

# On the proper treatment of spillover in real-time reading studies: Consequences for psycholinguistic theories

Shravan Vasishth  
University of Potsdam, Germany  
vasishth@acm.org

In recent psycholinguistic research, the effect of predictability in incremental processing has become an important theoretical issue. Dependency locality theory (Gibson, 2000), for example, assumes a monotonically increasing processing cost as a function of (inter alia) the number of new discourse referents intervening between a head and a dependent (e.g. a verb and its argument). Hawkins' Early Immediate Constituents (Hawkins, 1994) provides a similar metric, and in fact EIC's validity depends on the idea of locality being empirically confirmed. Given that experimental and corpus studies of English have repeatedly provided evidence for this idea, psycholinguists and syntacticians have come to believe that such distance-based effects provide a robust explanation for processing difficulty. Interestingly, not much attention is paid to the fact (see, e.g. (Hawkins, 2004)) that the explanation is simply wrong in the case of head-final languages like German (Konieczny, 2000), Hindi (Vasishth, 2003), and Japanese (Nakatani and Gibson, 2004).

Restricting our attention only to English then, one might ask, just how strong is the experimental evidence for this locality effect? Consider for example an important recent demonstration of non-locality by Grodner and Gibson (2005): In (1a) the verb *supervised* and its argument *nurse* are adjacent, but in (1b) and (1c) a PP and an RC intervene respectively. The locality hypothesis predicts increased processing difficulty at the embedded verb *supervised* and the main verb *scolded* if the interposed phrases contain new discourse referents.

- (1) a. The administrator who the nurse supervised scolded the medic while ...
- b. The administrator who the nurse *from the clinic* supervised scolded the medic while ...
- c. The administrator who the nurse *who was from the clinic* supervised scolded the medic while ...

Most experimental research on locality is faced with an interesting confound in the design of the stimuli: since the material preceding the critical region (the verb in this case) is not identical, reading times at the critical region are possibly confounded by spillover, defined by Mitchell (1984, 76) as follows:

In most immediate processing tasks the end of one response measure is immediately followed by the beginning of another, together with a new portion of text. In this situation any uncompleted processing will spill over from one response measure to the next. In others words, certain aspects of processing will be postponed and join a queue or buffer so that they can be dealt with later. . . . Here, the response measure will be influenced not only by the problems in the current display but also by any backlog or processing that may have built up in the buffer.

In other words, it is possible that the critical region of interest is swamped by processing continuing from the (immediately) preceding region. Since this preceding region differs in the local versus non-local conditions, any significant difference observed at the critical region could be a function, at least partly, of the preceding region's processing difficulty.

Resolving this issue is critical for psycholinguistic research because a large number of studies has targeted this question, all of them involving the confounding factor of spillover; a small sample is the work presented in (Christianson, 2002), (Grodner and Gibson, 2005), (Konieczny, 2000), (Nakatani et al., 2000), (Vasishth, 2003), (Warren and Hirotsu, 2005).

An anonymous reviewer for (Vasishth and Lewis, 2005) suggests that in order to resolve this issue, residuals rather than raw reading times be analyzed. This approach involves determining for each subject  $i$  a separate regression equation (1) that predicts reading time  $Y_i$  at a critical region  $n + 1$  from the reading time  $X_i$  at the immediately preceding region  $n$ . The error in prediction  $\epsilon_i$  for each subject (the residuals) is the unexplained variation, which can be used as the reading time that can be attributed to the locality manipulation.

$$Y_i = \beta_1 + \beta_2 \times X_i + \epsilon_i \quad (1)$$

Then, for each subject a set of residual scores can be calculated by subtracting each subject's regression equation estimates from the observed scores, and an analysis of variance carried out on the residuals. This approach is commonly used in psycholinguistics to factor out the effect of word length on a word's reading time (Ferreira and Clifton, 1986).

Although the residuals approach is reminiscent of carrying out an analysis of covariance (ANCOVA), there are several problems with it (Maxwell et al., 1985; García-Berthou, 2001), the most serious being that Type I error rates increase.

I show here that linear mixed-effects models (Pinheiro and Bates, 2000) provide a better and more informative approach. In such models, two classes of effects are distinguished, random and fixed. In the case of the spillover problem, the participants (and items) are the random effects (the experimental conditions being nested within these), and the experimental conditions and spillover are the fixed effects. More generally (and ignoring repeated measurements for simplicity of exposition), if  $y_{ij}$  is the  $j$ -th observation in the  $i$ -th group,  $x_{ij}$  is the corresponding value of the continuous covariate

(here the preceding region’s RT), a separate random effects term  $b_i$  can be defined for each observation (i.e. for each subject), and the main effect (in our example the locality manipulation) constitutes the intercept term  $\beta_1$  (equation (2)). (For nested effects in repeated measures settings, a further term must be included, see (Pinheiro and Bates, 2000) for details).

$$y_{ij} = \beta_1 + b_i + \beta_2 \times x_{ij} + \varepsilon_{ij} \quad (2)$$

$$(i = 1, \dots, M, j = 1, \dots, n_i, b_i \sim \mathcal{N}(0, \sigma_b^2), \varepsilon_{ij} \sim \mathcal{N}(0, \sigma^2))$$

I now reexamine Grodner and Gibson’s experimental data<sup>1</sup> by correcting for spillover using the linear-mixed effects model.<sup>2</sup> I show that spillover from the intervening region seems to be the reason for the slowdown observed in this Grodner and Gibson experiment. Once spillover is factored out, the locality effect disappears, at least in this experiment.

Interposed item	Locality Effect	Spillover Effect	Interaction
PP	<b>xx</b>	<b>x✓</b>	<b>✓✓</b>
RC	<b>xx</b>	<b>x✓</b>	<b>✓✓</b>

Table 1: Summary of linear mixed-effects model analysis at the embedded verb in Grodner and Gibson’s Experiment 2. Locality Effect refers to the predicted slowdown.

The mixed-effects analysis shows in addition that the effect of spillover is stronger than any slowdown predicted by the locality hypothesis. At the embedded verb, PP-interposition did not have a significant effect

( $F(1, 48) = 0.54, p = 0.5$ ;  $F(1, 29) = 0.46, p = 0.5$ ), but spillover showed an effect in the by-items analysis ( $F(1, 390) = 0.016, p = 0.9$ ;  $F(1, 428) = 5.71, p = 0.02$ ), and there was an intervention-spillover interaction ( $F(1, 390) = 6.0, p = 0.02$ ;  $F(1, 428) = 14.23, p = 0.0002$ ). RC-interposition showed a slight slowdown in the by-subjects analysis ( $F(1, 48) = 2.85, p = 0.1$ ;  $F(1, 29) = 2.96, p = 0.1$ ), and spillover showed an effect in the by-items ( $F(1, 390) = 1.79, p = 0.2$ ;  $F(1, 428) = 13.91, p = 0.0002$ ). A marginal interaction was seen in by-items ( $F(1, 390) = 2.33, p = 0.13$ ;  $F(1, 428) = 3.45, p = 0.06$ ). Table 1 summarizes these results.

As summarized in Table 2, at the main verb, PP-interposition had no detectable effect ( $F(1, 48) = 0.37, p = 0.6$ ;  $F(1, 29) = 0.37, p = 0.6$ ), and spillover had an effect in by-items ( $F(1, 390) = 0.62, p = 0.4$ ;  $F(1, 428) = 9.60, p = 0.002$ ). There was no interaction ( $F(1, 390) = 0.48, p = 0.5$ ;  $F(1, 428) = 1.13, p = 0.23$ ). The RC condition showed no intervention effect ( $F(1, 48) = 1, p = 0.33$ ;  $F(1, 29) = 0.91, p = 0.4$ ), a marginal spillover effect in by-items ( $F(1, 390) = 0.08, p = 0.8$ ;  $F(1, 428) = 3.37,$

<sup>1</sup>I thank Daniel Grodner for graciously providing me with the raw data.

<sup>2</sup>This reanalysis was also done for three other studies, and also compared with the standard residuals-based analyses, but for space reasons I do not discuss these results in this abstract.

Interposed item	Locality Effect	Spillover Effect	Interaction
PP	<b>xx</b>	<b>x✓</b>	<b>xx</b>
RC	<b>xx</b>	<b>xx</b>	<b>xx</b>

Table 2: Summary of linear mixed-effects model analysis at the main verb in Grodner and Gibson’s Expt. 2.

$p = 0.07$ ), and no interaction ( $F1(1, 390) = 0.087$ ,  $p = 0.8$ ;  $F2(1, 428) = 0.019$ ,  $p = 0.9$ ).

In sum, the mixed effects analysis suggests that spillover may play a dominant role in the processing slowdowns observed in experiments that manipulate locality. An important point to note is that the claim is not that locality plays no role. The argument is rather that such correction should be carried out in reading-time studies in order to avoid misleading results; it is entirely possible that even stronger evidence will emerge for locality where none was previously found (Warren and Hirotani, 2005). Furthermore, Grodner (personal communication) has suggested that the effect of position must also be factored out for a meaningful discussion of spillover effects. I am in the process of reanalyzing the data with this additional correction.

A further possibility is that spillover plays a bigger role in self-paced reading experiments compared to eyetracking studies. This is likely since self-paced reading forces the participant to maintain previously seen words in memory, and prevents him/her from previewing words to the right of the word currently being processed. In order to explore this possibility, an experiment with a locality manipulation was performed using both self-paced reading and eyetracking; the results of the locality manipulation after factoring out spillover will be discussed.

To conclude, this paper make two points. First, (psycho)linguists need to become aware of the well-known fact that residuals are inappropriate alternatives to ANCOVA, and a better alternative is available. Second, the evidence for locality and predictability in processing needs a careful reinvestigation by systematically taking into account the effect of spillover. Not doing so can lead to possibly misleading conclusions about the constraints on real-time parsing processes.

## References

- Christianson, K. T. (2002). Sentence processing in a “nonconfigurational” language. Ph.D. thesis, Michigan State University, East Lansing.
- Ferreira, F. and J. Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, **25**:348–368.
- García-Berthou, E. (2001). On the misuse of residuals in ecology: Testing regression residuals vs. the analysis of covariance. *Journal of Animal Ecology*, **70**:708–711.

- Gibson, E. (2000). Dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, and W. O'Neil, eds., *Image, Language, brain: Papers from the First Mind Articulation Project Symposium*. MIT Press, Cambridge, MA.
- Grodner, D. and E. Gibson (2005). Consequences of the serial nature of linguistic input. *Cognitive Science*, **29**:261–290.
- Hawkins, J. A. (1994). *A Performance Theory of Order and Constituency*. Cambridge University Press, New York.
- Hawkins, J. A. (2004). *Efficiency and Complexity in Grammars*. Oxford University Press.
- Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research*, **29**(6):627–645.
- Maxwell, S. E., H. D. Delaney, and J. M. Manheimer (1985). ANOVA of residuals and ANCOVA: Correcting an illusion by using model comparisons and graphs. *Journal of Educational Statistics*, **10**:197–209.
- Mitchell, D. C. (1984). An evaluation of subject-paced reading tasks and other methods of investigating immediate processes in reading. In D. E. Kieras and M.A. Just, eds., *New Methods in Reading Comprehension Research*. Erlbaum, Hillsdale, N.J.
- Nakatani, K., M. Babyonyshev, and E. Gibson (2000). The complexity of nested structures in Japanese. Poster presented at the CUNY Sentence Processing Conference, University of California, San Diego.
- Nakatani, K. and E. Gibson (2004). An online study of Japanese nesting complexity. MS.
- Pinheiro, J. C. and D. M. Bates (2000). *Mixed-Effects Models in S and S-PLUS*. Springer-Verlag, New York.
- Vasishth, S. (2003). *Working memory in sentence comprehension: Processing Hindi center embeddings*. Garland Press, New York. Published in the Garland series *Outstanding Dissertations in Linguistics*, edited by Laurence Horn.
- Vasishth, S. and R. L. Lewis (2005). Argument-head distance and processing complexity: Explaining both locality and anti-locality effects. Submitted to *Language*.
- Warren, T. and M. Hirotani (2005). Memory influences on the processing negative polarity items. In *Polarity meets psycholinguistics*. University of Potsdam.