A belief-updating model of adaptation and cue combination in syntactic comprehension

Dave F. Kleinschmidt[1], **Alex B. Fine**[1], and **T. Florian Jaeger**[1,2]
{dkleinschmidt, afine, fjeager} @ bcs.rochester.edu
[1]Department of Brain and Cognitive Sciences and [2]Department of Computer Science,
University of Rochester, Rochester, NY, 14607 USA

Abstract

We develop and evaluate a preliminary belief-updating model which links intermediate-term (i.e., over several days) syntactic adaptation to the joint statistics of syntactic structures and lexical cues to those structures. This model shows how subjects differentially depend on different cues to syntactic structure following changes in the reliability of those cues, as shown by Fine and Jaeger (2011). By relating syntactic adaptation and cue combination to rational inference under uncertainty, this work links learning and adaptation in sentence processing with adaptation in speech perception and non-linguistic domains.

Keywords: sentence processing, adaptation, Bayesian modeling, cue combination, rational analysis

Introduction

Humans must maintain a stable representation of the environment despite the fact that available sensory input changes across time: for example, over the course of a day, we recognize and grasp objects in a variety of lighting conditions; we execute accurate motor commands despite changes in our own motor systems due to fatigue, over-caffeination, etc.; and during linguistic communication, we process rapidly unfolding acoustic information that varies from talker to talker.

Variability within each of these different modalities changes the correlation between cues—whether visual, haptic, or linguistic—and the things in the world we wish to make inferences about *based on* those cues. How do our brains make use of these cues in spite of variability in the environment? One possibility, suggested by research across a number of domains, is that humans deal with variability in the environment by *adapting to* changes in the statistical properties of the environment (for examples from vision, motor planning, and speech perception, see respectively: Blakemore & Campbell, 1969; Koerding, Tenenbaum, & Shadmehr, 2007; Norris, McQueen, & Cutler, 2003).

While most work on adaptation has been concentrated in perception, the question of whether adaptation is operative in higher level cognition has recently received more attention, particularly in language processing research. For instance, a number of researchers have shown that, when given sufficient experience with a structure initially judged to be ungrammatical, listenders come to subsequently comprehend (Luka & Barsalou, 2005), generalize, and even produce (Kaschak & Glenberg, 2004) that structure. Similarly, recent work has shown that we fine-tune our expectations about which syntactic structures are likely to occur in a given context based on recent experience (Thothathiri & Snedeker, 2008; Farmer, Fine, & Jaeger, 2011).

Thus, behavioral evidence seems to suggest that adaptation, qualitatively speaking, is a very general feature of perception and cognition. A question that arises from all of this previous work is whether adaptation observed across all of these domains can be modeled within a single framework. The goal of this paper is to take a step in this direction. In particular, we model adaptation in syntactic comprehension in terms of Bayesian belief update. Modeling syntactic adaptation in a Bayesian framework is appealing because the same basic computational approach has been successfully pursued in a variety of perceptual and motor domains (e.g. Koerding et al., 2007) and, more recently, in speech perception (Kleinschmidt & Jaeger, 2011; Sonderegger & Yu, 2010).

Moreover, Bayesian belief update is ideally suited to explicitly model the fact that syntactic comprehension involves the combination of multiple cues. This offers the advantage of suggesting a single computational framework for adaptation and cue combination, since Bayesian approaches to cue combination have been successful in a number of domains including visually-guided grasping, audio-visual cue combination, and the weighting of cues to phonetic category. In particular, Bayesian approaches to cue combination in perception have provided a formal means of capturing the fact that humans are able to weight multiple cues (e.g., multiple cues to object depth, such as shading and texture) according to how *reliable* those cues are. We return to the relationship between adaptation and cue combination in the discussion.

The goal of the current study is to ask whether a rational model of adaptation–implemented in the form of Bayesian belief update–can account for behavioral evidence for adaptation in the syntactic domain. Here we model behavioral data originally reported in Fine and Jaeger (2011), which concerns how subjects adjust their expectations about different syntactic structures conditioned on lexical information. Specifically, we exploit temporary syntactic ambiguities as a window onto syntactic expectations. In sentences such as (1), the syntactic assignment of the noun phrase *the judge* is temporarily ambiguous, since it can be parsed as either the subject of a sentence complement (SC) clause, as in (1a), or as the the direct object (DO) of *acknowledged* (as in 1b).

(1) The lawyer acknowledged the judge ...

a. ... $\overbrace{\text{had been}}^{\text{disambiguation}}$ unfair to the defendant.

b. ... in the black robe.

The sentence is disambiguated towards the latter reading at *had been*. A great deal of previous work suggests that reading times at *had been* are a function of subjects' *expectations* about which syntactic structure is likely to occur, based on previous cues in the sentence, such as the verb, the combination of the verb and post-verbal noun phrase (e.g., Garnsey, Pearlmutter, Myers, & Lotocky, 1997), and whether or not the complementizer *that* occurs after the verb (e.g., *The lawyer acknowledged* ***that*** *the judge had been unfair to the defendant*). More recent work has explicitly quantified syntactic expectations in probabilistic terms (Hale, 2001; Levy, 2008). In other words, reading times at the disambiguating region (*had been*) provide information about subjects' subjective beliefs about the relative probability of the SC vs. the DO structures: If reading times are high, this indicates that subjects had assigned a relatively *low* probability to the SC structure; if reading times are low, then subjects had likely assigned a relatively *high* probability to the SC structure.

We model changes in reading times at the point of disambiguation as a consequence of syntactic expectation adaptation. Assuming that reading times in the disambiguating region in sentences such as (1a) reflect subjects' beliefs about the relative probabilities of different syntactic structures, we can interpret *changes* in reading times as changes in subjects' beliefs about the distribution of syntactic structures (at least in the context of the experiment, a point to which we return in the discussion). Syntactic adaptation, construed as the incremental adjustment of the subject's representation of a probability distribution over linguistic events, can therefore be naturally modeled in terms of Bayesian belief update.

In the following section, we briefly describe the behavioral data we set out to model. Next we present a Bayesian belief update model of this behavioral data, and assess the quality of the model's fit to the behavioral data. We conclude by summarizing the model's results and providing a discussion of the implications of this modeling work for our understanding of adaptation and cue combination.

Methods and Summary of Behavioral Results

The data we use to test the hypothesis that syntactic adaptation can be understood in terms of incremental Bayesian belief update comes from (Fine & Jaeger, 2011). We briefly describe the design of that experiment.

Experimental Procedure

In a between-subjects, multi-visit self-paced reading experiment, (Fine & Jaeger, 2011) investigated whether comprehenders update their estimates of the probability of the syntactic structures in (1) conditioned on the verb used in the sentence and the presence of the complementizer *that*. The All-SC group received evidence that SC-taking verbs always occur in sentences like (1a), while the 50-50 group was exposed to a 50/50 mix of SC (1a) and DO (1b) structures. For both groups, *that* occurred in 50% of all SC sentences. The experiment consisted of a pre-exposure session, three exposure sessions over at least 6 days, and a final post-exposure session at least 2 days after the last exposure session (cf. Wells,

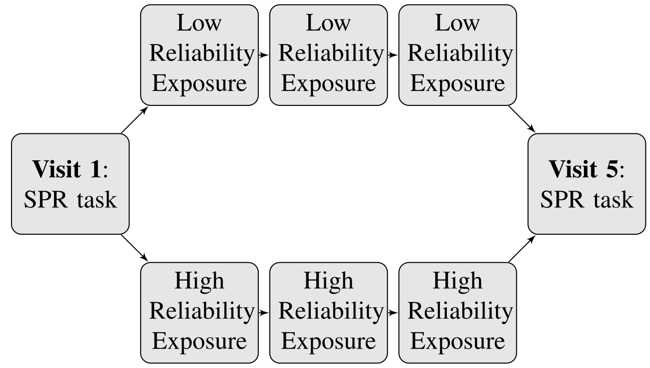

Figure 1: Schematic illustration of experimental procedure

Christiansen, Race, Acheson, & MacDonald, 2009). Subjects in both groups completed identical self-paced reading tasks in the first and final visits. A schematic representation of the experimental procedure is given in Figure 1. Given the design of the experiment, if reading behavior changes from visit 1 to visit 5 differentially across the groups, we can attribute this difference to the distributional properties of the exposure phase within each group.

This experiment allows us to ask two questions. First and foremost, do comprehenders track the distribution of syntactic structures in a given environment? That is, do comprehenders *adapt* to the statistical properties of novel linguistic situations? Second, given the distributional properties of that linguistic environment, do comprehenders *combine* multiple cues to syntactic structure in a way that is rational (i.e., by weighting cues according to their reliability; cf. Bates & MacWhinney, 1987; Anderson, 1990)? Specifically, for subjects in the 50-50 group, all verbs participating in the ambiguity in (1) become—in the experimental context—equally likely to occur with an SC or a DO. Thus the verb becomes, for this group, a very *unreliable* cue to syntactic structure. Qualitatively, according to rational models of cue combination, this should lead subjects in the 50-50 group to depend more on the complementizer *that* in the post-exposure reading task relative to subjects in the All-SC group. That is, the presence of the complementizer should more strongly influence reading times for the 50-50 group. In a regression framework, we therefore predicted a 3-way interaction between time (pre- vs. post-exposure), ambiguity (presence vs. absence of the complementizer *that*), and group (All-SC vs. 50-50). This three-way interaction was significant ($\beta = 2.4, SE = .9, p < .05$), and is visualized in Figure 2 (light bars). This figure shows the decrease in ambiguity effect from pre-test to post-test as a function of training: there is a greater decrease after high-reliability, All-SC training, where subjects do not need to rely on the complementizer as much.

Modeling framework

In constructing a belief updating model of syntactic adaptation and cue combination, there are two main considerations.

First, how can syntactic expectations be quantified, and second, how are those expectations related to the cues present in the linguistic environment and updated based on linguistic experience? In our model, syntactic expectations are quantified as discrete probability distributions over syntactic structures. In this case, the relevant syntactic structures are possible completions of sentences like (1), which we assume are limited to sentence complement completions (SC) and direct object completions, etc. (DO).

Syntactic expectations are related to relevant cues and in turn to linguistic experience via the conditional probability. SC completions are more common for some verbs than others, and are more common when the complementizer *that* is present. This dependence is captured by the conditional probability of $p(S = \mathrm{SC}|T,V)$, where T indicates the presence or absence of *that* and V indicates the identity of the main verb of the sentence. The conditional probability of $S = \mathrm{SC}$ is closely related to the joint probability via the base probability of the various possible combinations of cue values: $p(S,T,V) = p(S|T,V)p(T,V)$.

We can model this joint distribution of syntactic structures and cues via a multinomial distribution. A multinomial distribution assigns a probability to a group of observations, each of which is, in our case, a triplet of the form $S = i, T = j, V = k$, each with a known probability of $p(S = i, T = j, V = k) = \theta_{i,j,k}$. The likelihood of making a group of observations X, where each outcome occurs $n_{i,j,k}$ times, is

$$p(X|\theta) \propto \prod_{i,j,k} \theta_{i,j,k}^{n_{i,j,k}}$$

This provides a way of capturing syntactic expectations when the probability of each outcome is known with certainty. However, if the subject is truly certain about the probability of each outcome, then no adaptation should occur. Thus, in order to capture adaptation, or change in expectations, we must capture *uncertain* beliefs about such expectations, via a prior distribution over the probabilities $\theta_{i,j,k}$. The most natural choice is the conjugate prior for multinomial probabilities, the Dirichlet distribution:

$$p(\theta) \propto \prod_{i,j,k} \theta_{i,j,k}^{\alpha_{i,j,k}-1}$$

The primary advantage of using this prior distribution is that, after making observations X, the posterior is also Dirichlet, with parameters $\alpha_{i,j,k} + n_{i,j,k}$:

$$p(\theta|X) \propto p(X|\theta)p(\theta) \propto \prod_{i,j,k} \theta_{i,j,k}^{\alpha_{i,j,k}+n_{i,j,k}-1}$$

The parameters ($\alpha_{i,j,k}$) of the Dirichlet prior can thus be interpreted as the number of times each outcome was observed in prior experience. Intuitively, this can be seen just by looking at the equations for the prior and likelihood and noting that α and n appear in the same places.

Under this model, the conditional probability of SC, given specific $V = v$ and $T = t$ is

$$p(\mathrm{SC}|V = v, T = t, \theta) = \theta_{\mathrm{SC}|v,t} = \frac{\theta_{\mathrm{SC},v,t}}{\theta_{\mathrm{SC},v,t} + \theta_{\mathrm{DO},v,t}}$$

It can be shown that $\theta_{\mathrm{SC}|v,t}$ (and by definition $\theta_{\mathrm{DO}|v,t} = 1 - \theta_{\mathrm{SC}|v,t}$) follows a Beta distribution with parameters $(\alpha_{\mathrm{SC},v,t}, \alpha_{\mathrm{DO},v,t})$. Marginalizing over the distribution of $\theta_{i,j,k}$, the conditional probability of SC given verb v and complementizer cue t is:

$$p(\mathrm{SC}|v,t) = \int p(\mathrm{SC}|v,t,\theta)p(\theta|\alpha)\mathrm{d}\theta = \frac{\alpha_{\mathrm{SC},v,t}}{\alpha_{\mathrm{SC},v,t} + \alpha_{\mathrm{DO},v,t}}$$

This conditional probability is the major predictor of syntactic expectation and associated comprehension difficulty.

In order to make quantitative predictions from this general framework, it is necessary to specify the parameters of the prior distribution ($\alpha_{i,j,k}$) and likelihood function ($n_{i,j,k}$), and to relate the model output (conditional probability) to the behavioral measure (reading times). These are addressed in the next sections.

Determining the likelihood

The parameters of the likelihood function are the counts $n_{i,j,k}$ of how often each unique combination of syntactic structure $S = i$, verb $V = j$, and complementizer presence/absence $T = k$ was observed during training, and were set to the counts of the training phase.

Determining the prior

The prior parameters are the pseudo-counts $\alpha_{i,j,k}$ which are proportional to the joint prior probabilities. These probabilities are estimated based on a combination of corpus and norming data. The joint probability of syntactic structures, verbs, and complementizer presence $p(S,T,V)$ can be factored as

$$p(S,T,V) = p(T|V,S)p(S|V)p(V)$$

The verb frequency $p(V)$ is estimated from the British National Corpus, while the SC-bias of each verb (probability of SC completion) $p(S|V)$ and *that*-bias (probability of complementizer presence for SC completions of each verb) $p(T|V,S)$ are estimated based on a sentence-completion norming study (Garnsey et al., 1997).

Together, these factors provide an estimate of the relative prior frequency of each outcome, and thus the relative magnitudes of the $\alpha_{i,j,k}$ Dirichlet prior parameters, but say nothing about their absolute magnitude. The absolute magnitude $A = \sum_{i,j,k} \alpha_{i,j,k}$, corresponds to the degree of confidence in the prior beliefs: the higher A, the more the distribution over the modeled probabilities $\theta_{i,j,k}$ is peaked around the estimated prior frequency, and the less new observations will influence these beliefs. Since there is no way to determine the strength of the prior beliefs *a priori*, we treat A as a free parameter, which controls the degree of adaptation but does not change

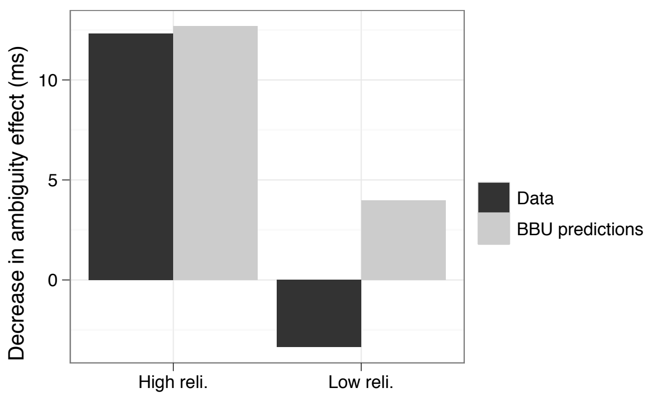

Figure 2: Behavioral results from Fine and Jaeger (2011) (light bars) and corresponding model predictions (dark bars), showing differential effect of high- and low-reliability training on ambiguity effect. Bars show decrease in ambiguity effect (difference in reading times for *that*-present vs. *that*-absent sentences) from pre-test to post-test.

its shape. This is the only free parameter of the model simulations reported here.[1]

Analysis

To evaluate the predictions of the model against the behavioral data of Fine and Jaeger (2011), we regressed the negative log conditional probability against length-corrected reading times. This measure is known as *surprisal*, and has been shown to be a good predictor of reading times in syntactic comprehension (Levy, 2008; Hale, 2001; Fine, Qian, Jaeger, & Jacobs, 2010).

However, there are many factors which influence reading times, of which syntactic expectation may be just one. This measure explicitly removes the influence of verb frequency and *that*-bias, which independently predict reading times for SCs. Also, reading times decrease in self-paced reading tasks just because subjects become better at "pushing the button", an effect which will confound any difference in reading times between pre- and post-test due to adaptation.

In order to control for these confounding effects and evaluate their relationship with our model's predictions about adaptation of syntactic expectations, we fit an increasingly complex series of linear mixed-effects regression models. For each regression model, we compared the baseline, with only the "standard" suite of predictors, with the belief-updating model which additionally includes the surprisal of each item as a predictor. Table 1 shows the factors that were included in each model.

[1]The probability of *that* occurring as a non-complementizer, $p(T = that|V, S = \mathrm{DO})$ cannot be determined from the same norming study, and in corpora it varies dramatically between spoken and written English. For the simulations reported here it was fixed at 0.00005, based on the very low but non-zero value observed in the Wall Street Journal corpus. This does not dramatically change the predictions or the best-fitting pseudocount A.

Model (adds)	Deviance	ΔDev.	df	$p(\chi^2)$
(1|Subj)+ (1|Verb)	28066	195	2	<0.0001
Verb Freq.	28059	195	2	<0.0001
Time (pre/post)	27452	25	2	<0.0001
That presence	27448	15	2	0.0006
Verb SC Bias	27438	18	2	0.0001
Training Group	27432	15	2	0.0005
(interactions)	27216	7	2	0.03

Table 1: Results of linear mixed-effects regression analysis of belief-updating model predictions of self-paced reading times. Each row reports the goodness of fit of a model with belief-updating-predicted surprisal and all of the fixed and random effects listed in its row and the preceding rows (that is, the third model includes random intercepts for subjects and verbs, and a fixed effect for verb log-frequency). In the final row, the baseline model has all fixed effects and their interactions, except for the predicted surprisal, which does not interact with any other effects. The deviance (negative two times model log likelihood) is reported for each model, along with the improvement in deviance over the corresponding baseline model, the additional degrees of freedom, and the result of a χ^2 test.

We found the overall best-fitting parameter values by fitting the whole series of belief-updating regression models using a range of parameter values ($A = 10^{-3}$ to 10^4). The parameter value with the best mean normalized goodness-of-fit (across the various regression models) was used for the results below. We compared both r^2 and deviance as measures of regression goodness-of-fit; both measures produce similar relative goodness-of-fit values but we chose to use deviance since it suffers less from ceiling effects in the most complex (and best-fitting) models. The best fitting prior pseudocount, used to generate the predictions evaluated below, was $A = 2.7$.

Results

Qualitatively, the belief-updating model predicts the three-way interaction between group, time, and ambiguity (presence or absence of *that*), Figure 2. The degree to which subjects rely on the complementizer as a cue to SC continuations—i.e., the strength of the effect of the complementizer on RTs—can be measured by the difference in reading times between complementizer-present and -absent sentences. The model predicts (dark bars), as is observed in the data (light bars), that ambiguity effects should decrease more after high-reliability training (All-SC group) and decrease less (if at all) after low-reliability training (50-50 group).

The results from the regression analysis of the belief udpating model predictions show that the model predictions generally provide a good fit for reading times in the disambiguating region. First, the model predictions alone (with random intercepts for subject and verb) account for 17% of the variance in reading times. This effect cannot be reduced to any of the other controls we evaluated (verb frequency, time, presence of *that*, verb SC-bias, and training condition group): the belief-updating predictions still significantly improve the fit of the model even after controlling for all of these fixed effects and all of their interactions ($\chi^2(2) = 7, p < 0.03$).

Of all of these control predictors, time (pre- vs. post-training) has a notably large effect, and Wells et al. (2009) observed a similar global speed-up between pre- and post-test, independent of effects on the form of the test sentences and the type of training the subject received. This speed-up reflects both increased familiarity with the self-paced reading task (demonstrated by the fact that when the Time predictor is added, the deviance accounted for by the belief-updating predictions is reduced) and the effects of simply having seen more SC structures than in typical written English (captured by the belief-updating predictions after some SC exposure; Fine et al., 2010).

Finally, the value of the prior confidence pseudocount parameter A which best fits the data corresponds to an effective sample size of 2.7 observations for the prior beliefs. This value is very low, but is consistent with results from other belief-updating models of rapid syntactic adaptation and phonetic adaptation (Fine et al., 2010; Kleinschmidt & Jaeger, 2011) and with the larger idea that comprehenders weight prior evidence less in novel situations where rapid adaptation is likely required. Higher values correspond to less adaptation, and produce a worse fit, but interestingly *lower* values produce a worse fit as well. If the good fit of this model were simply due to the qualitative pattern of cue reliability in the exposure statistics, then lower values of A, which result in post-test reading times which better approximate the exposure statistics, should produce better fits, which is not the case. This supports the idea that post-test reading times reflect a combination of prior knowledge and recent experience.

Discussion and conclusion

In this paper, we formally characterize syntactic adaptation as the incremental updating of a probability distribution over syntactic structures. We showed that such a model provides a good fit of behavioral data from a previously published study of syntactic adaptation (Fine & Jaeger, 2011). This model is a first step and leaves much open for future work. Because of how it tracks the co-occurrence statistics of cues and syntactic structures, it does not make meaningful predictions on a trial-by-trial basis for how the overall greater prevalence of SC structures in the experiment changes syntactic expectations for verbs not encountered during training (which influences reading times as well, Fine et al., 2010). Such on-line generalization is not in principle beyond the scope of the type of model presented here, but is omitted in favor of presenting a simple model which demonstrates the connections between adaptation, cue-combination, and statistical learning in syntactic comprehension.

Independent of the details of the particular model presented in this paper, characterizing syntactic adaptation in terms of Bayesian belief update is appealing for at least two reasons. First, by modeling syntactic adaptation as incremental Bayesian belief update, we provide a natural, formal connection between previous work on probabilistic models of expectation-based processing (e.g., Hale, 2001; Levy, 2008) and behavioral work on syntactic adaptation (or syntactic priming; e.g., (e.g., Thothathiri & Snedeker, 2008). Second, using this modeling approach has allowed us to take a step towards providing a single computational framework for adaptation phenomena in language processing, since the same approach has been successfully applied in phonetic adaptation (e.g., Kleinschmidt & Jaeger, 2011). Providing such a "common language" is an important step since this provides a way of bridging insights from multiple strands of psycholinguistic research that have previously been pursued in isolation from each other, notably syntactic priming in comprehension (e.g., Thothathiri & Snedeker, 2008) and perceptual adaptation in speech (e.g., Norris et al., 2003).

As we briefly mentioned in the introduction, the model reported here provides a way of formally describing both adaptation and cue combination. Bayesian models of perception have consistently suggested that, when multiple cues are available in a given task, the perceptual system *weights* those cues according to how *reliable* they are, or, more formally, how narrow or wide the variance is over inferences made *based on* those cues (Jacobs, 2002). In the exposure phase of the study modeled here, the reliability of the verb as a cue to syntactic structure is very high in the All-SC group, but very low in the 50-50 group; on the other hand, the complementizer *that* is a consistently good cue across both groups. Our model qualitatively captures the behavioral result that comprehenders in the 50-50 group come to rely *more* on the complementizer *that* in the post- relative to the pre-exposure phase, compared to the All-SC group (see Figure 2). Significantly, this result comes out of the model as a natural consequence of tracking the joint distribution over syntactic structures (DOs vs. SCs) and syntactic cues (complementizers, verbs). The model here therefore suggests a very close relationship between adaptation and cue combination, and provides a formal account for the classic observation that cues are weighted differentially according to their reliability in language acquisition and language processing (Bates & MacWhinney, 1987).

In general, the approach here is conceptually compatible with a sentence processing research emphasizing the role of experience and learning in language comprehension (e.g., MacDonald, 1999). Bayesian models provide a formal framework for capturing the assumption—shared by many experience-based accounts of processing—that comprehenders monitor and constantly integrate new evidence from the input in order to maintain accurate linguistic expectations, and to process language more efficiently (cf., Smith & Levy, 2008).

The results reported here raise a number of interesting questions that we intend to pursue going forward. First, we employ the same modeling framework and find results that are generally consistent with the modeling results reported in Fine et al. (2010). However, important differences in the experimental design between the two studies leave many questions open. Most significantly, the experiment in Fine et al. (2010) observed changes in reading behavior over a much shorter period of time (one half-hour experimental session) than the current study, which lasted several days. The modeling framework employed here could be extended to examine possible qualitative differences in adaptation over very different time courses, paralleling Bayesian accounts of the time course of adaptation in speech perception (Kleinschmidt & Jaeger, 2011).

Finally, the experiment and model reported here leave open the question of how much the changes in expectations which constitute adaptation *generalize* to novel situations (i.e., did the adaptation effects observed here persist, and influence the way subjects processed language outside the lab?). Rational models of linguistic adaptation generally predict that the extent to which comprehenders generalize adapted expectations should depend on their prior beliefs about the degree of similarity between different situations. This question has been addressed behaviorally in phonetic adaptation (Kraljic & Samuel, 2006, 2007) but remains virtually unexplored in other domains of language processing, and has not been quantitatively modeled. Answering the question of generalization is therefore a high priority for future work on adaptation.

Acknowledgements

This research was partially funded by NSF Graduate Research Fellowships to DK and ABF and NSF Grant BCS-0844472 as well as an Alfred P. Sloan Research Fellowship to TFJ.

References

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

Bates, E., & MacWhinney, B. (1987). Competition, Variation, and Language Learning. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 157–194). Hillsdale, NJ: Lawrence Erlbaum Associates.

Blakemore, C., & Campbell, F. (1969). On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *The Journal of Physiology*, *203*(1), 237–260.

Farmer, T. A., Fine, A. B., & Jaeger, T. F. (2011). Implicit Context-Specific Learning Leads to Rapid Shifts in Syntactic Expectations. In *The 33rd annual meeting of the cognitive science society (cogsci11).* Boston, MA.

Fine, A. B., & Jaeger, T. F. (2011). Language comprehension is sensitive to changes in the reliability of lexical cues. In *The 33rd annual meeting of the cognitive science society (cogsci11).* Boston, MA.

Fine, A. B., Qian, T., Jaeger, T. F., & Jacobs, R. A. (2010). Syntactic adaptation in language comprehension. In *Acl workshop on cognitive modeling and computational linguistics* (pp. 18–26).

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*(37), 58–93.

Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. *Second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies 2001 NAACL 01*, *2*, 1–8.

Jacobs, R. A. (2002). What determines visual cue reliability? *Trends in cognitive sciences*, *6*(8), 345–350.

Kaschak, M. P., & Glenberg, A. M. (2004). This construction needs learned. *Journal of experimental psychology. General*, *133*(3), 450–67.

Kleinschmidt, D., & Jaeger, T. F. (2011). A Bayesian belief updating model of phonetic recalibration and selective adaptation. In *2nd acl workshop on cognitive modeling and computational linguistics.*

Koerding, K. P., Tenenbaum, J. B., & Shadmehr, R. (2007). The dynamics of memory as a consequence of optimal adaptation to a changing body. *Nature Neuroscience*, *10*(6), 779–86.

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic bulletin & review*, *13*(2), 262–8.

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

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–77.

Luka, B., & Barsalou, L. (2005). Structural facilitation: Mere exposure effects for grammatical acceptability as evidence for syntactic priming in comprehension. *Journal of Memory and Language*, *52*(3), 436–459.

MacDonald, M. C. (1999). Distributional information in language comprehension, production, and acquisition: Three puzzles and a moral. In *The emergence of language* (pp. 177–196). Mahwah, NJ: Erlbaum.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology*, *47*(2), 204–238.

Smith, N., & Levy, R. (2008). Optimal processing times in reading: A formal model and empirical investigation. In *Proceedings of the 30th annual conference of the cognitive science society* (pp. 595–600). Austin, TX.

Sonderegger, M., & Yu, A. (2010). A rational account of perceptual compensation for coarticulation. In *Proceedings of the 32nd annual conference of the cognitive science society* (pp. 375–380).

Thothathiri, M., & Snedeker, J. (2008). Give and take: syntactic priming during spoken language comprehension. *Cognition*, *108*(1), 51–68.

Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. (2009). Experience and sentence processing: statistical learning and relative clause comprehension. *Cognitive psychology*, *58*(2), 250–71.