Language comprehension is sensitive to changes in the reliability of lexical cues

Alex B. Fine (afine@bcs.rochester.edu)
Department of Brain and Cognitive Sciences and Department of Linguistics, University of Rochester
Rochester, NY 14627 USA

T. Florian Jaeger (fjaeger@bcs.rochester.edu)
Department of Brain and Cognitive Sciences and Department of Computer Science, University of Rochester
Rochester, NY 14627 USA

Abstract

This paper tests the hypothesis that language comprehenders update their beliefs about the statistics of their language throughout the lifespan, and that this belief update allows comprehenders to combine probabilistic linguistic cues according to their reliability. We conduct a multi-day sentence comprehension study in which the reliability of a probabilistic cue to syntactic structure is manipulated between subjects. We find that as the reliability of one cue to syntactic structure decreases, comprehenders come to rely more on a second cue to syntactic structure. The results are consonant with rational models of cue integration in speech perception and in non-linguistic domains, thus suggesting a unifying computational principle governing the way humans use information across both perceptual and higher-level cognitive tasks.

Keywords: psycholinguistics; adaptation; sentence processing; cue combination

Introduction

In order to process language, humans must make *inferences* about intended messages in the face of uncertainty arising from noisy perceptual data and ambiguity inherent in the linguistic signal. Research in psycholinguistics suggests that humans accomplish this task partially by capitalizing on probabilistic cues in the linguistic as well as the non-linguistic context (Jurafsky, 1996; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). For instance, in sentences such as (1), *the judge* is temporarily interpretable as both the direct object of *acknowledged* and the subject of an embedded sentence complement. By-word reading times at the point at which the sentence is disambiguated (*had been*) are correlated with the conditional probability of the structural representation assigned to the incremental parse given a number of probabilistic cues. One such cue is the verb—the probability of a sentence complement occurring is greater given *assert* than *acknowledge*, based on corpus statistics and norming data (e.g., Trueswell, Tanenhaus, & Kello, 1993; MacDonald, Pearlmutter, & Seidenberg, 1994). A second cue is the post-verbal noun phrase—if a post-verbal noun phrase is unlikely to be a direct object of the verb, it is more likely to be the subject of an embedded clause, thereby increasing the probability of a sentence complement continuation (Garnsey, Pearlmutter, Myers, & Lotocky, 1997). Finally, if present, the complementizer *that* (e.g., *The lawyer acknowledged* ***that*** *the judge had been* . . .) also serves as a strong cue to syntactic structure. Indeed, comprehenders have been shown to rely on all of these cues during the incremental processing of sentences such as (1) (MacDonald et al., 1994).

(1) The lawyer acknowledged the judge had been unfair to the defendant.

Probabilistic cues provide comprehenders with information that can guide inferences during incremental language processing, contributing to *processing efficiency* (see Smith & Levy, 2008 for an explicit proposal along these lines). However, the cues relevant to comprehension are moving targets: probabilistic cues such as those mentioned above are context-dependent in that their *validity* (Bates & MacWhinney, 1987) may vary depending on speaker identity, context, and speaker dialect (see Tagliamonte, 2005 for a discussion of variability in syntax). Bates and MacWhinney (1987) define cue validity as the product of *cue availability* (how often a cue is present in the environment) and *cue reliability* (how often a cue leads to the correct inference, when present). How do comprehenders cope with this variability and maximize the usefulness of probabilistic cues? The current study addresses this question and tests a two-pronged hypothesis, framed in the spirit of *rational analysis* (Anderson, 1990):

- **A: Lifelong implicit learning**: Throughout adulthood, humans continuously update and adjust estimates of probabilistic cues relevant to language comprehension. We will refer to the results of this process as *adaptation* (cf. also Chang, Dell, & Bock, 2006 and references therein).
- **B: Rational linguistic adaptation**: Adaptation is *rational* in the sense that humans update the weight they assign to a particular cue based on changes in the *validity* of that cue.

Preliminary evidence for (A) comes from language comprehension studies at multiple levels of representation (at the phonetic level: Clayards, Tanenhaus, Aslin, & Jacobs, 2008; Kraljic & Samuel, 2007; at the syntactic level: Fine, Qian, Jaeger, & Jacobs, 2010; Wells, Christiansen, Race, Acheson, & MacDonald, 2009). Preliminary evidence for (B) comes primarily from speech perception (Clayards et al., 2008; Kraljic, Samuel, & Brennan, 2008), though these studies are not necessarily framed in terms of the hypotheses presented above. Of particular relevance is Clayards et al. (2008), who manipulated participants' experience with voice-onset time (VOT), a probabilistic cue to phonetic category membership. For participants in one group, the distribution over VOT values that emerged over the course of the experiment had a low variance; for participants in the other group, this distribution had a high variance. The rationale of the manipulation is that the *reliability* of a probabilistic cue can be quantified as the inverse of the variance of the distribution over values that the cue can take. Generally speaking, rational models of perception predict that cues should be weighted according to their reliability—the lower the reliability of a probabilistic cue, the less subjects should rely on that cue (e.g., Ernst & Banks, 2002).

Consistent with the predictions of rational models, participants in the low reliability (i.e. high variance) group displayed less certainty than subjects in the high reliability (i.e. low-variance) group in a categorization task that required making inferences based on VOT. In short, participants relied on a probabilistic cue to the extent that that cue was reliable.

It remains an open question whether the computational principles underlying participants' behavior in the Clayards et al. (2008) study pervade all levels of language processing, or whether these principles are limited to the interface between perceptual processes and linguistic categories, as in speech perception.

The goal of the present study is therefore to test the hypotheses that belief update and the rational weighting of probabilistic cues are general computational principles of language processing by asking whether these principles are operative in *sentence* comprehension. Following the terminology from Bates and MacWhinney (1987) introduced above, we focus on the effect of *cue reliability* in syntactic comprehension.

Methods

To explore these questions, we conducted a multi-day sentence processing study in which the reliability of a probabilistic cue to syntactic structure was manipulated by providing participants with experimentally controlled experience with that cue in a between-subjects design, following the logic of the experiment reported in Clayards et al. (2008).

Specifically, in sentences like (1), reading times at the point of disambiguation (*had been*) are sensitive to at least two cues. First, we consider the presence or absence of the complementizer *that* (e.g. *The lawyer acknowledged **that** the judge had been unfair to the defendant*), which, when present, disambiguates the post-verbal NP (*the judge*) as the subject of a sentence complement. However, because the complementizer cue is not always *available*, it is not a perfectly *valid* cue. Moreover, it is important to consider that linguistic perception is *noisy* (due to environmental noise, noise within the nervous system, etc.) such that there is uncertainty about what words have been perceived (for evidence, see Levy, Bicknell, Slattery, & Rayner, 2009). This suggests that even when the complementizer is present in a speaker's output it is not a perfect cue to syntactic structure due to the uncertainty as to whether it has been perceived. Second, we consider the verb itself (*acknowledged* in (1)) a probabilistic cue to syntactic structure. The verb contains information about the probability of different argument types (and hence different syntactic structures) following it. In that sense, the verb is a probabilistic cue that comprehenders can employ to make inferences about the incremental parse. Verbs such as *acknowledge*, *regret*, *confess*, etc. can take either a sentence complement (SC), as in (1), or a direct object (DO), as in *The woman acknowledged her own shortcomings*. Each of these argument types occurs with some probability, which can be estimated based on corpus statistics or norming studies. In the current study, we provide participants with experimentally controlled exposure to such so-called DO/SC verbs to directly manipulate participants' estimates of the reliability of the verb as a probabilistic cue in order to see whether this shifts how participants weigh each cue during parsing.

Participants were assigned to one of two groups. In both groups, participants read sentences containing DO/SC verbs over the course of three non-consecutive days. In one group, all verbs took SCs. In the other group, verbs occurred 50% of the time with DO arguments and 50% of the time with SC arguments. Because the variance of a binomial distribution is *minimized* when one event occurs with a probability of 1 and the other with a probability of 0 and *maximized* when both events are equally likely, we refer to the two groups as the *high reliability group* and the *low reliability group*, respectively. This between-participants manipulation is visualized in Figure (1).

The key prediction is that, as the variance of the distribution over argument types for the verbs increases (i.e., as the variance of $p(SC|v)$ increases), the reliability of the verb as a cue to syntactic structure decreases. Participants should therefore rely *more* on a second cue to syntactic structure—here, the complementizer *that*, since the reliability of that cue remains constant across groups—as the reliability of the verb cue decreases. In other words, the complementizer becomes a better cue to syntactic structure than the verb for participants in the low reliability group.

A tradeoff between multiple cues like the one predicted here has been demonstrated in vision (Knill & Saunders, 2003) and visual-haptic tasks (Ernst & Banks, 2002), to give just two examples.

Procedure

80 participants visited the lab on five non-consecutive days. Each visit took place no sooner than 48 hours after the previous one. The structure and time-course of the exposure phase in this experiment were closely modeled on that of Wells et al. (2009). Similar to their experiment, ours consisted of a pre-exposure self-paced reading task on the first visit, a second, post-exposure self-paced reading task on the fifth visit which was identical to the first, and three intervening visits that comprised the exposure phase of the study.

The procedure at each of these five visits is outlined below, and the overall experimental regimen is visualized in Figure (2), where each box corresponds to a different day in the experiment.

Visit 1: Pre-training self-paced reading task Participants were randomly assigned to either the low reliability or high

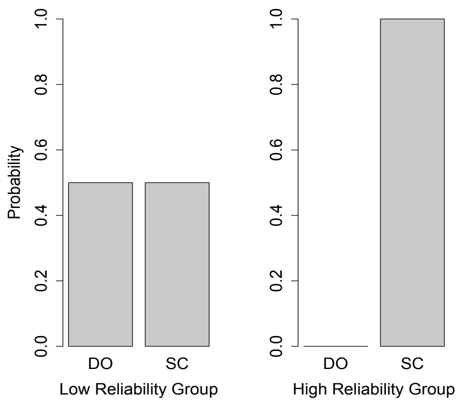

Figure 1: **Between-participant manipulation**: Distribution of DO vs. SC continuations that participants in the Low Reliability (left) and High Reliability (right) were exposed to during visit 2-4 (cf. Figure 2).

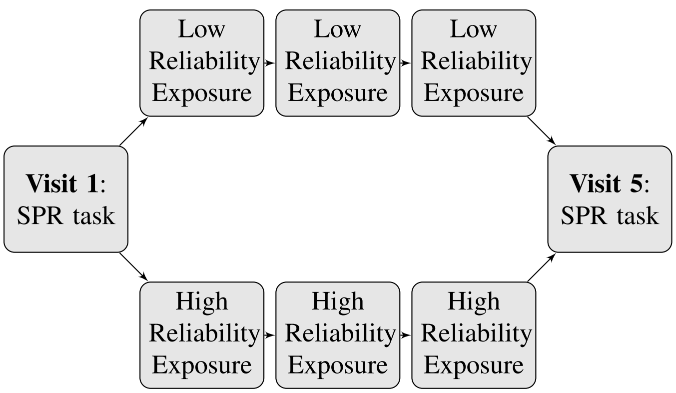

Figure 2: Schematic illustration of the exposure phase used in the experiment

reliability group. During the first visit, participants in both groups completed the same self-paced reading task. The materials for the self-paced reading task comprised a subset of those used in Garnsey et al. (1997). Participants read 36 target sentences containing DO/SC verbs, as well as 72 filler sentences containing a variety of syntactic structures.

To maximize the temporary ambiguity effect, the DO/SC verb was always followed by an NP that made a plausible DO continuation (e.g. *The talented photographer accepted the money could not be spent yet*). Target sentences occurred in one of 2 (temporarily ambiguous vs. not) x 3 (verb bias) conditions. In the self-paced reading task, all critical items contained SC continuations—that is, all verbs that *could* take a SC continuation did.

In the unambiguous condition, the complementizer *that* was present, as in (2a). In the ambiguous version, the complementizer *that* was absent, as in (2b), where the temporarily ambiguous NP (*the money*) and disambiguating region (*could not*) are underlined.

(2) The talented photographer accepted . . .
  a. . . . that <u>the money could not</u> be spent yet.
  b. . . . <u>the money could not</u> be spent yet.

Verb-bias was manipulated between items. Based on norming data from Garnsey et al. (1997), 12 target verbs were classified as SC-biased, 12 as EQ-biased, and 12 as DO-biased.

The goal of the first visit was to provide an initial, baseline measure of the effects of prior verb bias and ambiguity (complementizer presence/absence) on participants in each group, to which post-exposure self-paced reading times could be compared to assess the effect of exposure. Specifically, we expect the change in reading times from pre-exposure to post-exposure during the ambiguous and disambiguating regions (e.g., . . . *the money could not* . . .) to reflect the group manipulation.

Visits 2-4: Exposure Beginning with the second visit to the lab, participants received experimentally controlled exposure to DO/SC verbs. Of the 36 verbs included in the self-paced reading task from visit 1, 16 of these were included in the exposure phase. Of these, 8 were classified as SC-biased and 8 were DO-biased. The purpose of including only a subset of the verbs in the exposure phase was to assess the lexical-specificity of the effect of exposure, discussed below.

At each visit in the exposure phase, participants read a total of 64 sentences containing DO/SC verbs, with each of the 16 verbs appearing 4 times at each visit. In addition to these 64 sentences, participants read 64 filler sentences, randomly interspersed between critical sentences. Filler sentences contained a variety of syntactic structures, but none contained DO/SC verbs. Across the exposure phase visits, all DO/SC sentences and all fillers were unique (participants never read the same sentence twice). Moreover, the sentences containing DO/SC verbs included in the exposure varied in length and semantic content. Sentences were presented in block form (i.e. the entire sentence appeared on the screen, and participants pressed the space bar when they were done reading the sentence).

Participants assigned to both the low reliability and the high reliability group received exposure to the same 16 verbs, saw these 16 verbs an equal number of times, and read the same fillers. The crucial difference between the two groups' exposure lists was the proportion of sentences containing DO/SC verbs that involved DO continuations (as in (3)) versus SC continuations (as in (4)). For participants in the high reliability group, all sentences containing DO/SC verbs occurred with SCs. For participants in the low reliability group, DO/SC verbs occurred 50% of the time with DOs and 50% of the time with SCs.

(3) The lawyer acknowledged [$_{SC}$ the judge had been lying].

(4) The lawyer acknowledged [$_{DO}$ the judge in the red sweater].

For both groups, half of all SC sentences included the complementizer *that*.

Visit 5: Post-exposure self-paced reading task Participants in both groups returned to the lab and performed the exact same self-paced reading task they performed during visit 1. Additionally, each subject saw the same experimental list they saw during visit 1 (i.e., saw the same items in the same conditions), in order to make pre- and post-training reading times maximally comparable.

By hypothesis, then, for participants in the high reliability group, an estimate of $p(SC|v_i)$—i.e. the conditional probability of the SC structure given a particular verb, v_i, included in the exposure—that reflects the context-specific statistics of the input is $p(SC|v_i) = 1$, and for participants in the low reliability group, $p(SC|v_i) = .5$. Crucially, for both groups, $p(SC|that)$ is the same. Thus, the informativity or reliability of the verb cue differs between the groups, while the reliability of *that* remains identical for the two groups.

What would count as a *rational* estimate of $p(SC|v)$, if the goal of adaptation is efficient processing (as hypothesized in (B)), depends on a variety of as yet unknown factors: how much variability there is between speakers at the syntactic level, for example. Also, in our experiment, it is unknown whether speakers consider the visits to be all generated by the same "speaker" or at least a sufficiently consistent and stable "situation" that adaptation can be considered a rational strategy.

Even though these questions remain unanswered and we therefore do not know the rational estimate of $p(SC|v)$ for visit 5, we can still say that the verb will be a more reliable cue in the high reliability group compared to the low reliability group, and that the reliability of the complementizer will be higher than that of the verb cue in the low reliability group.

In sum, then, if participants in our experiment are updating their representations of probabilistic cues to syntactic structure in order to reflect the statistics of the (possibly experiment-specific) input, and are subsequently weighting these cues according to their reliability, participants in the low reliability group should rely more on the complementizer as a cue during the post-exposure self-paced reading task than in the pre-exposure task. These participants should also rely on this cue more relative to participants in the high reliability group.

Results and Discussion

First, we computed length-corrected reading times by regressing raw word-by-word reading times for all words in both critical and filler items onto word length. The residuals of this model then served as the dependent measure for all analyses reported. We analyzed length-corrected (residual) reading times for words in the critical region (underlined in (2a)-(2b)) across both pre- and post-exposure. We examine the effect of exposure for items which contained a verb included in the exposure, as well as the extent to which the effect of training was *lexically specific* by comparing the effect of exposure on verbs included in the exposure phase vs. those not included.

Effect on verbs in exposure

To test the hypotheses outlined in the introduction, length-corrected reading times at the critical region were regressed onto the full factorial design (i.e. all main effects and interactions) of time (pre- vs. post-training), group (high reliability vs. low reliability), ambiguity (complementizer present vs. absent), and SC-bias (SC- vs. DO-biased). The data were analyzed using linear mixed effects regression, with the maximum random effects structure justified by the data based on model comparison.[1]

Regardless of which group participants were assigned to, there was an overall speedup in reading times from pre- to post-exposure—i.e., a main effect of time ($\beta = -42, SE = 3.9, p < .001$). There was also a significant main effect of ambiguity, such that reading times were lower for unambiguous sentences (sentences with the complementizer *that*) than for ambiguous sentences (sentences without the complementizer). This interacted with time: the processing advantage conferred by the presence of the complementizer *that* was greater during the pre-exposure self-paced reading task than in the post-exposure self-paced reading task ($\beta = 3.8, SE = .92, p < .05$).

A significant ambiguity by SC-bias interaction was found, suggesting that the processing advantage conferred by the complementizer was diminished as the *a priori* bias of the verb to take SCs increased ($\beta = 2.3, SE = .89, p < .05$). This replicates previous studies using similar materials (e.g. Garnsey et al., 1997; Trueswell et al., 1993). Again, this effect interacted with time: the tradeoff between ambiguity and prior verb bias was diminished during the post-exposure self-paced reading task compared to the pre-exposure task ($\beta = -2.24, SE = .89, p < .05$).

Most crucially for the hypothesis that linguistic adaptation serves the purpose of allowing efficient communication, there was a significant time by group by ambiguity interaction ($\beta = 2.4, SE = .9, p < .05$): the degree to which the complementizer was exploited by participants changed over time, but more importantly, the nature of this change depended on the group's experience during the exposure phase.

To facilitate visualization of the interaction, we computed a difference score by subtracting post-exposure length-corrected RTs from pre-exposure length-corrected RTs in the critical region. Thus, a large change score means a large decrease from pre- to post-exposure reading times. As shown in Figure (3), participants in the high reliability group showed a greater decrease in reading times for ambiguous sentences than participants in the low reliability group, reflecting the relatively high degree of certainty for participants in the high reliability group that DO/SC verbs would take SCs, based on the statistics of the exposure phase.

[1] 4-way ANOVAs yield the same results as those reported below. The results do not depend on the particular statistical analysis performed.

Furthermore, the decrease in reading times for participants in the low reliability group was actually *greater* for sentences with the complementizer than for sentences without; and this decrease was greater for participants in the low reliability group than in the high reliability group. This pattern shows that participants in the low reliability group came to rely on the complementizer as a probabilistic cue *more* than the high reliability group, providing support for hypothesis (B) outlined in the introduction.

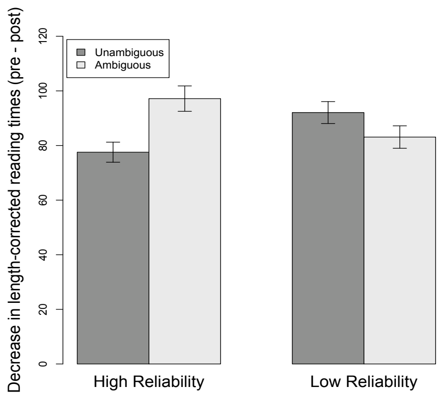

Figure 3: Group by Ambiguity interaction

Lexical Specificity of Exposure Effect

We also tested whether the effect of training was modulated by whether the verb for the item being read was included in the exposure phase (recall that only a subset of the 36 verbs in the pre- and post-exposure self-paced reading task were included in the exposure phase). Because all verbs included in the exposure were either SC- or DO-biased, models including both a term for exposure (in exposure phase vs. not) and for SC-bias did not converge, due to collinearity between these two predictors. Therefore, to test the lexical specificity of the exposure effect, we regressed length-corrected reading times at the critical region onto the full factorial design of time (pre- vs. post-exposure), exposure (in vs. not), ambiguity (temporarily ambiguous vs. unambiguous), and group (high vs. low reliability). The model included the maximum random effect structure justified by the data based on model comparison. Again, there was a main effect of time ($\beta = -38.5, SE = 4.3, p < .001$), and a main effect of ambiguity ($\beta = -7.8, SE = 1.1, p < .001$) as well as an interaction between these two predictors ($\beta = 4, SE = .8, p < .01$). All of these effects went in the same direction as in the previous analysis.

Most notably, the three-way time by ambiguity by group interaction reported above interacted with training: specifically, the differential weighting of the complementizer across the two groups *only held for experimental items containing verbs that appeared in the training phase* ($\beta = 1.8, SE = .9, p < .05$). For items with verbs not included in training, the effect of training was of a similar character across both groups, and there was no significant time by ambiguity by group interaction ($p > .1$).

Conclusion

The results reported here provide support for two related claims. First, the results support a view of language comprehension in which humans continuously update their estimates of the statistics of the language they speak via implicit learning. This extends previous work in speech perception to higher level aspects of language processing (Clayards et al., 2008; Kraljic & Samuel, 2007; Vroomen, Linden, Gelder, & Bertelson, 2007).

Together, these results support the hypothesis of lifelong learning (A), which is a central assumption of many connectionist accounts (Chang et al., 2006; Elman, 1990; Juola, 1999). This assumption is also supported by recent work suggesting effects of recent experience that go beyond short-term boosts in activation associated with the most recently processed relevant linguistic stimulus (e.g., in production: Kaschak, 2007; Snider & Jaeger, submitted; in comprehension: Fine et al., 2010; Wells et al., 2009). Since at least 10 days elapsed between the first and last visit, our results highlight the longevity of the effect of exposure, suggesting that the effect is not the result of short-term syntactic priming.

Second, our results suggest a possible *explanation* for lifelong linguistic adaptation: by maintaining accurate estimates of the statistics of the ambient language, comprehenders can exploit probabilistic linguistic cues in a way that maximizes the utility of these cues. In our experiment, as the reliability of one probabilistic cue to syntactic structure (the verb) decreased, participants came to depend more on a second cue (the complementizer *that*). Thus participants in our experiment showed a behavioral pattern consistent with rational models of cue combination (Ernst & Banks, 2002; Knill & Saunders, 2003).

The *lexical specificity* of the effect of exposure is noteworthy as well. Recall that the tradeoff between the verb and complementizer cues was observed only for items containing verbs included in the exposure phase, suggesting that participants in the experiment tracked very fine-grained statistics about the reliability of the verb as a cue to syntactic structure.

The results reported here thus go beyond previous work in important respects. Within sentence comprehension, in addition to finding that the use of probabilistic cues during sentence comprehension is sensitive to experience (supporting the claims made by, e.g., Wells et al., 2009), we find that the way in which *multiple cues* are used during sentence comprehension is guided by very specific details of the statistical structure of that experience.

Moreover, to the extent that subjects' behavior in our experiment depended on the *variance* of a cue to syntactic structure (i.e., on the variance of $p(SC|v)$), the results suggest that sentence comprehension may be guided by knowledge of entire probability *distributions*, rather than simple point estimates of those distributions, as is typically assumed (explicitly or implicitly) by previous work focusing on the role of probabilistic cues during sentence comprehension (e.g., Trueswell et al., 1993).

Beyond the domain of sentence comprehension, if humans rationally integrate cues to syntactic structure, this would suggest that the same computational principle governing cue combination demonstrated in speech perception (Bejjanki, Clayards, Knill, & Aslin, accepted; Clayards et al., 2008; Toscano & McMurray, 2010) and in non-linguistic domains (Ernst & Banks, 2002; Knill & Saunders, 2003) is at work in higher-level language processing, thus suggesting a *unifying computational principle* governing the way humans use information across both perceptual and higher-level cognitive tasks.

Acknowledgments

We wish to thank Jeremy Ferris for his dedicated and tireless efforts with subject recruitment and data collection, and Camber Hansen-Karr for comments on an earlier draft. We also thank five anonymous CogSci reviewers. This work was partially supported by NSF Graduate Research Fellowship to AF and NSF Grant BCS-0844472 and Alfred P. Sloan Research Fellowship to TFJ.

References

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum.

Bates, E., & MacWhinney, B. (1987). Mechanisms of language acquisition. In B. MacWhinney (Ed.), (chap. Competition, variation, and language learning). Hillsdale, NJ: Lawrence Erlbaum.

Bejjanki, V. R., Clayards, M., Knill, D. C., & Aslin, R. N. (accepted). Cue integration in categorical tasks: insights from audio-visual speech perception. *PLoS ONE*.

Chang, F., Dell, G., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, *113*(2), 234-272.

Clayards, M., Tanenhaus, M., Aslin, R., & Jacobs, R. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, *108*(3), 804-809.

Elman, J. (1990). Finding structure in time. *Cognitive Science*, *14*, 179-211.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429-433.

Fine, A. B., Qian, T., Jaeger, T. F., & Jacobs, R. (2010, July). Syntactic adaptation in language comprehension. In *Proceedings of the 2010 workshop on cognitive modeling and computational linguistics* (pp. 18–26). Uppsala, Sweden: Association for Computational Linguistics.

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

Juola, K. P. . P. (1999). A connectionist model of english past tense and plural morphology. *Cognitive Science*, *23*(4), 463-490.

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

Kaschak, M. (2007). Long-term structure priming affects subsequent patterns of language production. *Memory and Cognition*, *35*, 925-937.

Knill, D. C., & Saunders, J. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research*, *43*(24), 2539-2558.

Kraljic, T., & Samuel, A. (2007). Perceptual adjustments to multiple speakers. *Journal of Memory and Language*, *56*, 1-15.

Kraljic, T., Samuel, A., & Brennan, S. (2008). First impressions and last resorts: How listeners adjust to speaker variability. *Psychological Science*, *19*(4), 332-338.

Levy, R., Bicknell, K., Slattery, T., & Rayner, K. (2009). Eye movement evidence that readers maintain and act on uncertainty about past linguistic input. *Proceedings of the National Academy of Sciences*, *106*(50), 21086-21090.

MacDonald, M., Pearlmutter, N., & Seidenberg, M. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*, 676-703.

Smith, N. J., & Levy, R. (2008). Optimal processing times in reading: A formal model and empirical investigation. In *Proceedings of the 30th annual meeting of the cognitive science society*.

Snider, N., & Jaeger, T. F. (submitted). Syntax in flux: Structural priming maintains probabilistic representations.

Tagliamonte, S. (2005). No momentary fancy! the zero in english dialects. *English Language and Linguistics*(2), 289-309.

Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632-1634.

Toscano, J. C., & McMurray, B. (2010). Cue integration with categories: Weighting acoustic cues in speech using unsupervised learning and distributional statistics. *Cognitive Science*, 434-464.

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *19*(3), 528-553.

Vroomen, J., Linden, S. van, Gelder, B. de, & Bertelson, P. (2007). Visual recalibration and selective adaptation in auditory-visual speech perception: Constrasting build-up courses. *Neuropsychologia*, *45*, 572-577.

Wells, J., Christiansen, M., Race, D., Acheson, D., & MacDonald, M. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology*, *58*(2), 250-271.