Three Benchmarks for Distributional Approaches to Natural Language Syntax

Colin Phillips

Department of Linguistics
University of Maryland
College Park, MD 20782
colin@umd.edu

Abstract

Human language abilities are far richer than what is represented in the kinds of monolingual corpora that are standardly used to evaluate statistical models of language learning. This article summarizes a series of findings from language acquisition, cross-language typology, and language processing, that illustrate the challenges that any serious model of natural language syntax must meet. Even a putative ideal statistical learner of cooccurrences in corpora will struggle to meet the challenges of complexity, cross-language consistency, and causality, unless it is able to take advantage of the rich representational primitives motivated by linguistics and psycholinguistics.

1 Complexity

There has been a good deal of recent interest in statistical learning models for language (Manning & Schütze 1999) and in evidence that humans can learn and use at least some distributional statistics (MacDonald, Pearlmutter, and Seidenberg 1994; Saffran, Aslin, and Newport 1996). Although it has been shown that there are some quite simple statistical patterns that humans are not good at tracking (Newport and Aslin 2004; Pena et al. 2002), there appears to be widespread optimism in some circles that discovery of the right statistical learning procedure will ultimately provide an adequate account of human linguistic abilities, and obviate the need for the richly structured representations widely assumed in the linguistics tradition.

Rather than discussing the successes and limitations of current statistical models of language, the goal of this paper is to draw attention to problems that will remain even if it is possible to construct an 'ideal' learner of surface distributional facts. These problems can be divided into three types of challenges, each of which is likely to encompass a large variety of linguistic phenomena. *Complexity* refers to the fact that human linguistic abilities extend far beyond what one finds in the corpora that appear to be the standard benchmarks for current statistical learning models. Speakers are able to make fine-grained judgments about sentence structures that they have encountered rarely or never at all before. I will not discuss specific examples of the complexity of human syntactic abilities in this section, since the cases discussed in sections 2 and 3 already serve this role. *Consistency*, discussed in Section 2, refers to the contribution of linguistic typology. Cross-language variation is surprisingly limited, particularly when one looks beyond the simplest and most frequent phenomena, and it is unclear how powerful statistical learning models can explain this consistency, or the evidence for early emergence of universal constraints in language development. *Causality*, discussed in Section 3 refers to the fact that humans appear to be very sensitive not only to the fact *that* a reliable statistical co-occurrence obtains, but also to *why* it obtains.

2 Consistency

Linguists tend to be impressed by the consistency of cross-language variation. Many properties or constraints appear to hold in all languages, other sets of properties show consistent co-variation across languages (e.g., if a language has property X it also has property Y). Human languages seem to carve out just a small part of the space of conceivable languages. If this cross-language consistency is matched by consistency in surface distributional phenomena, then it may be possible to develop a learning model that relies on surface statistics. However, if cross-language consistency occurs in the face of very diverse surface distributions, then it is more likely that the consistency reflects constraints that are built into the learner (whether these are language-specific constraints or general cognitive constraints—my current arguments do not distinguish between these alternatives). Therefore, it is important to determine whether cross-language consistency is matched by consistent surface statistics, and whether the development of constraints is better predicted by the consistency of their surface patterns or by the consistency of their cross-language variation.

2.1 Backward anaphora

A pronoun stereotypically appears in a context where it is preceded in discourse by a full NP that serves as its *antecedent*. There are also fully natural cases where the pronoun precedes its antecedent, known as *backward anaphora*, as in (1a), where coreference is indicated by coindexation. However, there are specific environments where backwards anaphora is ruled out, as in (1b). A standard account of the impossibility of (1b) is that the pronoun c-commands its antecedent, in violation of Condition C of the Binding Theory (Chomsky 1981).[i]

(1) a. While he_i was eating an apple, $John_i$ read a book.

b. * He_i ate an apple while $John_i$ was reading a book.

This constraint is very real, and can be shown without reliance on conscious grammaticality judgments. For example, we have recently shown that Condition C applies immediately in language processing, using a self-paced reading study (Kazanina, et al. 2004). When the gender of the second subject NP *John* is manipulated in sentences like (1), such that it mismatches the gender of the pronoun (e.g., ... *he* ... *Mary* ...), there is an immediate slowdown in reading times at the mismatching NP in sentences like (1a), suggesting that speakers expect to find an antecedent for the pronoun in that position, as previously shown by van Gompel and Liversedge (2003). However, the same speakers show no similar gender mismatch effect at the corresponding NP in examples like (1b), suggesting that the NP is not even considered as a potential antecedent, as expected if Condition C acts as an immediate filter on structure building.

In language development studies Condition C appears to be active as early as it is possible to test children. English-speaking 3-year olds allow coreference in sentences like (1a) but not in sentences like (1b) (Crain and McKee 1985). In addition, Condition C is a good candidate for a cross-language universal, appearing in languages as diverse as English, where the surface facts are relatively straightforward, but also in languages such as Mohawk, where the facts of free word order and liberal argument omission make the effects of Condition C rather more difficult to detect (Baker 1991).

Many linguists are impressed by the correspondence between the early emergence of Condition C in development and the fact that its cross-linguistic distribution is more consistent than its surface realization across-languages, and conclude from this that it is a constraint that need not be learned. An alternative possibility is that future advances in statistical learning tools might someday succeed in deriving the contrast between (1a) and (1b) from the input available to 2-year olds, and will also succeed in deriving similar contrasts in Mohawk and other languages. Since universals of language have no privileged status for this putative learner, we can suppose that it would be sophisticated enough to also rapidly learn constraints on backward anaphora that are very similar to Condition C, but are not universal. Russian provides an interesting case that we have investigated in a project led by Nina Kazanina. Russian has similar word order properties to English, and like English it disallows coreference in Condition C contexts (2b), but unlike English it also disallows coreference in (2a), which is the counterpart of (1a). There is no across-the-board ban on backwards anaphora in Russian (Avrutin and Reuland 2002; Kazanina and Phillips 2001).

(2) a. * Poka on$_i$ chital knigu, Pooh$_i$ s'el yabloko.
while he was reading.imp the book Pooh ate.perf the apple

b. * On$_i$ s'el yabloko, poka Pooh$_i$ chital knigu.
he ate.perf the apple while Pooh was reading.imp the book

In a Truth Value Judgment Task study with Russian 3-6 year olds (Kazanina and Phillips 2001), similar to that conducted previously in English, we found that 5-6 year olds reliably respect both the universal constraint (100%) and the language particular constraint on backwards anaphora (78%), just like Russian adults. More interestingly, the Russian 3-year olds gave judgments that were almost identical to their English-speaking contemporaries. They consistently disallowed coreference in (2b) (85%), as required by Condition C, but they consistently allowed coreference in (2a) (13%), in violation of the Russian-specific constraint. 4-year olds gave a mixed pattern of judgments.

The substantial developmental delay between mastery of Condition C and the Russian-specific constraint is not surprising in an approach where universals need not be learned, but language specific constraints clearly must be learned. The developmental delay is rather more surprising if the two similar constraints must be learned by the same sophisticated statistical mechanism. There are no viable current proposals about how Condition C effects might be inferred from the input to learners by age 3, but if such an account were found, it would run the risk of predicting that Russian children would also master the Russian-specific constraint at a similar age.

2.2 Verb argument structure

It is well known that verbs with similar meanings have similar syntactic properties (‘argument structure’). Children and adults are able to use knowledge about these syntax-semantics links to infer information about novel verbs (Gillette et al. 1999; Gropen et al. 1991). Importantly, detailed studies of English show that the syntactic classification of verbs follows very specific semantic parameters, such as transfer-of-possession, manner-of-motion or change-of-state (Levin 1993), rather than general notions of semantic similarity. While some have suggested that the specificity of these semantic parameters reflects in-built biases in the learner (Pinker 1989), others have argued that these generalizations can be learned from surface statistics (Seidenberg 1997), and have argued that the existence of cross-language variation makes this necessary. In a project led by Meesook Kim, we tested these ideas using a detailed survey of the syntax of 'locative' verbs, such as *pour*, *fill*, and *load* in 20 languages (Kim 1999; Kim, Landau, and Phillips 1999; Phillips, Rabbin, and Kim 2001). We found that there is indeed cross-language variation, but that it is restricted and rather systematic. More importantly, languages appear to consistently group verbs along the same semantic parameters.

For example, change-of-state verbs like *fill* (e.g., *cover*, *decorate*) in English disallow the 'figure' frame (3a), but allow the 'ground' frame (3b). Many languages treat these verbs in the same way that English does, but in many other languages (e.g., Japanese, Turkish, Luganda, Hindi, Thai) these verbs are alternators, allowing both of the constructions in (3). Although these languages diverge from English in their more liberal treatment of the *fill* class, we have found no languages that diverge from English in their treatment of the *pour* class, the members of which encode manner-of-motion meanings (e.g., *spill*, *drip*, *shake*) and only allow the figure frame (4). Thus, the association between manner-of-motion verbs and the figure frame appears to be a universal.[ii] Furthermore, those languages that allow change-of-state verbs in the *fill* class to alternate are all languages that allow verb compounding ('serialization') and null objects, whereas none of the English-type languages allow these. Thus, variation in this verb class is quite systematic, and appears to reflect what linguists call a 'parameter', i.e., a set of properties that consistently co-occur across languages (Baker 2001).

(3) a. *He filled water into the glass.

b. He filled the glass with water.

(4) a. He poured water into the glass.

b. * He poured the glass with water.

Furthermore, we have often found that distinctions among verb classes that seem to disappear in the most common surface verb phrase forms sometimes re-emerge in more obscure corners of the grammars of individual languages. To take just one example, verbs like *pile* (e.g., *scatter*, *load*) are alternators in English, allowing either the figure or the ground frame (5). In Korean, the counterpart of (5b) is unavailable, and thus the distinction between the *pour* class and the *pile* class appears to be lost. However, the distinction re-emerges in Korean verb compounding constructions. When Korean verbs from the *pile* class are compounded with *put*, to form compounds such as *pile-put*, both figure and ground frames become available. Note that neither *pile* nor *put* allows the ground frame when the verb is used alone. Meanwhile, compounds involving verbs from the *pour* class, e.g. *pour-put* do not allow alternation (Kim 1999). In other words, compounding recreates exactly the same distinction seen in English between the *pour* class and the *pile* class. We found other instances in which a language appeared at first sight to conflate a verb class distinction found in English, but turned out to preserve the distinction in constructions that we had good reason to believe are not frequent in the input to learners.

(5) a. He piled the books onto the shelf.

b. He piled the shelf with books.

The fact that languages consistently classify verbs using the same semantic parameters, even when this is obscured in the frequently occurring constructions of the language, suggests that something other than classification based on surface statistics is at work. The cross-language regularities observed in our survey are expected under an approach that views syntax-semantics links in argument structure as reflections of fixed constraints on language learners. On the other hand, under the view where verb classes reflect distributional analyses of frequently occurring surface forms in individual languages, the cross-language consistency is more surprising.

3 Causality

A dependency between two elements in a structure may be viewed in purely statistical terms, encoding the fact that the two elements reliably cooccur. Let us assume for the sake of discussion that statistical models are good at extracting such regularities, and can readily make use of observations of the form *after an X occurs, a Y is sure to follow (at some point) thereafter*. However, structural dependencies are generally much more than statistical regularities. Most reliable cooccurrences exist for a reason, and a number of results indicate that speakers are very sensitive to the causes of structural regularities. A statistical model that is able to extract regular cooccurrences but is not able to encode the causes of the regularities will miss important generalizations.

3.1 *Wh*-Questions

Wh-dependencies are the most widely investigated type of unbounded linguistic dependency. The head of the dependency is a fronted *wh*-phrase, and the tail of the dependency is a gap in an argument position in some models (Gibson and Hickok 1993; Haegeman 1994), or a predicate that selects the fronted phrase in other models (Pickering and Barry 1991; Sag and Fodor 1994). For ease of exposition, I will adopt the terminology of gap-based approaches here. It is likely that in English and other languages, most naturally occurring *wh*-dependencies span just a single clause, for the simple reason that most questions contain only one clause. A well-known generalization about the processing of English *wh*-questions is that after encountering a fronted *wh*-phrase, speakers actively attempt to construct a gap in the first available position, without waiting for confirmation that the position is available, as shown by processing disruption when speakers encounter an overt NP in a position where a gap may have been anticipated, highlighted in (6) (Crain and Fodor 1985; Stowe 1986). In other words, it appears that speakers try to construct short *wh*-dependencies.

(6) My brother wanted to know who Ruth will bring **us** home to __ at Christmas.

The preference for shorter *wh*-dependencies could be explained in a number of ways. It may reflect a specialized subroutine of the parser that favors short dependencies (Frazier and Clifton 1989), or it may reflect statistical biases accrued from large amounts of experience with short dependencies. Alternatively, it may reflect the more abstract property that in English a short *wh*-dependency allows the *wh*-phrase to receive a thematic role and enter into compositional semantic interpretation more quickly (Pritchett 1992). This last alternative specifically makes reference to the cause of the gap in a *wh*-dependency: the gap marks the thematic position of the *wh*-phrase and allows it to enter into compositional interpretation with the verb.

In a project led by Sachiko Aoshima, we have found that the on-line processing of Japanese *wh*-fronting constructions looks strikingly different from English when viewed in terms of surface structures, but looks very similar to English when viewed in terms of the cause of *wh*-dependency formation. Japanese *wh*-phrases canonically appear in-situ in argument positions, but fronting of *wh*-phrases is also possible, in a process known as *wh-scrambling*. Short-distance scrambling creates a dependency that spans just one clause; long-distance scrambling creates a dependency that spans multiple clauses, as in a sentence like (7). In (7) the fronted dative *wh*-phrase is an argument of the embedded verb *give*. The fact that the *wh*-phrase occurs in sentence initial position does not mark the sentence as a direct question, as would be the case in English. Rather, the presence of the question particle *-ka* on the embedded verb marks the sentence as an indirect question.

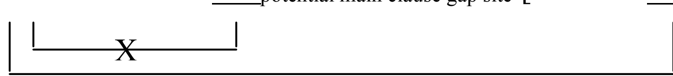
(7) Dare-ni John-wa [Mary-ga ___ sono hon-o ageta-ka] itta.

whom-dat John-top Mary-nom ___$_{gap}$ *that book-acc gave-Q said*

'John said who Mary gave that book to.'

In a series of 4 different experiments using Japanese bi-clausal sentences with a fronted *wh*-phrase we have found that Japanese speakers show a consistent bias for the long-distance scrambling construal of the fronted phrase (8) (Aoshima, Phillips, and Weinberg 2004; Yoshida, Aoshima, and Phillips 2004). Furthermore, our experiments show that this is not simply a consequence of the need to satisfy argument structure requirements of the embedded verb, as might be argued based on examples like (7). The same bias was found to be present in a comprehension task where the embedded verb does not require a dative argument, and also in a generation task where the participants were left to generate the embedded verb for themselves. Furthermore, on-line reading time profiles suggest that a gap is posited in the embedded clause before any verb is encountered.

(8) wh-dat NP-nom ___$_{\text{potential main clause gap site}}$ [NP-nom ___$_{\text{preferred gap site}}$ … V] … V



The bias for long-distance scrambling is unexpected under an account that incorporates a specific principle that favors gaps that are close to their antecedents. It is highly unlikely that the bias for long-distance scrambling reflects a predominance of long-distance scrambling in corpora of Japanese questions, although we have not specifically checked this statistic. The relevant configurations are quite rare in naturally occurring Japanese. However, the bias is expected under an approach that takes into account the syntactic and semantic motivation for completing a *wh*-dependency. The strongly verb-final property of Japanese means that a gap in the embedded clause provides the first opportunity for grammatical constraint satisfaction and compositional interpretation of the fronted phrase. Thus, there is good reason to posit a gap in the embedded clause where possible in Japanese. Under this approach, the same mechanism can account for generalizations in English and Japanese that appear quite different on the surface.

Although a statistical model that simply tracks surface co-occurrence probabilities is unlikely to capture the Japanese generalization (however sophisticated its learning algorithm might be), this does not mean that there is no role for a statistical model that uses sufficiently rich representations. It is probably true that in the majority of Japanese *wh*-questions, the grammatical requirements of the *wh*-phrase are satisfied at the first available opportunity, either because the question contains just one clause, or because the *wh*-phrase appears in situ in an embedded clause. A model that can abstract the motivation for *wh*-dependency formation in this way may be able to appropriately generalize to our findings about long-distance scrambling. On the other hand, a model that simply registers the frequent occurrence of local *wh*-dependencies will wrongly predict a bias for local scrambling. Furthermore, a model that attempts to explain our findings by ignoring the structural configurations of Japanese and blindly linking a sentence-initial *wh*-phrase with the first verb will also fail. The bias for long-distance scrambling is absent in configurations where Japanese grammar blocks long-distance scrambling, such as when the *wh*-phrase bears nominative case (Miyamoto and Takahashi 2003) or when the embedded clause is a relative clause (Yoshida et al. 2004).

3.2 Tense

Linguistic dependencies give rise to situations where the occurrence of one element reliably predicts the subsequent appearance of another related element, often many words later in the sentence. Many statistical models are rather good at picking out such regularities, and let us assume for the sake of discussion that there will be statistical models that are able to pick out any reliable co-occurrence. However, models that detect reliable co-occurrence do not typically encode *why* certain elements cooccur. In contrast, this is something that humans are quite good at. A recent ERP study of Hindi sentence processing conducted with Andrew Nevins provides a clear example of this (Nevins, Phillips, and Poeppel 2004). In this study, we were interested in two environments in Hindi that give rise to the same prediction, but for different reasons. The goal of the study was to test whether during sentence comprehension speakers carry forward just information about *what* material is predicted, or whether they also carry forward information about *why* that prediction was generated.

Like Japanese, Hindi exhibits primarily verb-final word order. This means that prior material in a clause may provide a variety of cues to the form of the verb. We investigated two types of reliable cues for the tense marking on the verb. As in other languages, the occurrence of a past tense adverbial in Hindi, such as *last week*, reliably predicts a past tense verb. This is a semantic cue. In addition, in transitive past tense clauses the case-marking on the subject NP serves as a predictor of past tense. Hindi has a split-ergative case system, in which present and future tenses are associated with nominative-accusative case marking, but past tense is associated with ergative-absolutive case marking. Therefore, the appearance of ergative case on a subject NP is a reliable predictor of past tense. This is a syntactic cue. The statistical properties of both dependencies appear to be the same. Adverbials and ergative case both reliably predict an upcoming past tense verb, and neither the adverbial nor the ergative case is necessary in a past tense clause. However, when speakers are presented with sentences containing tense violations, in which a past tense verb is substituted with a future tense verb, the ERP responses differ markedly, depending on whether the past tense was predicted by a semantic cue (9) or by a syntactic cue (10). Sentences were presented visually using devanagari script in an RSVP paradigm (650ms/word). ERPs were recorded from 30 scalp electrodes in 20 native speakers of Hindi. Violations of a semantically-cued past tense elicit an N400 response, which is characteristic of semantic violations (Kutas and Federmeier 2000; Kutas & Hillyard 1980). Violations of a syntactically-cued past tense elicit a P600 response, which is characteristic of syntactic violations (Hagoort, Brown, and Osterhout 1999).

(9) Haalanki pichle shaam vo rahgiir pathaar ke-upar giraa/ giregaa,

although last night that traveler-abs stone upon fall-pst/fall-fut

lekin use choT nahiin aayii

but to-him injuries not happen

'Although last night that traveler fell upon a stone, he was not injured.'

(10) Haalanki us bunkar-ne ek baRaa sveTar jaldi bunaa/ bunegaa

although that weaver-erg one big sweater quickly weave-pst/weave-fut

lekin graahak-ne sabhii-ki kimaat ek-hi dii

but customer-erg all-of prices same gave

'Although that weaver wove one big sweater quickly, the customer paid the same for all of them.'

Previous ERP studies of English have shown that locally cued violations involving tense/aspect morphology elicit a P600 response (e.g. **was eat*, Allen, Badecker, and Osterhout 2003), and there have been numerous studies that have contrasted N400 and P600 responses associated with semantic and syntactic violations, respectively (e.g., Friederici et al. 1993; Hagoort 2003; Neville et al. 1991). What is unusual about the Hindi study is the fact that the content of the violation is identical in both conditions, and that other potentially interfering factors are matched (e.g., distance from predictor to predicted, number of referents). Thus, the results suggest that Hindi speakers encode not only what they are predicting as they process a sentence, but also why they are predicting it.

The challenge that the Hindi results present is similar to the challenge provided by the example from Japanese *wh*-fronting. They show that a model that accurately captures the surface statistics of Hindi sentences will still fall short as a model of how speakers process sentences. On the other hand, a model that incorporates more abstract representations, and that can therefore encode the causes of structural dependencies, may be able to attain far greater success using simpler statistics.[iii]

4 Conclusion

Any model of language that aspires to account for human abilities, rather than simply to solve an engineering task, needs to take seriously the full range of human language abilities, many of which are not apparent in everyday speaking and writing. Although corpora can provide a reasonable first approximation to the training data available to the human learner, the ability to parse similar corpora is a rather poor criterion for a model of the end-state of language learning.

There are a number of reasons why the full complexities of human language may tend to be overlooked in current statistical models. In part, this may reflect the attraction of the objectivity of corpus-based benchmarks, such as the Penn Treebank, which allow for simple quantitative scoring of competing models. Another factor may be that there is skepticism about the reality of the complex phenomena that linguists describe. Finally, part of the problem may be that linguists and psycholinguists tend to focus on knowledge of *constraints,* which describe what a speaker cannot do. It is easy to assess whether a model successfully parses an individual sentence, but much harder to assess whether a model captures the effects of a constraint, since this requires detailed analysis of what structures the model systematically avoids building. However, none of these can justify ignoring the richness of natural language syntax. The complexities described by linguists are very real, and speakers' knowledge of constraints on language involves much more than just surface co-occurrence patterns, as the studies described above show.

In presenting the challenges offered by complexity, consistency, and causality, my goal is not to argue that statistics play no role in models of language learning, nor is it my goal to argue against the use of objective benchmarks for testing of models. The point is that the objective benchmarks that are used need to more accurately represent the scope and complexity of human language abilities, and that models are unlikely to succeed on stronger tests of this kind unless their designers are willing to incorporate richer representations.

Acknowledgments

This article is based upon a talk originally presented at the workshop on *Syntax, Semantics, and Statistics* at the Neural Information Processing Systems meeting in Whistler, BC, in December 2003, and I thank the editors of this volume for allowing me to include it here. This work was supported in part by grants from NSF, #BCS-0196004, the Human Frontier Science Program, #RGY-0134, and the McDonnell-Pew Cognitive Neuroscience Program, #99-31, and the University of Maryland General Research Board. Thanks to Sachiko Aoshima, Nina Kazanina, Meesook Kim, Barbara Landau, Ellen Lau, Moti Lieberman, Andrew Nevins, Beth Rabbin, Amy Weinberg, and Masaya Yoshida for their major contributions to the studies described here, and thanks to Raffaella Zanuttini and Héctor Campos for comments on an earlier draft.

References

Allen, M., Badecker, W., & Osterhout, L. 2003. Morphological analysis in sentence processing: An ERP study. *Language and Cognitive Processes* 18: 405-430.

Allen, J. & Seidenberg, M. 1999. The emergence of grammaticality in connectionist networks. In B. MacWhinney (ed.), *Emergentist approaches to language*, pp. 115-151 Mahwah, NJ: Erlbaum.

Aoshima, S., Phillips, C., & Weinberg, A. 2004. Processing filler-gap dependencies in a head-final language. *Journal of Memory and Language* 51: 23-54.

Avrutin, S. & Reuland, E. 2002. Backward anaphora and tense interpretation. Ms., Utrecht University.

Baker, M. C. 1991. On some subject-object non-asymmetries in Mohawk. *Natural Language and Linguistic Theory* 9: 537-76.

Baker, M. C. 2001. *The atoms of language*. New York: Basic Books.

Chomsky, N. 1981. *Lectures on government and binding*. Dordrecht: Foris.

Crain, S., & Fodor, J. D. 1985. How can grammars help parsers? In D. Dowty, L. Kartunnen, and A. M. Zwicky (Eds.), *Natural language parsing: Psycholinguistic, computational, and theoretical perspectives*, pp. 94-128. Cambridge, UK: Cambridge University Press.

Crain, S., & McKee, C. 1985. The acquisition of structural restrictions on anaphora. In S. Berman, J. Choe, & J. McDonough (eds.), *Proceedings of NELS15*, pp. 94-110. Amherst, MA: GLSA Publications.

Evans, G. 1982. *The varieties of reference*. Oxford, UK: Oxford University Press.

Frazier, L., & Clifton, C. Jr. 1989. Successive cyclicity in the grammar and the parser. *Language and Cognitive Processes* 4: 93-126.

Friederici, A. D., Pfeifer, E., & Hahne, A. 1993. Event-related brain potentials during natural speech processing: Effects of semantic, morphological, and syntactic violations. *Cognitive Brain Research* 1:183-192.

Gibson, E., & Hickok, G. 1993. Sentence processing with empty categories. *Language and Cognitive Processes* 8: 147-161.

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. 1999. Human simulations of vocabulary learning. *Cognition* 73: 135-176.

Gropen, J., Pinker, S., Hollander, M., Goldberg, R. 1991. Syntax and Semantics in the acquisition of locative verbs. *Journal of Child Language* 18: 115-151.

Haegeman, L. 1994. *Introduction to government and binding theory (2nd edn.)*. Oxford, UK: Blackwell.

Hagoort, P. 2003. Interplay between syntax and semantics during sentence comprehension: ERP effects of combining syntactic and semantic violations. *Journal of Cognitive Neuroscience* 15: 883-899.

Hagoort, P., Brown, C. M., & Osterhout, L. 1999. The neural architecture of syntactic processing. In C. M. Brown & P. Hagoort (eds.), *Neurocognition of language*, pp. 273-316. Oxford: Oxford University Press.

Kazanina, N., Lau, E., Lieberman, M., Phillips, C., & Yoshida, M. 2004. Use of grammatical constraints in the processing of backward anaphora. Ms., in preparation.

Kazanina, N. & Phillips, C. 2001. Coreference in child Russian: Distinguishing syntactic and discourse constraints. In A. Do, L. Dominguez, and A. Johansen (Eds.), *Proceedings of BUCLD 25*, pp. 413-424. Cascadilla Press.

Kim, M. 1999. *A cross-linguistic perspective on the acquisition of locative verbs*. Ph.D. dissertation, University of Delaware.

Kim, M., Landau, B., & Phillips, C. 1999. Cross-linguistic differences in children's syntax for locative verbs. In A. Greenhill, H. Littlefield, and C. Tano (Eds.), *Proceedings of BUCLD 23*, pp. 337-348. Somerville, MA: Cascadilla Press.

Kutas, M. & Federmeier, K. D. 2000. Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences* 4: 463-470.

Kutas, M. & Hillyard, S. A. 1980. Reading senseless sentences: Brain potentials reflect semantic anomaly. *Science* 207: 203-205.

Levin, B. 1993. *English verb classes and alternations*. Chicago: U. of Chicago Press.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. 1994. The lexical nature of syntactic ambiguity resolution. *Psychological Review* 89: 483-506.

Manning, C. D. & Schütze, H. 1999. *Foundations of statistical natural language processing*. Cambridge, MA: MIT Press.

Miyamoto, E. T., & Takahashi, S. 2003. Typing mismatch effects in the processing of wh-phrases in Japanese. Manuscript submitted for publication.

Neville, H., Nicol, J., Barss, A., Forster, K. I., & Garrett, M. I. 1991. Syntactically-based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience* 3: 151-165.

Nevins, A., Phillips, C., & Poeppel. D. 2004. Syntactic and semantic predictors of tense in Hindi: An ERP study. Talk presented at the 17th Annual CUNY Sentence Processing Conference, College Park, MD.

Newport, E. L. & Aslin, R. N. 2004. Learning at a distance I: Statistical learning of non-adjacent dependencies. *Cognitive Psychology* 48: 127-162.

Pena, M., Bonatti, L. L., Nespor, M., & Mehler, J. 2002. Signal-driven computations in speech processing. *Science* 298: 604-607.

Phillips, C., Rabbin, B., & Kim, M. 2001. Principles and parameters of locative verb syntax. Talk presented at workshop on verb classes, University of Maryland.

Pickering, M. J., & Barry, G. D. 1991. Sentence processing without empty categories. *Language and Cognitive Processes* 6: 229-259.

Pinker, S. 1989. *Learnability and cognition: The acquisition of argument structure*. Cambridge, MA: MIT Press.

Pritchett, B. L. 1992. *Grammatical competence and parsing performance.* Chicago, IL: The University of Chicago Press.

Reinhart, T. 1983. *Anaphora and semantic interpretation*. London: Croom Helm.

Saffran, J. R., Aslin, R. N., & Newport, E. L. 1996. Statistical learning by 8-month old infants. *Science* 274: 1926-1928.

Sag, I. A., & Fodor, J. D. 1994. Extraction without traces. In R. Aranovich, W. Byrne, S. Preuss, and M. Senturia (eds.), *Proceedings of the 13th Annual Meeting of the West Coast Conference on Formal Linguistics,* pp. 365-384. Stanford, CA: CSLI.

Seidenberg, M. 1997. Language acquisition and use: Learning and applying probabilistic constraints. *Science* 275: 1599-1603.

Stowe, L. A. 1986. Parsing *WH*-constructions: evidence for on-line gap location. *Language and Cognitive Processes* 1: 227-245.

van Gompel, R. P. G. & Liversedge, S. P. 2003. The influence of morphological information on cataphoric pronoun assignment. *Journal of Experimental Psychology: Learning, Memory, & Cognition* 29: 128-139.

Yoshida, M., Aoshima, S., & Phillips, C. 2004. Relative clause prediction in Japanese. Ms., in preparation.

[i] There are examples that appear to violate this generalization, e.g., *He then did what John always did in such situations….* Discussion of such cases goes back more than 20 years (Evans 1982; Reinhart 1983), and it is generally understood that they involve situations where multiple mental representations of 'guises' of the same individual are being compared, indicating that the constraint really applies to reference in a mental model, and not to objects in the world. Minimally different examples still disallow coreference, e.g. **He did what John had done half an hour earlier….*

[ii] Some languages do allow the equivalent of (4b), e.g., German and Russian. However, when they allow this, they use different verb roots for (4a) and (4b), e.g., German uses *giessen* for 'pour' in (4a), but *begiessen* in (4b).

[iii] Note that there have been attempts to capture the difference between syntactic ungrammaticality and semantic anomaly in the literature on simple recurrent backpropagation networks (e.g., Allen and Seidenberg 1999). However, such efforts rely on the assumption that syntactic and semantic violations will have different contents, an assumption that is contradicted by the Hindi examples.