Dialogue and Discourse 1(2) (2011) 59-82 doi: 10.5087/dad.2011.104

Locality in German

Shravan Vasishth VASISHTH@UNI-POTSDAM.DE
Department of Linguistics
University of Potsdam
Potsdam, 14476, Germany

Heiner Drenhaus DRENHAUS@COLI.UNI-SAARLAND.DE
Computational Linguistics and Phonetics
University of Saarland
Saarbruecken, 66123, Germany

Editor: David Schlangen and Hannes Rieser

Abstract

Three experiments (self-paced reading, eyetracking and an ERP study) show that in relative clauses, increasing the distance between the relativized noun and the relative-clause verb makes it more difficult to process the relative-clause verb (the so-called locality effect). This result is consistent with the predictions of several theories (Gibson, 2000; Lewis and Vasishth, 2005), and contradicts the recent claim (Levy, 2008) that in relative-clause structures increasing argument-verb distance makes processing easier at the verb. Levy's expectation-based account predicts that the expectation for a verb becomes sharper as distance is increased and therefore processing becomes easier at the verb. We argue that, in addition to expectation effects (which are seen in the eyetracking study in first-pass regression probability), processing load also increases with increasing distance. This contradicts Levy's claim that heightened expectation leads to lower processing cost. Dependency-resolution cost and expectation-based facilitation are jointly responsible for determining processing cost.

Keywords: Sentence processing; locality; surprisal; expectation-based sentence comprehension; eye-tracking; EEG; self-paced reading.

1. Introduction

A well-known claim in the psycholinguistic literature (Chomsky, 1965; Just and Carpenter, 1980, 1992; Gibson, 2000) states that completing a dependency between two linguistic units (such as a verb and an argument) is partly a function of the distance between them. An example is the self-paced reading study by Grodner and Gibson (2005), which showed increasing reading time at the verb *supervised* as a function of the distance between the subject *nurse* and the verb:

(1) a. The nurse supervised the administrator while…
b. The nurse from the clinic supervised the administrator while…
c. The nurse who was from the clinic supervised the administrator while…

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) directly address dependency resolution in sentence

 Submitted 1/2010; Accepted 12/2010; Published online 5/2011

comprehension in terms of memory retrieval (similar early approaches are the production-system based models of Anderson et al. 1977). Just and Carpenter developed a model of integration that involved activation decay (as 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 (Just and Carpenter, 1992, 133):

> 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.

The explanation for the distance effect in terms of activation decay was taken a great deal further in the Syntactic Prediction Locality Theory or SPLT (see Gibson (1998, 9) for a historical overview of the connection between decay and distance) and, more recently, the Dependency Locality Theory or DLT Gibson (2000). The DLT proposes (among other things) that the cognitive cost of assembling a dependent with a head is partly a function of the number of new intervening discourse referents that are introduced between the dependent and the head; see Figure 1 for an example. In effect, the DLT discretizes the concept of activation decay in the DLT complexity metric (Gibson, 2000, 103). The predictions of SPLT and DLT find quite good empirical support from online experiments involving English (e.g., Gibson and Thomas 1999, Grodner and Gibson 2005, Warren and Gibson 2005) and also Chinese (Hsiao and Gibson, 2003). At least one offline study involving Japanese is also consistent with the SPLT's (the precursor of DLT) predictions (Babyonyshev and Gibson, 1999).

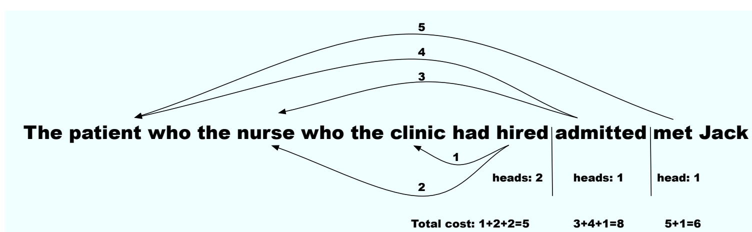

Figure 1: A schematic illustration of DLT's predictions for multiply embedded structures. Integration costs are labeled along the arcs that define the argument-head dependencies, computed by counting the number of intervening discourse referents. Another component of the theory is storage cost; these costs are presented under each verb for illustration. The storage costs is computed by counting the number of heads predicted at each point.

As mentioned above, locality cost is characterized by the DLT in terms of the number of discourse referents intervening between the dependent and the head. One may ask: what is so special about the number of new discourse referents? Why not count the number of intervening syntactic nodes, words, letters, syllables, etc.? The rationale within the DLT is that building discourse referents is computationally costly; independent evidence for this idea comes from studies showing that the accessibility of the intervening discourse referent (as defined by the accessibility hierarchy) can modulate retrieval difficulty (Warren, 2001; Warren and Gibson, 2005).

One interesting empirical problem is that, apart from the Grodner and Gibson (2005) results, locality does not seem to have much empirical support (indeed, Jaeger et al. (2008) have recently presented evidence that the locality constraint may not apply even in English, the language that has the most-attested instances of locality effects). Konieczny (2000) presented an important counterexample from German to the locality hypothesis. In a self-paced reading study involving center-embedded relative clauses, he showed that increasing argument-head distance, analogous to example 1 above, resulted in faster reading time at the verb, not slower (as predicted by locality-based accounts). Konieczny's explanation for the result was that the strength of prediction for the upcoming verb increases if more intervening material is present between the dependent and the head (he calls this the anticipation hypothesis).

Recently, Levy (2008) has proposed a related explanation for such antilocality effects. Under this view, antilocality effects could be explained by assuming that the material intervening between the dependent and head could serve to sharpen the expectation for the upcoming verb. As Levy (2008, 1144) puts it:

> "more preverbal dependents gives [sic] the comprehender more information with which to predict the final verbs identity and location, and comprehension should therefore be easier."

This expectation hypothesis is the suggested explanation for antilocality effects seen in German (Konieczny, 2000) and Hindi (Vasishth and Lewis, 2006).

An interesting prediction of Levy's expectation-based account is that English should also show antilocality effects. Indeed, Levy quotes a self-paced reading study by Jaeger and colleagues (also see the further studies in Jaeger et al. 2008) which confirmed this prediction. Jaeger et al. (2008) presented participants with sentences like 2. They found that reading time at the verb *bought* is faster as distance between the dependent *player* and the verb is increased. Jaeger and colleagues argue that this speedup occurs because "...end of [the relative clause (RC)] - and hence presence of matrix verb becomes more probable after each additional PP in RC." This is another example of how sharpened expectation of an upcoming verb results in faster reading time at the verb.

(2) a. The **player** [RC that the coach met at 8 o'clock] **bought** the house ...
    b. The **player** [RC that the coach met by the river at 8 o'clock] **bought** the house ...
    c. The **player** [RC that the coach met near the gym by the river at 8 o'clock] **bought** the house ...

Given the recent findings from English of Jaeger and colleagues, one might wonder whether locality can be dismissed altogether as a possible contraint on processing difficulty. However, there are at least two problems with dismissing the locality effect. First, Van Dyke and Lewis (2003) present indirect evidence for locality. They conducted a self-paced study involving sentences such as 3. One factor was ambiguity (presence/absence of the sentential complement that), and another was distance between an argument (here, the noun *student*) and verb (*was standing*).

(3) a. The assistant forgot that the student was standing in the hallway.
    b. The assistant forgot the student was standing in the hallway.
    c. The assistant forgot that the student who knew that the exam was important was standing in the hallway.
    d. The assistant forgot the student who knew that the exam was important was standing in the hallway.

The ambiguity manipulation ensures that reanalysis takes place at *was standing* – the NP *student* must be reanalyzed as the subject of a sentential complement rather than the object of *forgot*. The distance manipulation ensures that the reattachment of the NP as subject of *was standing* is affected by locality. The reanalysis requires an integration between the verb and the argument, which is either near or distant from the verb. Consequently, if a significantly greater reanalysis cost is observed in the intervening-items conditions 3c,d than in the non-intervening-items conditions 3a,b, this would be a locality effect, and it would be independent of spillover confounds because the comparison is no longer a direct one between conditions with differing regions preceding the critical verb. The interaction was in fact observed in the Van Dyke and Lewis study (i.e., the difference between 3c,d and 3a,b was significant). This result suggests that distance (or decay) can adversely affect processing.

The second problem with doing away with locality explanations is the fact that the experiment by Grodner and Gibson (2005) discussed above clearly shows locality effects; Levy's expectation-based account cannot explain this effect. Here, it is legitimate to question the replicability of the Grodner and Gibson result; after all, if their finding cannot be reproduced, perhaps the data should be doubted, not the expectation-based account. However, Bartek et al. (2010) present four replications of the Grodner and Gibson result using self-paced reading and eyetracking; this leads us to question the expectation-based account.

To summarize the discussion so far, both locality and antilocality effects have been attested in English, but to our knowledge only antilocality effects have been seen in German and Hindi. The goal of the present study was to determine whether locality effects can be observed in German in a signature syntactic configuration where Levy's expectation-based processing account predicts antilocality effects. Specifically, we investigated structures like (4), where the expectation for a verb should necessarily get sharper as dependent-head distance increases.

(4) a. Die Mutter von Paula und die Schwester von Sophie gruessten den **Direktor**, den
The mother of Paula and the sister of Sophie greeted the director whom
Maria und Franziska **ignoriert** hatten.
Maria and Franziska ignored had

'The mother of Paula and the sister of Sophie greeted the director whom Maria and Franziska had ignored.'

b. Paula und die Schwester von Sophie gruessten den **Direktor**, den Maria und die
Paula and the sister of Sophie greeted the director whom Maria and the
Mutter von Franziska **ignoriert** hatten.
mother of Franziska ignored had

'Paula and the sister of Sophie greeted the director whom Maria and the mother of Franziska had ignored.'

c. Paula und Sophie gruessten den **Direktor**, den die Schwester von Maria und die
Paula and Sophie greeted the director whom the sister of Maria and the
Mutter von Franziska **ignoriert** hatten.
mother of Franziska ignored had

'Paula and the sister of Sophie greeted the director whom Maria and the mother of Franziska had ignored.'

Here, once a relative clause begins after the argument *Direktor*, the appearance of the verb in the relative-clause verb is guaranteed (as in the stimuli in the experiments by Konieczny (2000). The prediction of the expectation-based account for these sentences is exactly the same as for the studies by Konieczny's (2000) study and Jaeger et al's (2008) studies, where increasing the length of the relative clause renders processing easier at the verb: delaying the appearance of such a clause-final verb should sharpen the expectation for a verb. By contrast, the locality-based explanations predict a slower reading time and greater processing difficulty at the verb as distance is increased. We present three experiments that evaluated these predictions.

2. Experiment 1: self-paced reading

2.1 Participants

Fifty-one students at the University of Potsdam, all of them native speakers of German, took part in the study for course credit or payment.

2.2 Method

A self-paced reading comprehension experiment (Just et al., 1982) was carried out in German at the University of Potsdam, Germany. Thirty target sentences with three conditions each (see examples 4 above) were presented in a counterbalanced manner, with 104 filler sentences pseudo-randomly interspersed between the target sentences. The counterbalancing meant that each participant saw each sentence only once. The experiment was run on a Macintosh computer using the Linger software developed by Doug Rohde (http://tedlab.mit.edu/∼dr/Linger/). All items for this and other experiments reported in the paper are available from the first author.

Participants read the introduction to the experiment on the computer screen. In order to read each word of a sentence successively in 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; no feedback was given as to whether the response was correct or not. One concern was that participants might develop a question-answering strategy without paying attention to the entire sentence; therefore, questions were designed to probe different argument-verb relations in the sentences. All stimulus items and accompanying questions, along with the expected correct answers, are available from the authors.

2.3 Results

Figure 2 shows all the reading times at all positions; note that differences in reading time at regions preceding position 16 are not comparable because the words differ in each condition. We compared reading time at three regions of interest: position 16, the word preceding the critical word *ignoriert*, 'ignored'; position 17, the critical word; and position 18, the word following the critical word (see 4). We defined no specific prediction for the pre-critical region, but we included it in the analysis because this was the first region that was comparable across the sentences. For the other two regions we predicted increased processing difficulty with increasing distance (position 18 was expected to show spillover effects possibly originating at the critical region, see Mitchell (1984)). The mean RTs and standard errors for the three regions analyzed are shown in Table 1.

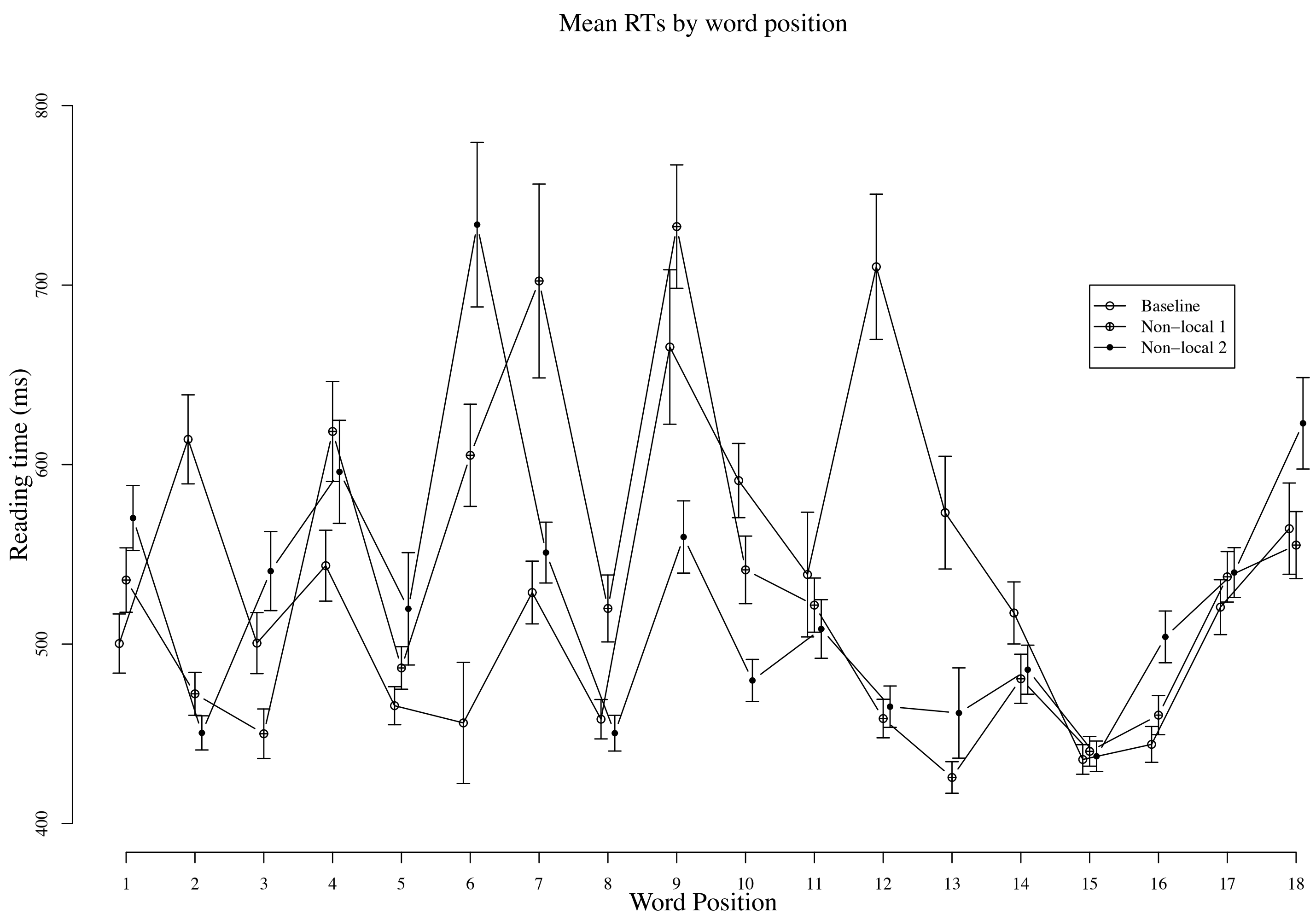

Figure 2: Mean reading time (ms) by word position in experiment 1 (self-paced reading), with standard error bars. This figure shows means and standard errors for the unreduced dataset; the data analysis was carried out after removing extreme values; see main text for details. The means for the positions 16–18 are shown in Table 1.

A linear mixed model was fit with items and participants as random intercepts, and Helmert contrasts (Venables and Ripley, 1999) were defined to investigate the effect of the locality manipulation. Helmert contrast coding was defined so that the first contrast involved comparing condition 4a with the average of 4b and 4c, and the second contrast compared 4b and 4c. This contrast coding has the advantage that a single statistical model gives us complete information about the relationship between the three conditions' means (the alternative would have been to carry out three pairwise t-tests, which would reduce the chances of detecting a true effect; standard statistics textbooks discuss this point, but also see Vasishth and Broe (2011) for a detailed discussion). In the model fits shown below, the coefficients are the differences between the means compared.

In order to reduce non-normality in the residuals, reading times less than 80 ms and greater than 2000 ms were removed from the data (this resulted in the removal of 0.05% of the data).[1] The data analysis was carried out on log-transformed reading times for reasons discussed in Baayen and Milin (2010).

As summarized in Table 2, Helmert contrasts showed that the average of conditions (b) and (c) was slower than (a) in all three regions; and that condition (c) was slower than (b) in the pre-critical and post-critical regions.

Table 1: Means (ms) and standard errors for the three regions analyzed. These are means with extreme values removed; see text for details.

condition	pre-critical	critical	post-critical
a	444 (10)	514 (14)	526 (13)
b	462 (11)	538 (14)	539 (15)
c	497 (13)	542 (14)	597 (19)

Table 2: Results of linear mixed model fit for experiment 1. Items and participants were crossed random factors. The asterisk represents statistical significance at $\alpha = 0.05$.

region	comparison	coef.	se	t-value
pre-critical	a vs b,c	0.0702	0.0204	3.4 *
	b vs c	0.0628	0.0235	2.7 *
critical	a vc b,c	0.0526	0.0219	2.4 *
	b vs c	0.0110	0.0252	0.4
post-critical	a vs b,c	0.05162	0.0262	2.0 *
	b vs c	0.0816	0.0302	2.7 *

1. Retaining all data points in the data analysis resulted in no effect at the critical region (the verb) for the comparison a vs b,c (t=1.8), and no effect in the post-critical region for the comparison a vs b, c (t=1.5).

2.4 Discussion

We found increased reading times in the critical and post-critical regions as distance between the verb (*ignoriert*) and the argument (*Direktor*) is increased. Interestingly, we also see increased reading time as a function of distance in the region preceding the verb; we discuss one possible implication of this result below.

The data clearly find support for the idea that increasing head-dependent distance increases processing load at the head. Thus, the result is consistent with locality predictions of models like Gibson's DLT and Lewis and Vasishth's ACT-R cue-based retrieval account. It is problematic for the expectation-based processing proposal by Levy, that increasing head-dependent distance in relative clauses renders the head more predictable and therefore easier to process.

Regarding the locality pattern seen in the pre-critical region, one possibility (a speculation at this point) is that the verb phrase (which is predicted from the moment that the relative clause begins) is already retrieved when the noun preceding the verb is processed; this assumption is consistent with Levy's expectation-based account. Once retrieved, the integration phase (retrieving the subject *Direktor*) could already begin while the noun is being processed. Assuming, as the DLT and the cue-based retrieval theory do, that this retrieval is costlier in the non-local conditions, the locality effect could appear even before the verb is processed. Clearly, this proposal needs further investigation.

Focusing now on only the regions for which we had well-defined a priori predictions, although we found clear evidence for locality accounts, one question arises: can both locality accounts and the expectation-based account co-exist? After all, as discussed above, evidence exists for both claims, and Boston et al. (2010) have shown that both the cue-based retrieval account of Lewis and Vasishth (2005) and the expectation-based account can independently explain reading data. We return to this question later in the paper.

Our next step is to attempt to replicate the above result using a different method. Replication is vital in order to verify the robustness of the finding, and a different method is necessary at the very least because it is important to determine whether the result in experiment is method-dependent or not. In this context, eyetracking is a potentially interesting method because, unlike self-paced reading, both first-pass and non-first pass measures can be investigated, and the reading task is more natural than self-paced reading.

3. Experiment 2: Eyetracking

3.1 Participants

Thirty-five native-speaker students at the University of Potsdam took part in the experiment, either for course credit or for payment.

3.2 Method

Participants were seated approximately 50 cm from a 19-inch color LCD monitor with 1024×768 pixel resolution; twenty-eight pixels equaled about one degree of visual angle.

The eyetracker used was an SR Research Eyelink 1000 eyetracker running at 500 Hz sampling rate with 0.01 degree tracking resolution and a gaze position accuracy of < 0.5 degree. Although viewing was binocular, only data from the right eye was used in analyses. For stability, participants were asked to place their head on a chin-rest and a forehead-rest. Participants were instructed to avoid large shifts in position throughout the experiment. Responses were recorded by a 7-button

Microsoft Sidewinder game pad. The presentation of the materials and the recording of the responses was controlled by a notebook Windows PC running EyeTrack, which was interfaced with the eyetracker. Each stimulus sentence was presented on two lines with the line break placed consistently after the comma (*Direktor,*). The space between the two lines was 221 pixels. This space was introduced in order to prevent participants from previewing the second line of the target sentence parafoveally while reading the first line.

Each participant was randomly assigned one of three stimuli lists which comprised different item-condition combinations according to a Latin Square. The trials per session were randomized individually per participant. At the start of the experiment, the experimenter performed the standard EyeLink calibration procedure, which involves participants looking at a grid of thirteen fixation targets in random succession. Calibration was 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 fixation target appeared 30 pixels distant from the left screen border, the position where the text would be aligned. The stimulus was presented only after the subject fixated on this target. After the participant had finished reading the sentence, they fixated a dot in the lower right corner of the screen and simultaneously pressed a specific button on the game pad. This triggered the presentation of a simple comprehension question which the participant had to answer either with the left ('no') or the right trigger button ('yes'). The text stimuli were presented using a Courier New font, printed in black on a white background. The characters (including spaces) were all the same width, approximately 9 pixels or 0.32 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 upper and lower boundaries were 15 pixels above and below the top and bottom of the line of text. The line of text was 17 pixels high.

3.3 Materials

We used sixty items, each with the three conditions shown in (4). In addition, forty filler sentences were pseudo-randomly interspersed with the target items. We did not use a larger number of filler items because in our experience more than 100 items results in significant participant fatigue.

3.4 Dependent measures in eyetracking

As mentioned earlier, eyetracking data has the great advantage that it provides a record of both first-pass and second pass reading; however, the richness of data is not necessarily an improvement over simpler methods such as self-paced reading; a central problem with calculating statistics on all available eyetracking dependent measures (as is often done in psycholinguistics) is that they tend to be highly correlated, making much of the statistical calculation uninformative. For completeness, in Appendix A we provide summary statistics of the major dependent measures, where we also include the definitions of the various dependent measures used. Since we were asked by a reviewer to compute statistics for all dependent measures, we did so, but we found no effects except in dependent measures related to re-reading.

In this section we provide only two sets of critical analyses, which relate to first-pass processing (first-pass regression probability) and second-pass processing (re-reading probability and re-reading time). First-pass regression probability at any word n is the probability of the eye moving leftward to a preceding word $n - k$ (where $k \geq 1$), after the word n has been fixated at least once. Re-reading probability for a word n is the probability of revisiting that word after having having made a first-pass through that word (in other words, it is total reading time minus first-pass reading time; see Appendix A for definitions). Re-reading time is the amount of time spent revisiting a word after having having made a first-pass through that word. As also discussed in Appendix A, there are two ways to define re-reading time; one (the standard approach) includes zeroes (Sturt, 2003) and the other only looks at pure second-pass time, excluding zero re-reading time (Vasishth et al., 2010). We will refer to the former as re-reading time, and the latter as second-pass time.

Re-reading time, as defined above, is considered the standard definition presumably because it also includes non-re-reading, i.e., the full dataset. It appears that second-pass time (as defined above) is never presented in papers; presumably, researchers believe that this measure yields no information about processing difficulty. This assumption is probably well-founded when the proportion cases where re-reading occurred is small (if nothing else, the low power that would result would make null results meaningless), but it makes little sense when a large proportion of re-reading is present in the data. Two published papers where re-reading proportion was high are Vasishth et al. (2008) and Vasishth et al. (2010). In the latter work, the authors systematically compared second-pass time as defined above with self-paced reading data from identical materials across German and English. These studies showed that SPR reading time and second-pass time had comparable results; in other words, it may be a mistake to ignore second-pass time as an informative dependent measure in eyetracking. For this reason, we include second-pass time in the analyses below. In the present dataset, re-reading proportion was low, and therefore power is low. Nevertheless, as shown below, the results are informative.

Re-reading probability is to our knowledge not used in the literature, but it is highly correlated to re-reading time (trivially so) and is useful in the present paper because it allows us to compare probabilities in both first-pass (regression probability) and second-pass, rather than probabilities in first-pass and reading times in second pass. The broad conclusions remain unchanged when we use re-reading time instead of re-reading probability, as shown in the next section.

3.5 Results

Table 3 shows first-pass regression probability, Table 4 shows re-reading probability, and Table 5 shows re-reading time and second-pass time. Interestingly, first-pass regression probability shows lower regression proportions in conditions (c) versus (b) at the pre-critical region as well as the critical region; this is consistent with the expectation-based account, but not with locality based accounts.

However, as shown in Table 4, second-pass or re-reading probability shows evidence consistent with locality-based predictions: re-reading probabilities are significantly higher in the post-critical region in the comparison (b,c) versus (a) and the comparison (b) versus (c). In addition, re-reading probability is marginally higher in the (b,c) versus (a) comparison in the pre-critical and critical regions, with p-values somewhat greater than 0.05.

Finally, as shown in Table 5, log re-reading time showed locality effects in the pre-critical region in the comparison involving (b,c) versus (a), and in the post-critical region for the comparison involving (c) versus (b); in the post-critical region the (b,c) versus (a) comparison was essentially significant, with t=1.98 (the distinction between a t-value of 1.98 and the critical t-value of 2 is

Table 3: Analyses for first-pass regression probability (experiment 2).

region	contrast	coef.	se	z-score	p-value
pre-critical	a vs b,c	-0.0624	0.2750	-0.23	0.82
	b vs c	-1.0574	0.3251	-3.25	<0.01 *
critical	a vs b,c	-0.0813	0.2392	-0.34	0.734
	b vs c	-0.5074	0.2827	-1.79	0.073
post-critical	a vs b,c	-0.0326	0.2481	-0.13	0.90
	b vs c	0.0345	0.2881	0.12	0.90

Table 4: Analyses for re-reading probability (experiment 2).

region	contrast	coef.	se	z-score	p-value
pre-critical	a vs b,c	0.331	0.190	1.75	0.08
	b vs c	-0.321	0.206	-1.56	0.12
critical	a vs b,c	0.2914	0.1632	1.78	0.074
	b vs c	0.2841	0.1828	1.55	0.120
post-critical	a vs b,c	0.3509	0.1702	2.06	0.040 *
	b vs c	0.4627	0.1875	2.47	0.014 *

small enough to consider this effect significant). Second-pass reading time showed a locality effect in the pre-critical region in the comparison (c) versus (b); this was in spite of the low power that resulted from removing 82.7% of the cases where no re-reading occurred (in the critical region 69.3% involved no re-reading and in the post-critical region 76.3%).

3.6 Discussion

Since re-reading probability and re-reading time deliver essentially the same results, we focus our discussion on the former. The central findings in the eyetracking study are that (i) first-pass regression probability shows a facilitation in processing with increasing distance, but only in the pre-critical and critical regions (in the latter, the effect is marginal, p=0.07); and re-reading probability shows locality cost at all three regions, with marginal effects in the pre-critical (p=0.08) and critical regions (p=0.07), and a statistically significant effect in the post-critical region.

Thus, in experiment 2, we find evidence favoring the expectation-based account in a first-pass measure, and evidence favoring locality accounts in a measure that includes second pass. This result is in harmony with the idea, proposed in the discussion section of experiment 1, that both expectation-based facilitation and locality cost play a role in determining processing cost but that these two factors operate at different stages of processing. Specifically, at the region (the noun) preceding the verb, it is plausible that the parser already retrieves the verb-phrase as soon as the noun is processed, and that this process completes faster in the long-distance conditions due to the higher expectation for the upcoming verb phrase—this is completely in line with Levy's (2008) proposal. Once the verb phrase is retrieved, a dependency must be established between the subject and the verb phrase; here, the dependency completion cost is affected by decay. Since this process occurs after the verb phrase is retrieved, the locality cost expresses itself in second-pass reading

Table 5: Analyses for log re-reading time and log second-pass time (experiment 2).

		log RRTs			log Second-pass time		
region	contrast	coef.	se	t-value	coef.	se	t-value
pre-critical	a vs b,c	0.2643	0.1317	2.01 *	-0.01155	0.0871	-0.1
	b vs c	-0.2314	0.1521	-1.52	0.1606	0.0798	2.0 *
critical	a vs b,c	0.27505	0.1519	1.81	0.0182	0.0698	0.3
	b vs c	0.27543	0.1753	1.57	-0.0670	0.0767	-0.9
post-critical	a vs b,c	0.2856	0.1441	1.98 *	-0.0749	0.0693	-1.1
	b vs c	0.4227	0.1664	2.54 *	0.0519	0.0748	0.7

measures. A similar proposal was made by Sommerfeld et al. (2007), based on eyetracking and ERP data from experiments involving locality manipulations (the design in those studies was different from the present paper's).

We turn now to a third study, where we use the EEG methodology to replicate these results in a different setting; as discussed above, our motivation for doing the EEG study was to determine whether the results are robust and whether they can be replicated in methods other than self-paced reading and eyetracking.

4. Experiment 3: Event-related potentials

4.1 Participants

Twenty-two undergraduate students from the University of Potsdam participated in the ERP study.

4.2 Materials

There were a total of 360 sentences (180 critical items and 180 of unrelated filler sentences). The sentences were split in three versions to avoid repetition of lexical material. Each subjects saw 60 critical sentences (see example 4) intermixed with 60 unrelated sentences which makes a total of 120 sentences. The sentences were presented in a pseudo-randomized order.

4.3 Procedure

Twelve training sentences (4 in each of the critical conditions, see above) were presented to the participants. After this training set, the 60 critical sentences and the 60 unrelated filler sentences were randomly presented in the center of a 17" computer screen, with 400 ms (plus 100 ms inter-stimulus interval) for each word. 500 ms after the last word of each sentence a comprehension question was presented on the screen for 2500 ms. The questions were designed to have 50% *yes* answers and 50% *no* answers, respectively. The task for the subjects was to answer the questions within a maximal interval of 3000 ms by pressing one of two buttons. 1000 ms after their response, the next trial began.

The EEG was recorded by means of 25 Ag/AgCl electrodes with a sampling rate of 250Hz (impedances $< 5\text{k}\Omega$) and were referenced to the left mastoid (re-referenced to linked mastoids offline). The horizontal electro-oculogram (EOG) was monitored with two electrodes placed at the outer canthus of each eye and the vertical EOG with two electrodes above and below the right eye.

4.4 Data preprocessing

Only the trials that did not have artifacts were selected for the ERP analysis. The data were filtered with 0.2 Hz (high pass) to compensate for drifts. Single subject averages were computed in a 1000 ms window relative to the onset of the critical item (*ignoriert* 'ignored') and aligned to a 200 ms pre-stimulus baseline. One time window was analyzed: 300-500 ms.

4.5 Results

The ERP patterns from the onset of the critical item (the verb *ignoriert* 'ignored', onset at 0 ms) up to 1000 ms thereafter are displayed for the electrodes FC5, CP5 and C3, in figure 3, which shows the grand average ERPs for the control condition compared to the two more complex conditions at one electrode. Visual inspection shows that the more complex conditions (4b,c) are more negative-going compared to the control condition (4a).

Table 6: Results of linear mixed model fit for experiment 3. Participants were random factors; the time window is 300-500 ms, and the electrodes are FC5, C3, CP5. Random slopes by subject were fit for the two Helmert contrasts described in the text.

contrast	coefficient	se	t-value
a bs b,c	-1.184	0.373	-3.17 *
b vs c	0.596	0.431	1.38

As summarized in Table 6, for three electrodes (FC5, C3, and CP5), we see a significant negativity for the contrast (b,c) vs (a). The coefficient in Table 6 is negative because of that fact that the ERP response is negative-going; by contrast, in the SPR and eyetracking studies, the coefficients had a positive sign because more complex conditions tended to show longer reading times and higher re-reading probability.

The ERP pattern and topology is reminiscent of the Left Anterior Negativity, which has been argued to be triggered in long-distance dependency resolution (Kluender and Kutas, 1993a). If only the electrode FC5 is analyzed, we see only a marginal negativity (t=-1.9); in addition, no effects were seen in other analyses involving the three frontal (F3, FZ, F4), central (C3, CZ, C4), and posterior (P3, PZ, P4) electrodes.

4.6 Discussion

In the ERP study, we find a negative-going potential in the non-local versus local conditions; this pattern is seen in the 300–500 ms window from the onset of the verb ignoriert, ignored.

The negativity has a fronto-central distribution; the topology is not dissimilar from the Left Anterior Negativity, which has been seen in long-distance wh-dependency resolution (Kluender and Kutas, 1993b,a). In these studies it was interpreted as indexing a "looking back" process which was triggered by the attempt to integrate syntactic information with material occurring earlier in the structure. Additionally, other researchers have found a negativity on the verb when the parser was checking for an appropriate subject and was trying to integrate this information (Osterhout and Holcomb, 1992; King and Kutas, 1995; Vos et al., 2001). Thus, consistent with the self-paced reading study and the eyetracking experiment, the ERP study also provides evidence for greater integration cost as argument-head distance increases.

5. General Discussion

We begin by discussing the evidence from the three experiments for the locality hypothesis, and then we turn to the evidence in the eyetracking data for expectation-based processing. The evidence for locality is summarized across the three methods in Table 7.

It is clear from Table 7 that we see greater processing difficulty in the non-local conditions mostly at the critical and post-critical regions; in the SPR study we see a slowdown in the non-local conditions in the pre-critical region as well. The latter effect could of course be a Type I error;

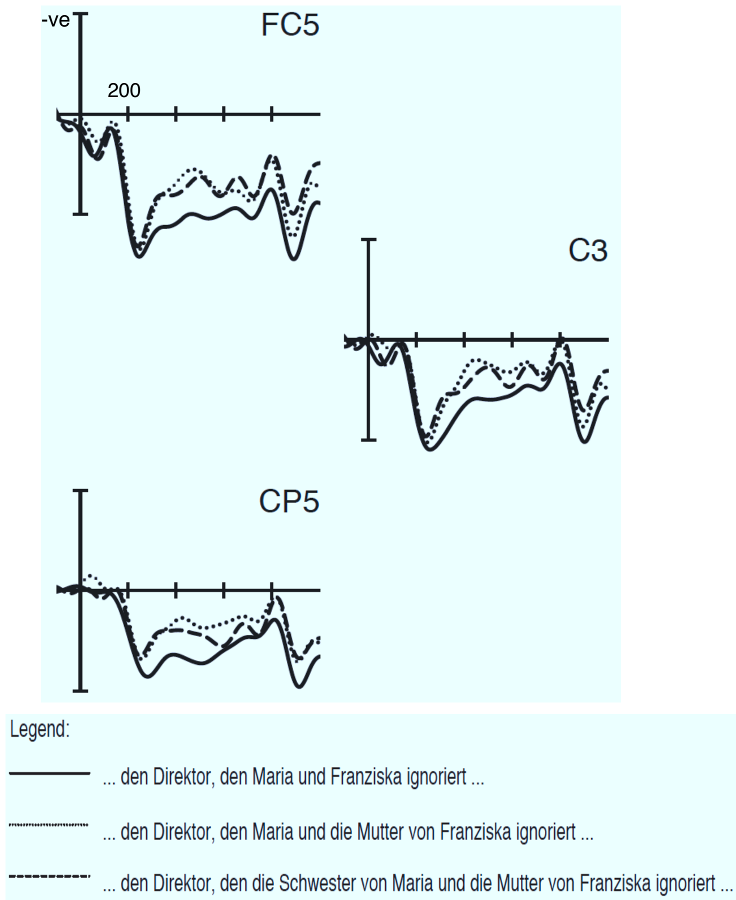

Figure 3: ERP voltage averages for the three conditions: control condition (solid), the more complex condition (dotted), and the most complex condition (dashed) at the left fronto-central electrode FC5. Time onset of the critical stimulus (the verb) at 0 s. Negativity is plotted upwards. For presentation purposes only, ERPs were filtered off-line with 8 Hz low pass.

Table 7: A summary of the locality effects found across the three experiments. Only statistically significant effects are shown.

	Experiment 1 (SPR)			
region	comparison	coef.	se	t-value
pre-critical	a vs b,c	0.0702	0.0204	3.4 *
	b vs c	0.0628	0.0235	2.7 *
critical	a vc b,c	0.0526	0.0219	2.4 *
post-critical	a vs b,c	0.0516	0.0262	2.0 *
	b vs c	0.0816	0.0302	2.7 *
	Experiment 2 (Eyetracking)			
	contrast	coefficient	se	z-value
post-critical	a vs b,c	0.3509	0.1702	2.06 *
	b vs c	0.4627	0.1875	2.47 *
	Experiment 3 (ERP)			
	contrast	coefficient	se	t-value
critical	a bs b,c	-1.184	0.373	-3.17 *

indeed, we did not expect an effect before the critical region in SPR. However, as mentioned earlier, if we assume that the effect is actually present, a plausible explanation would be that the parser may have retrieved the predicted verb phrase as soon as the pre-critical region is processed, and a retrieval of the grammatical subject may have been started as soon as the parser retrieves the verb phrase (or even while it is retrieving the verb phrase—nothing speaks against parallel execution of the prediction-integration steps). Such a mechanism could account for the slowdown seen in the pre-critical region.

Taken together, these findings deliver considerable evidence consistent with the locality hypothesis (Gibson, 2000; Lewis and Vasishth, 2005) and go directly against the expectation-based account of Levy (2008). Specifically, the statement by Levy (2008, 1144) that "more preverbal dependents gives [sic] the comprehender more information with which to predict the final verbs identity and location, and comprehension should therefore be easier" is incorrect. It may well be true that the interveners give more information about the final verb's identity and location, but it is not true that comprehension is necessarily easier: the costs associated with integration processes cannot be avoided.[2]

Why did our materials show a locality effect when previous work (e.g., Konieczny (2000)) has consistently shown antilocality effects? We believe that this comes from a particular property of our materials. All the items contain multiple instances of proper names in all three conditions; this makes the retrieval of *Direktor* (the argument of the verb) more difficult across all conditions due to generally high interference Lewis and Vasishth (2005); Van Dyke (2007). The presence of multiple candidate noun phrases that could be a subject of the verb *ignoriert* could result in this generally increased difficulty in retrieving the correct noun at the verb. Thus, in these stimuli, integration cost dominates over expectation-based facilitation.

2. For an alternative explanation for locality effects in terms of Good Enough processing, see Christianson and Luke (2010). Our data are consistent with this view as well; the central idea is that in long-distance dependencies the parser builds only partial trees, not complete structures as Gibson (2000) and Lewis and Vasishth (2005) assume. This may well be the correct way to characterize locality effects.

Note that we do not deny a role for expectation-based facilitation of the type that Levy advocates. In fact, the eyetracking record provides direct evidence for the expectation-based view. Specifically, in first-pass regression probability we see, in the pre-critical region a difference between condition (c) and (b), such that regression probability is lower in the non-local condition. We have independent evidence that regression probability has widely been acknowledged to index increased processing load. For example, Boston et al. (2008) and Boston et al. (2010) show, using the Potsdam Sentence Eyetracking Corpus (Kliegl et al., 2006) and two distinct types of grammars (probabilistic context free grammars and dependency grammars) derived from a treebank corpus, that both expectation cost (quantified as surprisal) and retrieval cost independently explain reading times.

If the above explanation is correct, this suggests that expectation-based facilitation and locality interact in an interesting manner: expectation plays a dominant role only when working memory load is relatively low. A related idea has been suggested independently by Gibson (2007); he suggested that expectation-based costs may dominate in relatively simple structures and/or more frequently occurring constructions, and distance-based costs may dominate in more complex and/or less frequently occurring constructions.

In sum, the present work limits the extent to which the expectation-based account can explain processing difficulty to very specific situations. This is consistent with the remark by Levy (2008) that expectation-based accounts are not an alternative account of dependency-resolution phenomena, but rather an additional constraint. One consequence of treating integration cost and expectation-based facilitation as two (largely) orthogonal factors is that one can dominate over the other depending on factors such as working memory load. Further support for this view comes from recent work by Boston et al. (2010), as discussed above.

The contribution of this paper is to provide novel empirical evidence from three methods demonstrating how integration cost can be detected independent of any advantage due to sharpened expectations for upcoming word-types.

In this appendix, we provide detailed summary statistics for experiment 2. We begin by providing definitions of the most common dependent measures used in reading research.

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.

First-pass regression probability is the probability of the eye moving leftward from a currently fixated word to a preceding word; it has been argued to index increased processing load.

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 (integration) processes (Gordon et al., 2006, 1308). Most researchers (e.g., Birch and Rayner (1997), Sturt (2003)) include zero re-reading times in their calculation of re-reading time; this appears to be the standard method in eyetracking research.

A minority (Vasishth et al. (2010, 2008), and possibly also Gordon et al. (2006)) exclude zero reading times in re-reading time. The approach, although not standard, has been shown to yield results comparable to self-paced reading times in English and German (Vasishth et al., 2010) (to our knowledge there has been no similar cross-method comparison of the ‘standard’ re-reading time, i.e., re-reading time including zeroes). In the present paper, we use a binary version of re-reading time, re-reading probability. This is a binary response variable, 1 for re-reading, 0 for no re-reading. This measure is essentially identical to using the ‘standard’ version of re-reading time except that the response is a binomial variable. Using a binomial version of re-reading has the advantage in the present study that we can compare two probabilities rather than probabilities and reading time: regression probability (a first-pass measure) and re-reading probability (a second-pass measure).

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.

Table 8: Re-reading probability.

Word pos	(a)	(b)	(c)
1	0.06 (0.01)	0.26 (0.02)	0.28 (0.02)
2	0.38 (0.03)	0.11 (0.02)	0.09 (0.02)
3	0.13 (0.02)	0.21 (0.02)	0.44 (0.03)
4	0.32 (0.02)	0.44 (0.03)	0.64 (0.03)
5	0.05 (0.01)	0.15 (0.02)	0.25 (0.02)
6	0.12 (0.02)	0.42 (0.03)	0.55 (0.03)
7	0.36 (0.03)	0.56 (0.03)	0.11 (0.02)
8	0.07 (0.01)	0.24 (0.02)	0.25 (0.02)
9	0.35 (0.03)	0.49 (0.03)	0.37 (0.03)
10	0.52 (0.03)	0.07 (0.01)	0.11 (0.02)
11	0.2 (0.02)	0.29 (0.02)	0.26 (0.02)
12	0.42 (0.03)	0.07 (0.01)	0.02 (0.01)
13	0.09 (0.01)	0.13 (0.02)	0.1 (0.02)
14	0.26 (0.02)	0.33 (0.03)	0.31 (0.02)
15	0.06 (0.01)	0.1 (0.02)	0.04 (0.01)
16	0.22 (0.02)	0.28 (0.02)	0.24 (0.02)
17	0.29 (0.02)	0.31 (0.02)	0.36 (0.03)
18	0.22 (0.02)	0.24 (0.02)	0.31 (0.02)

Table 9: First-fixation duration.

Word pos	(a)	(b)	(c)
1	44 (5)	116 (7)	110 (6)
2	164 (5)	48 (5)	49 (5)
3	78 (6)	97 (6)	203 (7)
4	195 (7)	194 (6)	243 (6)
5	30 (4)	71 (6)	99 (7)
6	109 (7)	193 (7)	211 (6)
7	198 (6)	249 (6)	64 (6)
8	63 (6)	106 (7)	104 (7)
9	192 (7)	230 (6)	167 (6)
10	241 (6)	55 (6)	81 (6)
11	108 (8)	140 (6)	185 (7)
12	227 (7)	44 (5)	29 (4)
13	56 (6)	101 (7)	102 (7)
14	128 (6)	181 (6)	169 (6)
15	39 (4)	79 (6)	53 (5)
16	170 (7)	170 (6)	165 (6)
17	222 (5)	238 (6)	232 (6)
18	188 (6)	195 (6)	187 (6)

Table 10: First-pass reading time.

Word pos	(a)	(b)	(c)
1	86 (7)	209 (14)	200 (10)
2	234 (11)	74 (5)	76 (5)
3	109 (7)	144 (7)	254 (9)
4	244 (9)	292 (12)	361 (12)
5	48 (5)	103 (7)	154 (7)
6	141 (8)	255 (10)	409 (19)
7	307 (13)	351 (11)	136 (9)
8	87 (6)	168 (8)	180 (9)
9	246 (9)	391 (16)	261 (12)
10	353 (11)	154 (10)	121 (7)
11	152 (8)	232 (10)	229 (9)
12	394 (16)	63 (5)	46 (5)
13	147 (9)	133 (8)	135 (8)
14	199 (8)	235 (10)	208 (8)
15	56 (5)	109 (7)	76 (6)
16	207 (8)	203 (8)	204 (8)
17	288 (9)	297 (8)	291 (8)
18	211 (6)	220 (7)	216 (7)

Table 11: Regression path duration.

Word pos	(a)	(b)	(c)
1	86 (7)	209 (14)	200 (10)
2	291 (14)	105 (9)	113 (9)
3	136 (10)	171 (11)	337 (18)
4	325 (19)	389 (21)	458 (22)
5	78 (10)	148 (15)	239 (15)
6	161 (10)	369 (24)	871 (46)
7	367 (17)	413 (17)	138 (10)
8	115 (11)	274 (27)	248 (13)
9	311 (16)	931 (51)	361 (16)
10	423 (19)	170 (18)	157 (13)
11	236 (17)	364 (18)	280 (13)
12	815 (44)	93 (12)	60 (7)
13	152 (11)	175 (13)	171 (15)
14	304 (14)	323 (24)	277 (20)
15	92 (10)	142 (12)	100 (10)
16	270 (22)	265 (15)	242 (14)
17	335 (17)	363 (18)	331 (15)
18	269 (17)	290 (20)	280 (19)

Table 12: Re-reading time.

Word pos	(a)	(b)	(c)
1	19 (5)	132 (16)	119 (15)
2	167 (16)	32 (6)	23 (5)
3	36 (7)	89 (14)	217 (20)
4	166 (18)	267 (25)	362 (22)
5	13 (4)	43 (6)	82 (10)
6	40 (8)	209 (20)	292 (22)
7	196 (21)	307 (22)	37 (7)
8	18 (4)	73 (9)	75 (10)
9	148 (15)	260 (20)	162 (18)
10	268 (21)	25 (5)	26 (5)
11	62 (8)	119 (15)	104 (12)
12	222 (21)	15 (3)	4 (2)
13	25 (5)	33 (5)	32 (7)
14	100 (14)	132 (14)	129 (14)
15	14 (3)	24 (4)	10 (3)
16	75 (10)	113 (13)	100 (12)
17	114 (15)	123 (14)	131 (12)
18	73 (9)	74 (9)	99 (11)

Table 13: Total reading time.

Word pos	(a)	(b)	(c)
1	105 (10)	341 (24)	318 (20)
2	401 (20)	106 (9)	99 (8)
3	145 (11)	233 (17)	471 (24)
4	410 (20)	559 (29)	723 (26)
5	60 (7)	146 (10)	237 (14)
6	182 (12)	464 (23)	700 (29)
7	503 (25)	658 (25)	172 (14)
8	105 (8)	241 (13)	255 (14)
9	394 (18)	652 (25)	423 (22)
10	621 (24)	179 (12)	147 (10)
11	214 (13)	351 (18)	333 (16)
12	616 (26)	78 (7)	50 (5)
13	171 (12)	166 (10)	166 (11)
14	298 (16)	367 (18)	336 (17)
15	69 (7)	133 (9)	85 (7)
16	283 (14)	316 (17)	303 (16)
17	403 (18)	420 (16)	422 (15)
18	284 (12)	294 (12)	315 (14)

References

J. R. Anderson, P. Kline, and C. Lewis. A production system model for language processing. In M. Just and P. Carpenter, editors, *Cognitive Processes in Comprehension*. Lawrence Erlbaum Associates, Hillsdale, NJ, 1977.

R.H. Baayen and P. Milin. Analyzing Reaction Times. To appear in International Journal of Psychological Research, 2010.

M. Babyonyshev and E. Gibson. The complexity of nested structures in Japanese. *Language*, 75(3): 423–450, 1999.

B. Bartek, R. L. Lewis, S. Vasishth, and M. Smith. In search of on-line locality effects in sentence comprehension. Submitted, 2010.

S. L. Birch and K. Rayner. Linguistic focus affects eye movements during reading. *Memory and Cognition*, 25(5):653–60, 1997.

M. F. Boston, J. T. Hale, U. Patil, R. Kliegl, and S. Vasishth. Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. *Journal of Eye Movement Research*, 2(1):1–12, 2008.

Marisa F. Boston, John T. Hale, Shravan Vasishth, and Reinhold Kliegl. Parallelism and syntactic processes in reading difficulty. *Language and Cognitive Processes*, 2010.

N. Chomsky. *Aspects of the Theory of Syntax*. MIT Press, Cambridge, MA, 1965.

K. Christianson and S. G. Luke. The ham stands alone: A Good Enough processing account of locality and local coherence. in revision, 2010.

E. Gibson. Locality and anti-locality effects in sentence comprehension. Workshop on processing head-final languages, 2007.

E. Gibson. Dependency locality theory: A distance-based theory of linguistic complexity. In Alec Marantz, Yasushi Miyashita, and Wayne O'Neil, editors, *Image, Language, brain: Papers from the First Mind Articulation Project Symposium*. MIT Press, Cambridge, MA, 2000.

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

E. Gibson and J. Thomas. Memory limitations and structural forgetting: The perception of complex ungrammatical sentences as grammatical. *Language and Cognitive Processes*, 14(3):225–248, 1999.

P. C. Gordon, R. Hendrick, M. Johnson, and Y. Lee. Similarity-Based Interference During Language Comprehension: Evidence from Eye Tracking During Reading. *Journal of Experimental Psychology: Learning Memory and Cognition*, 32(6):1304–1321, 2006.

D. Grodner and E. Gibson. Consequences of the serial nature of linguistic input. *Cognitive Science*, 29:261–290, 2005.

F. Hsiao and E. Gibson. Processing relative clauses in Chinese. *Cognition*, 90:3–27, 2003.

A. W. Inhoff. Two stages of word processing during eye fixations in the reading of prose. *Journal of verbal learning and verbal behavior*, 23(5):612–624, 1984.

F. T. Jaeger, E. Fedorenko, P. Hofmeister, and E. Gibson. Expectation-based syntactic processing: Antilocality outside of head-final languages. In *CUNY Sentence Processing Conference*, North Carolina, 2008.

M.A. Just and P. A. Carpenter. A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99(1):122–149, 1992.

M.A. Just and P.A. Carpenter. A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87(4):329–354, 1980.

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

J.W. King and M. Kutas. Who did what and when? Using word-and clause-level ERPs to monitor working memory usage in reading. *Journal of cognitive neuroscience*, 7(3):376–395, 1995.

R. Kliegl, A. Nuthmann, and R. Engbert. Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General*, 135:12–35, 2006.

R. Kluender and M. Kutas. Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience*, 5(2):196–214, 1993a.

R. Kluender and M. Kutas. Subjacency as a processing phenomenon. *Language and Cognitive Processes*, 8(4):573–633, 1993b.

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

R. Levy. Expectation-based syntactic comprehension. *Cognition*, 106:1126–1177, 2008.

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

D. C. Mitchell. An evaluation of subject-paced reading tasks and other methods of investigating immediate processes in reading. In D. E. Kieras and M.A. Just, editors, *New Methods in Reading Comprehension Research*. Erlbaum, Hillsdale, N.J., 1984.

L. Osterhout and P.J. Holcomb. Event-related brain potentials elicited by syntactic anomaly. *Journal of memory and language*, 31(6):785–806, 1992.

K. Rayner and A. Pollatsek. Eye movements in reading: A tutorial review. *Attention and performance XII: The psychology of reading*, pages 327–362, 1987.

E. Sommerfeld, S. Vasishth, P. Logačev, M. Baumann, and H. Drenhaus. A two-phase model of integration processes in sentence parsing: Locality and antilocality effects in german. In *Proceedings of the CUNY sentence processing conference*, La Jolla, CA, 2007.

P. Sturt. The time-course of the application of binding constraints in reference resolution. *Journal of Memory and Language*, 48:542–562, 2003.

J. Van Dyke. Interference effects from grammatically unavailable constituents during sentence processing. *Journal of Experimental Psychology: Learning Memory and Cognition*, 33(2):407–30, 2007.

J. Van Dyke and R. L. Lewis. 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, 2003.

S. Vasishth and M. Broe. *The Foundations of Statistics: A Simulation-based Approach*. Springer, Berlin, 2011.

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

S. Vasishth, S. Bruessow, R. L. Lewis, and H. Drenhaus. Processing polarity: How the ungrammatical intrudes on the grammatical. *Cognitive Science*, 32(4), 2008.

S. Vasishth, K. Suckow, R. L. Lewis, and S. Kern. Short-term forgetting in sentence comprehension: Crosslinguistic evidence from head-final structures. *Language and Cognitive Processes*, 25(4): 533–567, 2010.

W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S-PLUS*. Springer, New York, 1999.

S.H. Vos, T.C. Gunter, H.H.J. Kolk, and G. Mulder. Working memory constraints on syntactic processing: An electrophysiological investigation. *Psychophysiology*, 38(01):41–63, 2001.

T. C. Warren. *Understanding the Role of Referential Processing in Sentence Complexity*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 2001.

T. C. Warren and E. Gibson. Effects of NP-type on reading English clefts. *Language and Cognitive Processes*, pages 89–104, 2005.