride presented the report at the gathering to her colleagues and hoped for a promotion.
23. The lecturer that provoked the dean during a party left the university without any fear in early May but he returned the following year.
24. The messenger that summoned the knight before the ball read the letter to the King during the reception but he left out the last sentence.
25. The linguist that criticized the historian at the conference proposed a modification for the argument for the theory but his collaborator did not approve.
26. The biker that passed the driver in San Francisco made the turn at the crossing near Berlin Avenue and nearly hit the street light.
27. The author that met the president some time ago wrote the book before the elections in the US and it became a bestseller right away.

28. The guest that kissed the host last Thursday night stole the vase in the corridor during the party because he has always suffered from kleptomania.
29. The clerk that saw the boss in Palo Alto wanted the job at the counter in the Marriot and so, he filled out the application.
30. The artist that robbed the traveler during the night owned a horse in the forest by the lake and so reached the mountains in no time.
31. The trumpeter that loved the drummer before his marriage formed the band during high school fifteen years ago but unfortunately he was never successful.
32. The intern that distrusted the boss for many years disregarded the messages because of rumors about the party but she never admitted it.
33. The roommate that annoyed the landlord at Magazine Street slammed the door during a sighting of the apartment but he refused to pay for the damage.
34. The player that met the coach at eight o'clock bought the house near the gym by the river but his wife didn't like it very much.
35. The mayor that called the advisor after ravaging reviews requested an update for more comments about the project but he never received it.
36. The librarian that angered the teacher with rude remarks misplaced the dictionary again and again during the summer but she kept insisting it was checked out.

Stimuli of Experiment 2 in the ORC with *that* and 3 PP condition (for an example item in all 6 conditions, see (10); the items are based on Experiment 1, but slight changes had to be made to make the PPs compatible with topicalization)

1. The janitor that the plumber saw on the street outside the bar near the stadium misplaced the key but he later found it.
2. The hairdresser that the beautician hired for the fair in New York in late July remodeled the salon but he soon regretted it.
3. The lecturer that the dean provoked after the conference in early May during a party left the university but he returned the following year.

4. The trumpeter that the drummer dated in high school fifteen years ago before the accident formed the band but unfortunately he was never successful.
5. The tenant that the landlord annoyed during a sighting in the apartment on Magazine Street slammed the door but he refused to pay for the damage.
6. The player that the coach met near the gym by the river at eight o'clock saw the suspect but he couldn't remember any details.
7. The mayor that the advisor called in early November before the election after the poll_results updated the report but it soon was out of date again.
8. The librarian that the administrator interviewed during an exhibition in a museum in downtown Manhattan damaged a painting and she had to pay for it.
9. The pharmacist that the assistant helped in the afternoon after the conversation with the patient forgot the order and so it has not yet arrived.
10. The waitress that the bartender hugged in the kitchen during the dinner on Saturday night dropped the tray but nobody saw any of that.
11. The client that the retailer contacted in the spring before the takeover about hazardous chemicals broke the contract and this lead to a strike.
12. The detective that the spy recognized around seven o'clock early yesterday night in Eastern Europe attended a meeting so he missed an important phone call.
13. The journalist that the editor complimented in the hallway after the meeting in the morning smoked a cigarette while his assistant edited an article.
14. The legislator that the senator visited despite unforeseen problems during the trip in his home_state falsified the documents but the fraud was discovered last week.
15. The soldier that the rebel wounded during the occupation after the war for no reason got a medal but never was able to forget.
16. The officer that the judge interrogated on multiple occasions during court sessions at the city_hall told a lie but nobody ever noticed it.
17. The reporter that the cameraman followed several months ago during the journey through Saudi Arabia broke the camera but he still was able to get some footage.

18. The understudy that the agent telephoned in the office in Los Angeles an hour ago shared the story and felt relieved immediately.
19. The consultant that the programmer confronted in a rage after the discussion late yesterday night broke the computer and sent the parts back to the manufacturer.
20. The entrepreneur that the stockbroker congratulated after the merger in South Korea five days ago sold the shares and was looking forward to a long holiday.
21. The singer that the producer hired at the show in Las Vegas some weeks ago cancelled the recording "and so, his reputation was damaged."
22. The acrobat that the clown mocked at a show in the circus during the break performed the trick and the audience was amazed.
23. The partner that the businessman introduced at the gathering in the conference-room not without pride presented the report and hoped for a promotion.
24. The messenger that the knight received during the reception before the ball at the festival read the letter but he omitted important details.
25. The linguist that the historian criticized at a lunch-meeting before the talk at the conference insulted the organizer but he apologized immediately.
26. The biker that the driver passed at the crossing near Berlin Avenue during rush hour greeted a passer-by and nearly ran into a tree.
27. The author that the president honored before the elections in the US some time ago wrote the book and it became a bestseller right away.
28. The guest that the host kissed in the corridor during the party last Thursday night stole the vase because he has always suffered from kleptomania.
29. The clerk that the boss found at the counter in the Marriot in Palo Alto counted the money and afterwards he went out for a cigarette.
30. The artist that the traveler robbed in the forest by the lake during the night accused the locals and that didn't buy him much sympathy.

References

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition, 73*(3), 247-264.

Baayen, R. H. (2007). *Analyzing Linguistic Data: A practical introduction to Statistics*. Cambridge: Cambridge University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items.

Bartek, B., Smith, M., Lewis, R., & Vasishth, S. (in prep). Linking Working Memory Processes with Self-paced Reading and Eye-tracking Measures: How Lexical Processing Load and SPR Exaggerate effects.

Bates, D. (2007). lme4: Linear mixed-effects models using S4 classes. *Journal*,

Christiansen, M. H., & Chater, N. (1999). Toward a Connectionist Model of Recursion in Human Linguistic Performance. *Cognitive Science, 23*, 157-205.

Christianson, K. T. (2002). *Sentence processing in a 'nonconfigurational' language.* Michigan State University, East Lansing.

Demberg, V., & Keller, F. (to appear). Data from Eye-tracking Corpora as Evidence for Theories of Syntactic Processing Complexity. *Cognition*.

Ferreira, F., & Henderson. (1993). eading Processes During Syntactic Analysis and Reanalysis. *Canadian Journal of Experimental Psychology, 47*, 247-275.

Francis, W., & Kucera, H. (1961). The Brown corpus of standard American English.

Garnsey, S. M., Perlmutter, N. J., Meyers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language, 37*, 58–93.

Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition, 68*(1), 1-76.

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In Y. Miyashita, P. Marantx & W. O'Neil (Eds.), *Image, language, brain* (pp. 95-112). Cambridge, MA: MIT Press.

Gibson, E., & Wu. (submitted).

Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory Interference During Language Processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(6), 1411-1423.

Gordon, P. C., Hendrick, R., & Johnson, M. (2004). Effects of noun phrase type on sentence complexity. *Journal of Memory and Language, 51*(1), 97-114.

Grodner, D. J., & Gibson, E. (2005). Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive Science, 29*, 261-290.

Hakes, D. T., & Cairns, H. S. (1970). Sentence comprehension and relative pronouns. *Perception & Psychophysics, 8*, 5-8.

Hakes, D. T., Evans, J. S., & Brannon, L. L. (1976). Understanding sentences with relative clauses. *Memory & Cognition, 4*(3), 283-290.

Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. In *North American Chapter Of The Association For Computational Linguistics (NAACL)* (pp. 1-8). Morristown, NJ: Association for Computational Linguistics.

Harrell, F. E. (2001). *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*: Springer.

Hsiao, F., & Gibson, E. (2003). Processing relative clauses in Chinese. *Cognition, 90*(1), 3-27.

Jaeger, T. F. (2007). *Rational Speakers: speakers help processing when it is most necessary.* Paper presented at the 13th AMLaP, Turku, Finland.

Just, M. A., & Carpenter, P. A. (1980). A Theory of Reading: From Eye Fixations to Comprehension. *Psychological Review, 4*(329-354).

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General, 111*(2), 228-238.

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49*(1), 133-156.

Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research, 29*(6), 627-645.

Konieczny, L., & Döring, P. (2003). Anticipation of clause-final heads. Evidence from eye-tracking and SRNs. In P. P. Slezak (Ed.), *Proceedings of the 4th International Conference on Cognitive Science, July 13-17, 2003, University of New South Wales, Sydney, Australia* (pp. 330-335).

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition, 106*(3), 1126-1177.

Levy, R., Fedorenko, E., & Gibson, E. (2007). *The syntactic complexity of Russian relative clauses*. Paper presented at the 20th CUNY Conference on Human Sentence Processing.

Lewis, R. L. (1998). Reanalysis and limited repair parsing: Leaping off the garden path. In J. D. Fodor & F. Ferreira (Eds.), *Reanalysis in Sentence Processing* (pp. 247-185). Dordrecht: Kluwer Academic Publishers.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science, 29*(3), 375-419.

Lewis, R. L., Vasishth, S., & Van Dyke, J. A. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Science, 10*(10).

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). Lexical Nature of Syntactic Ambiguity Resolution. *Psychological Review, 191*, 676-703.

McDonald, S. A., & Shillcock, R. C. (2003). Low-level predictive inference in reading: the influence of transitional probabilities on eye movements. *Vision Research, 43*(16), 1735-1751.

Nakatani, K., & Gibson, E. (2008). Distinguishing theories of syntactic expectation cost in sentence comprehension: Evidence from Japanese. *Linguistics, 46*, 43-86.

Narayanan, S., & Jurafsky, D. (2002). A Bayesian Model Predicts Human Parse Preference and Reading Times in Sentence Processing. In *Advances in Neural Information Processing Systems (NIPS) 14: Proceedings of the 2002 Conference*.

Ni, W. (1996). Sidestepping Garden Paths: Assessing the Contributions of Syntax, Semantics and Plausibility in Resolving Ambiguities. *Language and Cognitive Processes, 11*(3), 283–334.

Pinheiro, J. C., & Bates, D. M. (2004). *Mixed Effects Models in S and S-Plus*. New York City, NY: Springer Verlag.

R Development Core Team. (2008). *R: A Language and Environment for Statistical Computing*.Unpublished manuscript, Vienna, Austria.

Race, D. S., & MacDonald, M. C. (2003). The use of "that" in the production and comprehension of object relative clauses. *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*.

Rohde, D. L. T. (2002). *A Connectionist Model of Sentence Comprehension and Production.* Carnegie Mellon University, Pittsburgh PA.

Spivey-Knowlton, M. J., Trueswell, J. C., & Tanenhaus, M. K. (1993). Context effects in syntactic ambiguity resolution: Discourse and semantic influences in parsing reduced relative clauses. *Canadian Journal of Experimental Psychology, 47*(2), 276–309.

Spivey, M. J., & Tanenhaus, M. K. (1998). Syntactic Ambiguity Resolution in Discourse: Modeling the Effects of Referential Context and Lexical Frequency. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 24*, 1521-1543.

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive psychology, 45*(4), 447–481.

Staub, A., & Clifton, C. J. (2006). Syntactic Prediction in Language Comprehension: Evidence From Either… or. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 32*(2), 435-436.

Tabor, W., Juliano, C., & Tanenhaus, M. K. (1997). Parsing in a dynamical system: An attractor-based account of the interaction of lexical and structural constraints in sentence processing. *Language and Cognitive Processes, 12*, 211-271.

Van Dyke, J. A., & 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*(3), 285-316.

Vasishth, S. (2003). *Working memory in sentence comprehension: Processing Hindi center embeddings.* . Routledge, NY: Garland Press.

Vasishth, S., & Lewis, R. (2006). Argument-head distance and processing complexity: Explaining both locality and anti-locality effects. *Language, 82*(4), 767-794.