Frequency effects in the processing of unbounded dependencies

Marten van Schijndel (vanschm@ling.osu.edu)
Department of Linguistics
The Ohio State University

William Schuler (schuler@ling.osu.edu)
Department of Linguistics
The Ohio State University

Peter W. Culicover (culicove@ling.osu.edu)
Department of Linguistics
The Ohio State University

Abstract

The results of several self-paced reading and eye-tracking studies (e.g., Pickering and Traxler, 2003) have been interpreted to suggest that readers do not use verb subcategorization preference as an early means for completing unbounded dependencies. Subsequent papers (e.g., Arai and Keller, 2013) have hypothesized that this finding may actually be due to the frequency of larger syntactic configurations. This paper utilizes a robust statistical model and finds evidence in support of the latter interpretation: apparent lack of sensitivity to subcategorization preference is shown to be explainable by a confounding frequency effect of syntactic configuration.

Keywords: Sentence processing, Frequency effects, Probabilistic models, Subcategorization

Introduction

Unbounded dependencies (e.g., between *the book* and *about* in the noun phrase *[the book]$_i$ the author wrote about t_i*) consist of a filler (*the book*) and an attachment site or gap (t_i) which can be separated by an unbounded number of words. Since gaps are not overtly represented in sentences, their locations can be temporarily ambiguous (e.g., after *wrote* or after *about*). Some researchers have suggested that maintaining such dependencies introduces additional processing difficulty (Chomsky & Miller, 1963; Gibson, 2000). In order to quickly resolve ambiguous unbounded dependencies and ease any potential difficulty, one might expect readers to make full use of the information at their disposal to complete unbounded dependencies as soon as possible.

Several self-paced reading and eye-tracking studies have explored whether readers make use of subcategorization preferences of verbs in order to immediately restrict the hypothesis space of unbounded dependency attachments (Mitchell, 1987; Pickering, Traxler, & Crocker, 2000; van Gompel & Pickering, 2001; Pickering & Traxler, 2003). Subcategorization preference or bias may be ascertained by observing how frequently a verb appears with given argument types (a verb that appears very frequently with a noun phrase (NP) direct object but occasionally without any direct object argument would be deemed an optionally transitive verb with a transitive bias to take NP arguments). The following sentences from Pickering and Traxler (2003) are representative of the stimuli used in such experiments:

(1) That's the plane that the pilot landed carefully behind in the fog at the airport.

(2) That's the truck that the pilot landed carefully behind in the fog at the airport.

These authors claim that if readers use subcategorization frequency in processing, the implausibility in (2) of *truck* as an argument of *landed* should not cause readers to slow since *landed* prefers a prepositional phrase (PP) argument to a noun phrase (NP) argument. Instead, readers of (2) do slow down compared to readers of (1) after reading *landed*, which the authors claim suggests that they have difficulty with the implausible interpretation of (2) that arises from the attachment of the unbounded dependency to the verb in spite of its subcategorization bias (pilots don't usually land trucks). These previous studies have interpreted such results as an indication that subcategorization frequency is not used by readers when initially resolving unbounded dependencies; rather, readers seem to employ a simple early-attachment heuristic.

This paper reviews recent articles from the psycholinguistics literature which suggest an alternative, frequency-based explanation for this finding. It then goes on to show how a probabilistic context free grammar (PCFG) can be constructed from corpora annotated with unbounded dependencies and used to estimate the frequency effects involved in unbounded dependency processing. This analysis shows that the slow-down in sentences such as (2) may be explained by the frequencies of non-preterminal syntactic configurations which may have a much stronger impact than subcategorization preferences.

Background

Since Pickering and Traxler (2003), a number of studies have revisited the claim that subcategorization is not used by readers in initial attachment of unbounded dependencies. For example, Staub, Clifton, and Frazier (2006) conducted two experiments that explored the time course of processing a particular kind of unbounded dependency: heavy-NP shift constructions. They found that sentences containing an optionally transitive verb with a transitive bias (e.g., *The teacher helped immediately [the confused student]*) were processed more slowly upon encountering the shifted region than sentences containing an obligatorily transitive verb (e.g., *The teacher corrected immediately [the unusual answer]*). They interpret their results as evidence that readers adopt a parsing heuristic that disprefers a heavy-NP shift interpretation rather than purely relying on the subcategorization bias of the verb.[1]

[1] While Staub et al. (2006) argue for a serial model of language processing, this paper remains agnostic with respect to whether processing is serial or parallel.

Otherwise, verbs with a transitive bias would force an initial transitive reading to be adopted and so would not yield this pattern of slowing. They point out, however, that their results may have been driven by the infrequency of heavy-NP shift as a construction. That is, the infrequency of heavy-NP shift may overwhelm any transitivity bias of the verb. The PCFG analysis described in this paper may be construed as a formalization of this analysis.

Arai and Keller (2013) suggest a similar frequency explanation of the findings of Pickering and Traxler (2003) based on visual world experiments similar to those of Altmann and Kamide (1999), Kamide, Altmann, and Haywood (2003), and Kamide, Scheepers, and Altmann (2003). By observing where subjects' eyes moved as they listened to sentences containing transitive or intransitive verbs, they found evidence that subjects do take selectional information into account at the verb. Specifically, plausible arguments and complements are fixated on more frequently than implausible ones when verbs with either subcategorization bias are heard. Based on this finding, they speculated that the findings of studies such as Pickering and Traxler (2003) could be due to the frequency of main clause direct object constructions when compared with the alternative constructions supported by verbal subcategorization, but they did not evaluate this claim.

Finally, Staub (2007) conducted three self-paced reading studies in an attempt to remove confounds from Pickering et al. (2000) and similar studies. By separating the intransitive verbs used in those studies into unaccusative verbs (e.g., *erupt*), which can never take a direct object argument, and unergative verbs (e.g., *sneeze*), which can take direct object arguments under particular circumstances, Staub was able to construct a set of sentences that were truly obligatorily intransitive. When reading a sentence containing an obligatorily intransitive (unaccusative) verb, readers did not show any evidence of attaching the filler of the unbounded dependency to the verb, unlike in the unergative case where there was a slight chance of obtaining a transitive interpretation. This finding indicates that any possibility of transitive interpretation, even when that possibility is very slight, can cause readers to adopt implausible analyses, which suggests that the frequency of a direct object interpretation can overwhelm the lexically-specific bias of a verb.

This paper directly investigates these recent claims that earlier findings of insensitivity to verb subcategorization bias may be due to syntactic configuration frequency. If the probability of a syntactic configuration is defined as the product of the probabilities of its component syntactic configurations and its lexical items, a very small or very large syntactic probability (e.g., that of heavy-NP shift, or the prevalence of direct object complements) could overwhelm verb-specific argument biases.

Probabilistic Grammars

Probabilities for syntactic configurations can be obtained by assigning probabilities to grammar rules. For example, a prepositional phrase (PP) usually generates a preposition (P) and a noun phrase (NP). Each such rule in the grammar may be assigned a conditional probability based on the frequency with which that parent category generates those child categories in large corpora. The resulting probability-weighted grammar is called a *probabilistic context-free grammar* (PCFG, Booth & Thompson, 1973). The probability of a syntactic configuration can then be estimated as the product of these conditional rule probabilities.

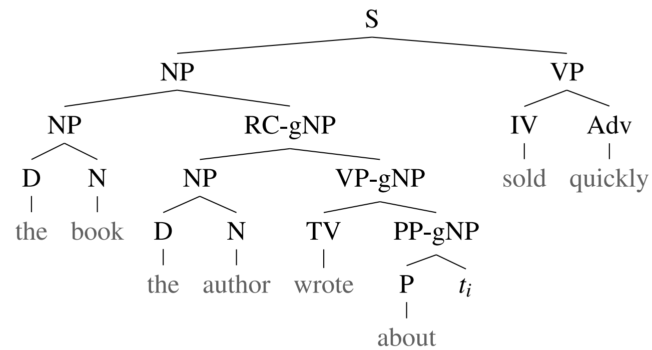

Figure 1: An example syntactic analysis of *The book the author wrote about sold quickly* with a GCG-like treatment of unbounded dependencies. The gap is annotated with t_i in the figure, only. Each category is sensitive to whether it has an unresolved gap within its subtree.

Well-studied algorithms exist for finding and refining PCFGs from data (Petrov, Barrett, Thibaux, & Klein, 2006), and PCFGs have been shown to be useful as a basis for information-theoretic accounts of garden path effects and reading time delays (Jurafsky, 1996; Hale, 2001, 2003, 2006; Levy, 2008). Usually, however, PCFG models have excluded unbounded dependency information because of its inherent complexity. In order to capture unbounded dependency information and still use existing algorithms for obtaining highly accurate PCFGs, this paper uses a *generalized categorial grammar* (GCG) (Bach, 1981; Nguyen, van Schijndel, & Schuler, 2012), which passes unbounded dependencies from parents to children and so makes the propagation of a gap into a category context-free (solely dependent on whether a gap exists in the parent category and on whether the preceding sibling could serve as a filler).

The Nguyen et al. (2012) GCG encodes gap information using a **-g** operator added to categories that contain a gap (see Figure 1), so a verb phrase (VP) with a gapped NP argument would be assigned the category VP-gNP and would expand to a child transitive verb (TV) and a gap associated with an NP. To link this gap to the correct filler, this GCG propagates the **-g** from the sibling category of the filler to each appropriate child in the syntax tree in a fashion similar to the SLASH category of Head-driven Phrase Structure Grammar (HPSG) (Pollard & Sag, 1994) and other HPSG-like context-free gap notations (Hale, 2001; Lewis & Vasishth, 2005).

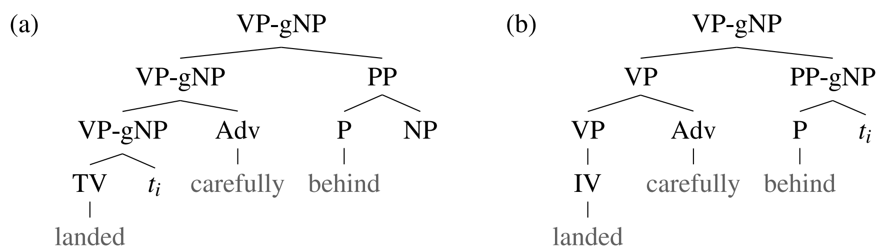

Figure 2: Alternative parses of a portion of *That's the plane/truck that the pilot landed carefully behind in the fog at the airport,* shown immediately after observing the word *behind*. The predicted syntactic category of the next observation is shown, and gaps are annotated with t_i. Parse (a) corresponds to a transitive NP interpretation and Parse (b) corresponds to an intransitive PP interpretation.

Interpretation	Grammar rule	Prob
Transitive	VP-gNP→VP-gNP PP	0.17
Intransitive	VP-gNP→VP PP-gNP	0.01

Table 1: The probability of the grammar rules associated with transitive and intransitive interpretations during incremental resolution of unbounded dependencies as calculated from the Wall Street Journal text corpus. These numbers are based on the 2,355 occurrences of VP-gNP in the corpus.

Evaluation

Aside from verb subcategorization bias, the difference between the transitive and intransitive interpretations of the sentences in Pickering and Traxler (2003) and related studies is that a transitive interpretation hypothesizes the gap as the complement of the main verb, whereas an intransitive interpretation hypothesizes the gap as the complement of the preposition (see Figure 2). In order to quantify the frequency interactions that may be behind the findings of studies such as Pickering and Traxler (2003), the Wall Street Journal (WSJ) portion of the Penn Treebank corpus of English (Marcus, Santorini, & Marcinkiewicz, 1993) is reannotated using a GCG as described by Nguyen et al. (2012).

Counts of each syntactic configuration in this reannotated corpus indicate that the intransitive interpretation is much less frequent than the transitive interpretation (see Table 1). Since the parse of each interpretation is otherwise equivalent up to the verb, the probability of subjects entertaining each possible interpretation may be computed by taking the product of each grammar rule probability in Table 1 and the probability of a subsequent rule generating the verb from a given preterminal category (since all other relevant grammar rules are common to both interpretations). The probability of a verb being generated from a given preterminal category (TV or IV) is proportional to (∝) the subcategorization bias of the verb divided by the prior probability of the preterminal:[2]

$$\begin{aligned} \text{P(Trans)} &= \text{P(VP-gNP} \rightarrow \text{VP-gNP PP)} \cdot \text{P(}\textit{verb} \mid \text{TV)} \\ &\propto \text{P(VP-gNP} \rightarrow \text{VP-gNP PP)} \cdot \frac{\text{P(TV} \mid \textit{verb}\text{)}}{\text{P(TV)}} \quad (1) \\ \text{P(Intrans)} &= \text{P(VP-gNP} \rightarrow \text{VP PP-gNP)} \cdot \text{P(}\textit{verb} \mid \text{IV)} \\ &\propto \text{P(VP-gNP} \rightarrow \text{VP PP-gNP)} \cdot \frac{\text{P(IV} \mid \textit{verb}\text{)}}{\text{P(IV)}} \quad (2) \end{aligned}$$

Counts from the reannotated corpus reveal that transitive verbs (TV) are *a priori* 2.6 times as likely as intransitive verbs (IV), so this is used as a normalizing factor in the evaluation.

For ease of comparison, this paper makes use of the bias frequencies obtained during the verb norming study of Pickering and Traxler (2003), which were obtained by asking 90 subjects to write sentences containing the relevant verbs and counting the number of times a verb appeared in a given configuration.[3] As an example, *landed* appeared with an NP 25% of the time, a PP 40% of the time, and with neither 35% of the time. Using the above formula of *rule·bias/prior*, this means the probability of *landed* inducing a transitive NP complement interpretation in subjects is $0.17 \cdot 0.25/2.6 = 0.016$ compared with the probability of *landed* inducing the intransitive PP complement interpretation in subjects, which is $0.01 \cdot 0.4/1 = 0.004$. The NP complement interpretation of *landed* is thus 300% more likely for subjects to adopt than a PP complement interpretation, despite the prima facie bias for *landed* to take a PP complement.

[2] An additional adverb (e.g., *carefully*) is sometimes used to increase the ambiguous region of each sentence during reading experiments. Although the probabilities for rules with and without the adverb are different, including the probabilities of adverbial rules (VP → VP Adv and VP-gNP → VP-gNP Adv) and the probabilities of preterminal rules (VP → IV and VP-gNP → TV) does not change the direction of the effect reported in this paper and generally increases the magnitude (with preterminal rules, the probability of transitive interpretation: 0.046 and intransitive interpretation: 0.0001; with adverbial rules, probability of transitive interpretation: 0.0078 and intransitive interpretation: 0.0001), so they are omitted for clarity.

[3] Pickering and Traxler (2003) also determined the subcategorization bias of each verb using other norming studies, but the study that yielded the results used in this paper had the largest subject pool. Using one of their other sets of bias results does not significantly affect the results of this paper.

This disparity directly arises from the substantially greater probability of propagating a gap dependency to a VP child than to a PP child. On average, the PP-biased verbs used in Pickering and Traxler (2003) have an intransitive bias of 0.52 and a transitive bias of 0.14, which means the average PP-biased verb is nearly twice as likely to induce a transitive interpretation than an intransitive interpretation in subjects (NP interpretation: $0.17 \cdot 0.14/2.6 = 0.009$; PP interpretation: $0.01 \cdot 0.52 = 0.005$). Even the second most PP-biased verb used by Pickering and Traxler (2003), *searched*, which appeared with an NP 15% of the time and with a PP 75% of the time, is more likely to receive an NP interpretation than a PP interpretation (NP interpretation: $0.17 \cdot 0.15/2.6 = 0.01$; PP interpretation: $0.01 \cdot 0.75 = 0.008$). Lacking a representative number of verbs with a strong enough subcategorization bias to induce a PP-interpretation, it is unsurprising that such studies have failed to observe an effect of verb subcategorization bias.

Discussion

A possible criticism of using frequency probabilities derived from the WSJ corpus is that the lexicon or the distribution of syntactic configurations may not generalize well to other domains (Sekine, 1997; McClosky, 2010). However, the lexeme-specific probabilities used in this study were determined experimentally by Pickering and Traxler (2003), so they do not depend on the WSJ lexicon. Only the syntactic rule probabilities are derived from the WSJ corpus; however, Nguyen et al. (2012) showed that a parser trained only on a GCG-reannotated WSJ corpus can achieve state-of-the-art parsing accuracy for unbounded dependencies in a variety of domains (news, narrative, etc). This finding suggests the distribution of unbounded dependencies in the WSJ corpus is representative of the distribution of English unbounded dependencies as a whole.

The same probability model given in this paper can be used to account for the findings of Staub (2007) that readers do not mistakenly attach fillers to unaccusative verbs (e.g., *erupt*). Since unaccusative verbs cannot take an NP argument, the probability of *erupt* inducing an NP transitive interpretation is $0.17 \cdot 0.0/2.6 = 0.0$.

Further, this model can account for the findings of Staub et al. (2006) regarding reading times of heavy-NP shift constructions. Though it is beyond the scope of this paper to detail the syntactic analyses that are involved, heavy-NP shift constructions are less frequent than unshifted constructions. Using a similar analysis to that given in this paper, it may be shown that this model replicates the findings of Staub et al. (2006): obligatorily transitive verbs should cause readers to slow at the inserted material in shifted constructions (because shifted constructions are less frequent than unshifted constructions) and optionally transitive verbs should cause readers to slow at the shifted NP (because the infrequency of shifted constructions outweighs all but the strongest transitive biases). Interestingly, preliminary results exploring heavy-NP shift with this model indicate that optionally transitive verbs with a transitive bias of around 87% may yield a slow-down at the inserted adverb (when compared with adverbs in unshifted, optionally transitive constructions) rather than the object noun since that optional transitive bias should outweigh the bias of intransitive constructions but not completely outweigh the preference to not shift. Such a finding was not observed by Staub et al. (2006), but their optionally transitive verbs did not approach this level of transitive bias.[4]

The effectiveness of this model at accounting for a variety of experimental findings has potential implications for theories of human sentence processing. For example, this model assumes that subcategorization information (e.g., the number and type of required arguments) is present immediately during parsing, regardless of its regularity. In contrast, some theories of sentence processing like the Garden Path Model (Frazier, 1987) or Construal (Frazier & Clifton, 1996) posit that only regular grammatical patterns (e.g., transitive verbs) are immediately available to the parser, whereas irregular exceptions only become available during a later stage in processing. Such theories have typically been supported by findings (e.g., Pickering & Traxler, 2003, and Pickering et al., 2000) that subcategorization is not immediately used during sentence processing. While the present study does not rule out multi-stage processing models altogether, it does show that processing can make immediate use of subcategorization biases and still replicate findings which had been interpreted as showing that subcategorization is not used immediately during processing. Therefore, a pool of supporting evidence that was previously thought to strongly favor multi-stage processing models should no longer be considered to do so.

Conclusion

While it may be true that verbs have specific subcategorization preferences, this paper has shown that the overwhelming bias to propagate a gap into a verb phrase rather than a prepositional phrase sibling will override all but the strongest subcategorization preferences during online processing. In fact, an optionally transitive verb would have to appear with a PP 6.5 times for every 1 NP (87% intransitive bias) in order to have an even chance of inducing a PP complement interpretation compared with an NP complement interpretation. This work, therefore, provides quantitative evidence in support of recent suggestions (Staub et al., 2006; Staub, 2007; Arai & Keller, 2013) that previous findings of reader insensitivity to verb subcategorization preference may be due to the frequencies of the syntactic configurations involved. This finding highlights the need to account for frequency at multiple levels of processing rather than simply in terms of lexical biases.

[4] A script to replicate all findings in this paper is available upon request. The replication script also confirms the preliminary heavy-NP shift analysis given here.

Acknowledgements

Thanks to Matthew Traxler and the attendees of the Ohio State University syntactic processing seminar for feedback on the original idea. Further thanks to Shari Speer who gave feedback on an earlier draft of this paper. This work was funded by an Ohio State University Department of Linguistics Targeted Investment for Excellence (TIE) grant for collaborative interdisciplinary projects conducted during the academic year 2012-13.

References

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

Arai, M., & Keller, F. (2013). The use of verb-specific information for prediction in sentence processing. *Language and Cognitive Processes*, *28*(4), 525–560.

Bach, E. (1981). Discontinuous constituents in generalized categorial grammars. *Proceedings of the Annual Meeting of the Northeast Linguistic Society (NELS)*, *11*, 1–12.

Booth, T. L., & Thompson, R. A. (1973). Applying probability measures to abstract languages. *IEEE Transactions on Computers*, *C-22*(5), 442–450.

Chomsky, N., & Miller, G. A. (1963). Introduction to the formal analysis of natural languages. In *Handbook of mathematical psychology* (pp. 269–321). New York, NY: Wiley.

Frazier, L. (1987). Sentence processing: A tutorial review. In M. Coltheart (Ed.), *Attention and performance 12: The psychology of reading* (pp. 559–586). Hillsdale, NJ: Erlbaum.

Frazier, L., & Clifton, C., Jr. (1996). *Construal*. Cambridge, MA: MIT Press.

Gibson, E. (2000). The dependency locality theory: A distance-based theory of linguistic complexity. In *Image, language, brain: Papers from the first mind articulation project symposium* (pp. 95–126). Cambridge, MA: MIT Press.

Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. In *Proceedings of the second meeting of the north american chapter of the association for computational linguistics* (pp. 159–166). Pittsburgh, PA.

Hale, J. (2003). *Grammar, uncertainty and sentence processing*. Unpublished doctoral dissertation, Cognitive Science, The Johns Hopkins University.

Hale, J. (2006). Uncertainty about the rest of the sentence. *Cognitive Science*, *30*(4), 609–642.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science: A Multidisciplinary Journal*, *20*(2), 137–194.

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.

Kamide, Y., Scheepers, C., & Altmann, G. T. M. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research*, *32*(1), 37–55.

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

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

Marcus, M. P., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: the Penn Treebank. *Computational Linguistics*, *19*(2), 313–330.

McClosky, D. (2010). *Any domain parsing: Automatic domain adapatation for parsing*. Unpublished doctoral dissertation, Computer Science Department, Brown University.

Mitchell, D. C. (1987). Lexical guidance in human parsing: Locus and processing characteristics. In M. Coltheart (Ed.), *Attention and performance xii: The Psychology of Reading* (pp. 601–618). Hillsdale, NJ: Erlbaum.

Nguyen, L., van Schijndel, M., & Schuler, W. (2012). Accurate unbounded dependency recovery using generalized categorial grammars. In *Proceedings of the 24th international conference on computational linguistics (coling '12)* (pp. 2125–2140). Mumbai, India.

Petrov, S., Barrett, L., Thibaux, R., & Klein, D. (2006). Learning accurate, compact, and interpretable tree annotation. In *Proceedings of the 44th annual meeting of the association for computational linguistics (COLING/ACL'06).*

Pickering, M. J., & Traxler, M. J. (2003). Evidence against the use of subcategorisation frequency in the processing of unbounded dependencies. *Language and Cognitive Processes*, *18*(4), 469–503.

Pickering, M. J., Traxler, M. J., & Crocker, M. W. (2000). Ambiguity resolution in sentence processing: Evidence against frequency-based accounts. *Journal of Memory and Language*, *43*, 447–475.

Pollard, C., & Sag, I. (1994). *Head-driven phrase structure grammar*. Chicago: University of Chicago Press.

Sekine, S. (1997). The domain dependence of parsing. In *Proceedings of the Fifth Conference on Applied Natural Language Processing* (pp. 96–102). Association for Computational Linguistics.

Staub, A. (2007). The parser doesn't ignore intransitivity, after all. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 550–569.

Staub, A., Clifton, C., & Frazier, L. (2006). Heavy NP shift is the parser's last resort: Evidence from eye movements. *Journal of Memory and Language*, *54*, 389–406.

van Gompel, R. P. G., & Pickering, M. J. (2001). Lexical guidance in sentence processing: A note on Adams, Clifton, and Mitchell (1998). *Psychonomic Bulletin and Review*, *8*, 851–857.