BLS 35, No 1 2009. DOI: http://dx.doi.org/10.3765/bls.v35i1.3594
(published by the Berkeley Linguistics Society and the Linguistic Society of America)

Correcting the Incorrect: Local Coherence Effects Modeled with Prior Belief Update

KLINTON BICKNELL,[1] ROGER LEVY,[1] and VERA DEMBERG[2]
[1]*University of California, San Diego,* [2]*University of Edinburgh*

0. Introduction

In the past four decades, the field of sentence processing research has generated a number of models of the incremental operation of the human sentence processor. One assumption common to most of these theories is that the difficulty of a word is determined by the possible syntactic structures of the preceding words, and thus, the difficulty of a word should be unaffected by ungrammatical analyses of the preceding words. Put formally, the difficulty of word w_n of a sentence is determined only by the grammatical syntactic structures covering the preceding words $w_1 \cdots w_{n-1} (\equiv w_1^{n-1})$. Recent results by Tabor, Galantucci, and Richardson (2004), however, appear to show evidence of a case where a syntactic structure which is not possible given w_1^{n-1} nevertheless influences the difficulty of w_n. They attribute such effects to 'merely locally coherent' syntactic structures and term such impossible structures *local coherences*. Follow-up studies by Konieczny (2005) and Konieczny and Müller (2006, 2007) provide further evidence that these impossible structures are being constructed and even semantically evaluated. These results have been taken to support a small class of dynamical systems models of sentence processing (e.g., Tabor and Hutchins 2004), in which, crucially, structures which are not possible given the current input are nevertheless constructed and compete with other, tenable structures. Unfortunately, the existing theories in this class have a large number of interacting free parameters, making interpretation somewhat difficult and leaving unspecified how to scale up such a system to make broad-coverage reading time predictions. This paper fills two gaps in the literature on local coherences. First, it demonstrates from two experiments with an eye-tracking corpus that effects of local coherences are evident in the reading of naturalistic text. Second, it describes a new computational model of local coherences that is motivated by a view of sentence processing as updating prior beliefs over syntactic structures.

0.1. Local Coherences: The Initial Result

The first study to report effects of local coherences is described in Tabor, Galantucci, and Richardson (2004). In Experiment 1, they use a self-paced reading task and materials containing relative clauses (RCs) attached to nouns in non-subject

position, as in (1).

(1) a. The coach smiled at the player tossed a frisbee by ...
b. The coach smiled at the player who was tossed a frisbee by ...
c. The coach smiled at the player thrown a frisbee by ...
d. The coach smiled at the player who was thrown a frisbee by ...

Their experimental design crossed RC reduction with verb ambiguity. RCs are either reduced (1a,c) or unreduced (1b,d), and the RC verb is either lexically ambiguous between a past tense active and a past participle (1a–b), or is unambiguously a past participle (1c–d).

Tabor, Galantucci, and Richardson point out that in one of these four conditions (1a) there is a locally coherent string *the player tossed a frisbee*. Out of context (e.g., if it were starting a sentence) this string would have a likely parse in which *the player* is the agent of *tossed* and *a frisbee* is the theme. Given the preceding context, however, *the player* is in non-subject position and thus this parse is impossible. That is, given the preceding context, *the player tossed the frisbee* must begin a reduced RC, and there is no local ambiguity. Thus, so long as ungrammatical analyses are not considered, (1a) should be no more difficult than the other examples, except insofar as ambiguous verbs are harder than unambiguous verbs, and reduced RCs are harder than unreduced RCs. That is, the prediction for reading times in the *tossed a frisbee by* region from most theories of sentence processing would be to get the two main effects of RC reduction and verb ambiguity.

Tabor, Galantucci, and Richardson, however, predict an interaction such that (1a) will have added difficulty above and beyond these two effects, because of the interference from the locally coherent parse of *the player tossed a frisbee*. Concordant with their predictions, they find an interaction in the *tossed a frisbee by* region, such that reading times for (1a) are super-additively high, suggesting that ungrammatical analyses are considered by the human sentence processor.

0.2. Local Coherences: Theories

With the results showing effects of local coherences in mind, we can ask the question of what sorts of theories predict these effects. This section briefly describes two recent examples of such theories. The first involves dynamical systems models to explain the effects, while the second uses a mathematical model of the combination of bottom-up and top-down probabilistic information.

Tabor and Hutchins (2004) describes the SOPARSE (self-organized parse) model, in which reading a word activates a set of lexically anchored tree fragments. These tree fragments then compete, spreading activation to compatible fragments and inhibiting incompatible fragments, such that the system eventually stabilizes to the correct parse. Reading times for each word can then be modeled as the time the system takes to stabilize after reading a word. Stabilization takes longer for locally coherent regions because the locally coherent parse is created and competes with the globally grammatical parses, thus nicely explaining the results on local coherences.

There are, however, unsolved issues with this model. The model has a number of free parameters, relating to the equations used for the competition, the method by which links between fragments are formed, as well as the question of precisely what tree fragments a given word will activate. While Tabor and Hutchins (2004) works out these questions in detail for the types of sentences they model, it is unclear how to scale the model up to make predictions for arbitrary types of sentences. That is, there is no principled system for setting the three types of parameters mentioned, and no clear interpretation of their values. The model put forward in this paper is an attempt to remedy this situation.

A recent proposal by Gibson (2006) can also explain some of the local coherence results. Gibson's proposal is that part-of-speech ambiguities have a special status in parsing; in effect, lexical part-of-speech ambiguities can be thought of as one-word local coherences. In this model, a lexical bias (LB) is created for each part-of-speech tag t_i of word w by multiplying together the context-independent probability of t_i given the word w (the *bottom-up* component) by a smoothed probability of the tag given the context (the *top-down* component):

(2) $$LB(t_i) = P(t_i|w)P_s(t_i|\text{context})$$

P_s is smoothed by adding .01 to the probability of every tag $t \in T$, such that it no longer sums to one, and is thus not a true probability function. Then, a true probability is calculated for each tag t_i by normalizing the LB terms:

(3) $$P(t_i) = \frac{LB(t_i)}{\sum_{t \in T} LB(t)}$$

Gibson describes two ways in which the resultant probabilities can be used to predict difficulty, one for serial and one for parallel models. For serial models, the parser stochastically selects a part-of-speech for the current word from the $P(t)$ distribution. When the part-of-speech it selects cannot be integrated into the current syntactic representation, difficulty occurs from reanalysis. In a parallel model, the parser maintains all possibilities for the part-of-speech of the word, weighted by $P(t)$. In cases where multiple parts of speech have positive probabilities, competition ensues.

Because the top-down probabilities are smoothed to allow for all possible parts-of-speech, any word which is lexically ambiguous will be more difficult to process, regardless of whether it is ambiguous or not in its context. This can thus explain some of the difference between the ambiguous and unambiguous verbs in Tabor, Galantucci, and Richardson (2004). It is not clear, however, under such a model why the super-additive interaction would obtain. Furthermore, such a theory cannot at all explain the semantic effects of local coherences, such as those described in Tabor, Galantucci, and Richardson's (2004) Experiment 2, or the visual world results of Konieczny and Müller (2006, 2007). In addition, Gibson's model is a bit underspecified: he does not discuss how the top-down probabilities are calculated, nor

what the precise linking hypothesis is between the final $P(t)$ and reading times. Finally, it is not at all clear why the top-down expectations should be smoothed, since the smoothing actually has negative consequences on the processor's performance.

0.3. Goals

The goals of this paper are twofold. The first goal concerns the empirical status of effects of local coherences. All of the extant results on the phenomenon involve controlled experiments, most of which crucially involve very rare types of constructions. For example, the result of Tabor, Galantucci, and Richardson (2004) relies on reduced relative clauses formed from a passivization on the recipient of a ditransitive construction. Such a type of sentence is quite rare in English, and thus might not give useful insight into the normal operation of the sentence processor. This paper presents the results of two experiments with a corpus of eye-tracking data from the reading of newspaper articles demonstrating effects of local coherences in the reading of naturalistic sentences. This establishes the ecological validity of the study of local coherences, and underscores the need for a theory of local coherences which makes broad-coverage predictions. The second goal of this paper is to present a model of the effects of local coherences that combines the strengths of Gibson's (2006) and Tabor and Hutchins's (2004) models. This model accounts for phrasal-level effects of local coherences (as Tabor and Hutchins), but does so using general quantities that can be calculated for any sentence type (as Gibson) by using a general probabilistic parser that can operate on any SCFG. The remainder of this paper is divided into four sections. The next two sections present the two corpus experiments. Following that, we present our model and conclude.

1. Experiment 1

The basic strategy of the two corpus experiments is to build a regression model of the reading times on each word in an eye-tracking corpus. Included in the regression model for each experiment is a factor quantifying the occurrence of local coherences. Establishing that local coherences have an effect on reading times is then merely a matter of assessing the significance of the local coherences factor in the model, and assessing the size of that effect is merely a matter of inspecting the coefficient estimate.

The local coherences factor in Experiment 1 is meant to start simple by measuring the effect of one-word local coherences. Although *prima facie*, one-word local coherences do not seem to look much like the materials in Tabor, Galantucci, and Richardson (2004), the reasoning for calling them local coherences is as follows: we take the definition of a local coherence to be a string of words w that out of context would suggest one very likely parse, and that parse is impossible (or at least highly unlikely) in context. We can scale this down to the case where w is a string of size 1; that is, out of context, a word w suggests a very likely parse (e.g., a part-of-speech tag) that is very unlikely or impossible in context. Because a word only has one part-of-speech in a given sentence, this means we can invert this statement to say that a one-word local coherence occurs when the only possible part-of-speech tag

for the word in context is highly unlikely out of context. By making the assumption that the only possible part-of-speech tag for a word in context can be approximated by the actual part-of-speech tag the word has in the sentence, we can calculate our one-word local coherence factor to be an estimate of the context-independent probability of the actual part-of-speech tag t_i for a word w_i given just the word $P(t_i|w_i)$. This factor will thus be low when there is a strong one-word local coherence.

This particular type of one-word local coherence is predicted to have an effect by both Gibson's (2006) and Tabor and Hutchins's (2004) models. This probability is actually one of the components in the Gibson theory, which would predict that – all else being equal – a word w_i would be read more slowly as $P(t_i|w_i)$ decreases. Just as Gibson's theory would predict, this factor would assign lower probability to *tossed* tagged as a past participle than it would *thrown* tagged as a past participle. A dynamical systems model such as Tabor and Hutchins's makes the same prediction if we assume that the strength of the lexically-anchored tree fragments corresponding to each part-of-speech vary in strength in proportion to $P(t_i|w_i)$, which seems to be a reasonable interpretation of what their model would involve. Of course, this factor doesn't capture local coherences at a phrasal level, as Tabor and Hutchins would predict. The next experiment remedies this situation somewhat by scaling up this factor by conditioning on two words, and the model given in the paper's next section completely eliminates this objection by specifying a theory predicting phrasal-level local coherences of an arbitrary length.

1.1. Methods

1.1.1. Data

This experiment makes use of the Dundee corpus (Kennedy and Pynte 2005) of eye-movement data from 10 participants reading 51,000 words each of *The Independent* (a British newspaper). To get part-of-speech tags for the corpus, we parsed it using the Charniak parser (Charniak 2000). From the eye-tracking record given in the corpus, we calculated our dependent measure of first pass times for each word, defined as the total duration of all fixations on a word prior to having fixated anything to its right.

1.1.2. Model

We tested the local coherence factor in a linear mixed-effect model (Pinheiro and Bates 2000; for a psycholinguistic introduction see also Baayen, Davidson, and Bates 2008) of the first pass times on each word, containing 11 fixed effect control factors and participant as a random effect, as in Demberg and Keller (2008). Coefficient estimates and significance levels were estimated by Markov chain Monte Carlo (MCMC) sampling (Baayen, Davidson, and Bates 2008).

1.1.3. Control Variables

We took our control factors from Demberg and Keller (2008). They included linguistic properties such as word length in characters, the logarithm of word frequency per million as estimated from the British National Corpus (BNC), bigram probability ($P(w_i|w_{i-1})$; also estimated from the BNC), and position in the sen-

tence in words. In addition, they included lexicalized and unlexicalized syntactic surprisal ($-\log P(w_i|w_1^{i-1})$), as well as eye movement properties such as the landing position with respect to the word, the number of characters between last fixation and current fixation, and whether the previous word was fixated.

1.1.4. Factor Estimation

For each word-tag pair in the Dundee corpus, we estimated $P(t_i|w_i)$ from a Charniak-parsed version of the BNC. We used two versions of the factor: P_m was the maximum likelihood estimate (MLE) and P_s was a smoothed version. P_s was calculated by smoothing the MLE with a type-averaged distribution over part-of-speech tags. Specifically, a type-averaged distribution P_{pr} was calculated for a given tag t_i as

$$P_{pr}(t_i) = \frac{\sum_w P_m(t_i|w)}{|w|} \tag{4}$$

where w ranges over word types (as opposed to tokens). The smoothed probability P_s of a tag t_i given a word w_i is then calculated to be

$$P_s(t_i|w_i) = \frac{c(t_i, w_i) + \beta(P_{pr}(t_i))}{c(w_i) + \beta} \tag{5}$$

where $c(t_i, w_i)$ returns the count of w_i tagged as t_i in the corpus, and β is set to minimize Dundee corpus perplexity.

1.1.5. Log Transform

In addition to the probabilities themselves, the base-2 logarithms of both versions of the factor, P_m and P_s, were also entered into the regression.

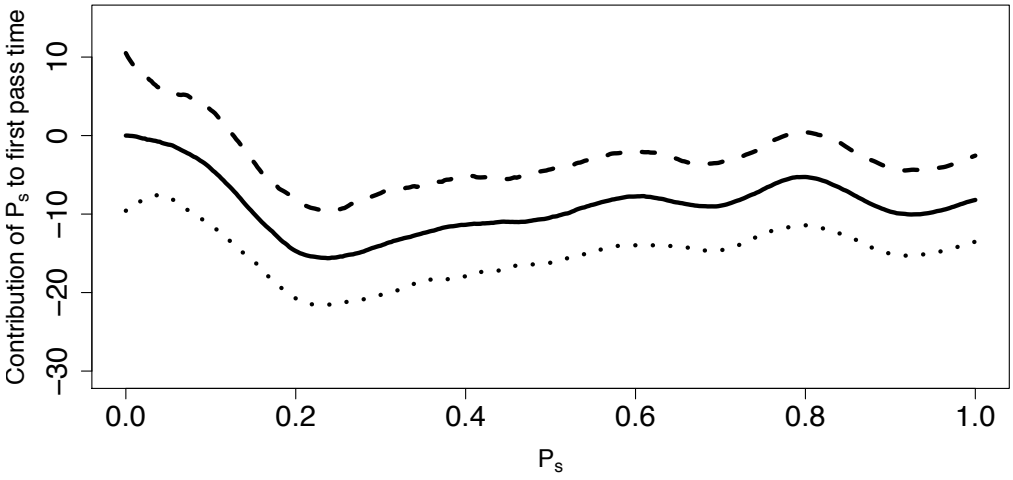
1.1.6. Data Selection

We excluded from the analysis any word in the Dundee corpus that had punctuation, contained numbers, did not contain letters, occurred as the first or last word of a line, or that did not occur in the BNC. In addition, we excluded the first-pass times on any word which had a first-pass time of zero.

1.2. Results

The better fit to the data was achieved by the log-transformed versions of the factor. The $\log P_m$ factor had a coefficient estimate of -0.71 ($p < .0001$), and the $\log P_s$ factor had a coefficient estimate of -0.80 ($p < .01$). By contrast, the linear fit version of P_m had an insignificant coefficient estimate of -0.84 ($p = .42$), and P_s had a coefficient estimate of -2.31 ($p < .05$). To better visualize the results, a natural spline regression was performed on P_s with 11 equally spaced knots. The result is shown in Figure 1 with bootstrapped 95% confidence intervals.

Figure 1: Natural spline regression on $P_s(t_i|w_i)$ with 11 equally spaced knots. 95% confidence intervals are bootstrapped.



1.3. Discussion

Since the better fit was achieved using the logarithmic version of the factors, we focus here on their interpretation. For both P_m and P_s, doubling the probability of a tag reduces the first-pass time by about 7 or 8 tenths of a millisecond. Looking at the spline regression in Figure 1 reveals that most of the differences it is accounting for exist for probabilities under 0.2. While this seems to be a somewhat small effect, the significance levels of these factors reveal that the effects are reliable. This provides the first evidence for the effects of local coherences (albeit local coherences consisting of one word) in the reading of naturalistic text.

2. Experiment 2

The second experiment is very similar to the first. In this case, however, we test for effects of two-word local coherences, again at the part-of-speech tag level, using as our factor an estimate of $P(t_i|w_{i-1}^i)$. To see how this factor is a measure of two-word local coherences, consider again the definition of local coherence effects we used above: a string of words w that out of context would suggest one very likely parse which is impossible (or at least highly unlikely) in context. If we again invert that definition, because part-of-speech tags are mutually exclusive, we see that local coherence effects occur when the only possible part-of-speech tag for a word in a sentential context is highly unlikely out of that context. Once again, we are using the actual part-of-speech tag of a word as a crude estimate of the only possible part-of-speech tag.

Take as an example a two-word sequence from Tabor, Galantucci, and Richardson, *player tossed*. Out of context this string is likely to have a parse where *tossed* is a past tense verb and very unlikely to have a parse where *tossed* is a past participle. Thus, this factor would predict reading *tossed* as a past participle to be especially difficult given that the previous word was *player*. Dynamical systems theories such

as Tabor and Hutchins (2004) would also predict a word to be read more slowly as the $P(t_i|w_{i-1}^i)$ decreases, since the lexically-anchored tree fragments for the two words should cooperate to cause a large amount of interference to the globally correct parse. While Gibson's theory would not predict the previous word to have an effect, this model still looks very similar to the sort of integration process he proposes, and may be one natural way to scale his theory up to the multi-word case.

2.1. Methods

2.1.1. Data, Model, and Control Variables

The data, model, and control variables used for Experiment 2 are the same as for Experiment 1.

2.1.2. Factor Estimation

For each word-word-tag triplet in the Dundee corpus, we estimated $P(t_i|w_{i-1}^i)$ from the Charniak-parsed BNC. As in Experiment 1, we used two versions of the factor: P_m was the maximum likelihood estimate (MLE) and P_s was version smoothed from the MLE using the same method as in Experiment 1.

2.1.3. Log Transform

As before, the base-2 logarithms of P_m and P_s were also entered into the regression.

2.1.4. Data Selection

As in Experiment 1, we excluded from the analysis any word in the Dundee corpus that had punctuation, contained numbers, did not contain letters, or occurred as the first or last word of a line. Words were also excluded when the bigram of that word and the previous word did not occur in the BNC. In addition, as in Experiment 1, we excluded the first-pass times on any word which had a first-pass time of zero.

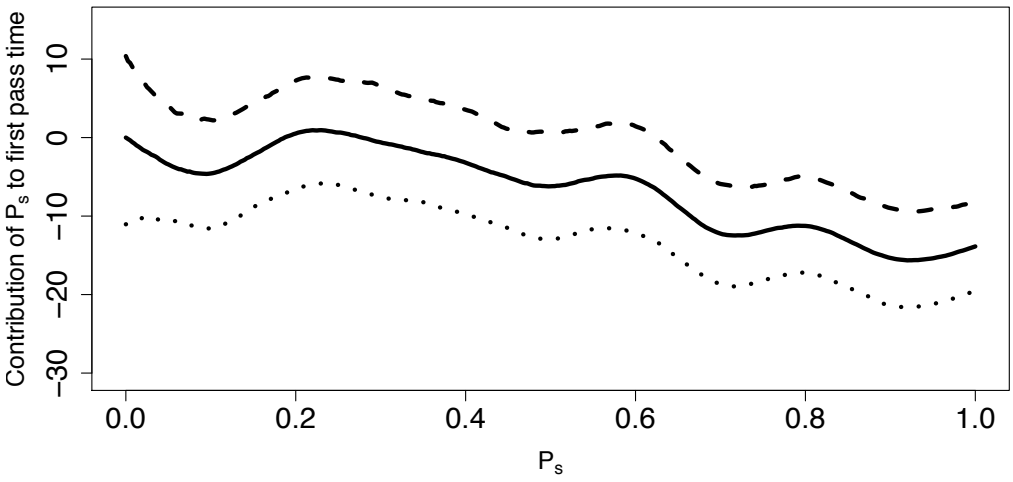
2.2. Results

As in Experiment 1, the better fit to the data was achieved by the log-transformed versions of the factors. The $\log P_m$ factor had a coefficient estimate of -0.43 ($p < .0001$), and the $\log P_s$ factor had a coefficient estimate of -4.10 ($p < .0001$). The linear fit version of P_m had a coefficient estimate of -2.94 ($p < .05$), and P_s had a coefficient estimate of -17.56 ($p < .0001$). As before, a natural spline regression performed on P_s is shown in Figure 2 with bootstrapped confidence intervals.

2.3. Discussion

Inspecting the coefficients for the better-fitting logarithmic versions of the factor reveals that this factor has a much larger effect that of Experiment 1. The coefficient estimate for the smoothed version indicates that doubling the probability of a tag reduces first-pass time by only about 4 tenths of a millisecond. The reason that this coefficient is even smaller than in Experiment 1 is probably simply because the probability function we are estimating is much more sparse than before, and thus smoothing is necessary. The coefficient estimate for the smoothed version of the factor indicates that doubling the probability of a tag reduces the first-pass time

Figure 2: Natural spline regression on $P_s(t_i|w_{i-1}^i)$ with 11 equally spaced knots. 95% confidence intervals are bootstrapped.



by over 4 ms. Inspecting the results of the spline regression in Figure 2 indicates that this trend is true across the range of probability. Again, the significance levels indicate that this effect is highly reliable in this dataset. This provides evidence for effects of multi-word local coherences in the reading of naturalistic text, and, because of the effect size, suggests that such effects are an even more important part of sentence processing than effects of single-world local coherences.

3. The Model

The demonstration in Experiment 2 that the effects of multi-word local coherences appear in the reading of naturalistic text underscores the need for a theory of phrasal-level local coherences which can make broad-coverage predictions. This section presents one such model. The basic intuition behind it is that incrementally processing a sentence can be conceptualized as a process of updating one's beliefs. Such an analogy has been used to motivate surprisal-based theories of sentence processing (Hale 2001; Levy 2008), where beliefs about the structure of a sentence after seeing the first $i-1$ words in the sentence w_1^{i-1} are updated upon encountering w_i. In this case, the *surprisal* of a word ($-\log P(w_i|w_1^{i-1})$) is equivalent to the Kullback-Leibler divergence of the beliefs after w_i from the beliefs before (Levy 2008). Our model focuses on another belief-update process in sentence processing: updating beliefs about the structures that a string of words is likely to have independent of context to beliefs about what structures it is likely to have in context.

A bit more formally, it views the process of integrating a string of words w_i^j into a sentence as beginning with a 'bottom-up' prior distribution of syntactic structures likely to span w_i^j and integrating that with 'top-down' knowledge from the previous words in the sentence w_1^{i-1} in order to reach a posterior distribution conditioning on w_1^j over which structures actually can span w_i^j. This belief update process can be viewed as a rational reconstruction of the Tabor and Hutchins (Tabor and

Hutchins 2004) model, where – instead of the system dynamics of competition between arbitrary tree fragments – differences between prior and posterior probability distributions over syntactic structures determine processing difficulty.

More formally still, when integrating w_i^j into a sentence, for each syntactic category X, we can define the prior probability conditioned only on w_i^j that w_i^j will form the beginning of that category, i.e., that an X exists which begins at index i and spans at least through j:

(6) Prior: $P(X_i^{k \geq j} | w_i^j)$

It is important to note here that this prior probability is conditional only on the value of w_i^j and not the values of i or j; that is, in the prior probability, i and j should be interpreted merely as a way to coindex the start and end points of the string of words being integrated with a category X potentially spanning them, and not as making reference to position in the full sentence string.

For each category X, this prior probability will be updated to the posterior probability of that category spanning w_i^j given all the words seen so far:

(7) Posterior: $P(X_i^{k \geq j} | w_1^j)$

In the equation for the posterior, of course, the indices i and j are positions in the sentence string, and not merely coindices.

Given these prior and posterior beliefs, we predict difficulty to arise in cases where the prior requires substantial modification to reach the posterior, that is, cases in which the prior and posterior make substantially different predictions for categories. A strong local coherence will have sharply different prior and posterior distributions, causing difficulty. We measure M_{ij}, the amount of modification required, as the K-L divergence of the prior from the posterior summed over syntactic categories. That is, if N is the set of non-terminal categories in the grammar, the size of the belief update is modeled as

(8) $$M_{ij} \stackrel{\text{def}}{=} \sum_{X \in N} D\left(P(X_i^{k \geq j} | w_1^j) \,||\, P(X_i^{k \geq j} | w_i^j)\right)$$

In Bicknell and Levy (2009), we show how to compute M_{ij} by using Bayesian inference on quantities calculated in ordinary probabilistic incremental Earley parsing with a stochastic context-free grammar (SCFG). Furthermore, we present the results of a computational experiment showing that our model makes the correct predictions on the original local coherences experiment of Tabor, Galantucci, and Richardson (2004).

4. Conclusion

This paper has made two contributions to the study of local coherences: a set of corpus experiments and a new model. The two novel corpus experiments showed evidence that effects of local coherences consisting of one or two words occur in the reading of naturalistic text. The first experiment showed a reliable effect of single-

word local coherences, such as those predicted by the model of Gibson (2006), and compatible with dynamical systems models such as that of Tabor and Hutchins (2004). The second experiment showed an even larger effect of two-word local coherences, such as those predicted by dynamical systems models such as Tabor and Hutchins (2004). Such results give ecological validity to the study of local coherences and demonstrate that they are not merely artifacts in the processing of very rare sentence types. Furthermore, the results suggested that two-word local coherences appear to be stronger than single-word coherences.

This latter observation led to the description of a mathematical model predicting where effects of local coherences will occur in arbitrary sentences. The fundamental insight of this model is that effects of local coherences can be described in terms of updating prior beliefs about the structures a new string of words is likely to take independent of context into posterior beliefs about what structures it is likely to take given contextual information. This model predicts local coherence effects to occur whenever prior and posterior beliefs are substantially different.

In contrast to Gibson's model, this model can account for all existing results on phrasal-level local coherences. In contrast to the dynamical systems models, it does not require assuming a rather arbitrary parsing mechanism with a large number of free parameters, but rather is described in terms only of probabilities in a grammar (which can be estimated in a principled and straightforward way). Future work will test that model's predictions on the reading of naturalistic text in a similar way to Experiments 1 and 2.

Acknowledgments

This work has benefited from useful discussion with Nathaniel Smith and from feedback from audiences at the the 2008 annual CUNY Conference on Human Sentence Processing and the 2009 Annual Meeting of the LSA. The research was supported by NIH Training Grant T32-DC000041 from the Center for Research in Language at UC San Diego to the first author.

References

Baayen, R.H., D.J. Davidson, and D.M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4):390–412.

Bicknell, Klinton and Roger Levy. 2009. A model of local coherence effects in human sentence processing as consequences of updates from bottom-up prior to posterior beliefs. In *Proceedings of the North American Chapter of the ACL: Human Language Technologies (NAACL HLT) 2009 Conference*, 665–673. Association for Computational Linguistics.

Charniak, Eugene. 2000. A maximum-entropy-inspired parser. In *Proceedings of the First Meeting of the North American Chapter of the ACL (NAACL)*, 132–139. San Francisco: Morgan Kaufmann Publishers Inc.

Demberg, Vera and Frank Keller. 2008. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition* 109(2):193–210.

Gibson, Edward. 2006. The interaction of top-down and bottom-up statistics in the resolution of syntactic category ambiguity. *Journal of Memory and Language* 54:363–388.

Hale, John. 2001. A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics*, Volume 2, 159–166. New Brunswick, NJ: Association for Computational Linguistics.

Kennedy, Alan and Joël Pynte. 2005. Parafoveal-on-foveal effects in normal reading. *Vision Research* 45:153–168.

Konieczny, Lars. 2005. The psychological reality of local coherences in sentence processing. In B. G. Bara, L. Barsalou, and M. Bucciarelli, eds., *Proceedings of the 27th annual meeting of the Cognitive Science Society*, 1178–1183. Mahwah, NJ: Lawrence Erlbaum Associates.

Konieczny, Lars and Daniel Müller. 2006. Local coherence interpretation in spoken language: Evidence from a visual world experiment. Presented at AMLaP (Architectures and Mechanisms for Language Processing) 2006, Nijmegen, the Netherlands.

Konieczny, Lars and Daniel Müller. 2007. Local coherence interpretation in written and spoken language. Presented at the 20th Annual CUNY Conference on Human Sentence Processing. La Jolla, CA.

Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106:1126–1177.

Pinheiro, José C. and Douglas M. Bates. 2000. *Mixed effects models in S and S-Plus*. New York: Springer Verlag.

Tabor, Whitney, Bruno Galantucci, and Daniel Richardson. 2004. Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language* 50:355–370.

Tabor, Whitney and Sean Hutchins. 2004. Evidence for self-organized sentence processing: Digging-in effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30(2):431–450.

Klinton Bicknell & Roger Levy
University of California, San Diego
Department of Linguistics
9500 Gilman Dr., Mail Code 0108
La Jolla, CA 92093-0108

{kbicknell,rlevy}@ling.ucsd.edu

Vera Demberg
Informatics Forum
10 Crichton Street
Edinburgh EH8 9AB
Scotland, UK

v.demberg@ed.ac.uk