

Integrating surprisal and uncertain-input models in online sentence comprehension: formal techniques and empirical results

Roger Levy

Department of Linguistics
University of California at San Diego
9500 Gilman Drive # 0108
La Jolla, CA 92093-0108
rlevy@ucsd.edu

Abstract

A system making optimal use of available information in incremental language comprehension might be expected to use linguistic knowledge together with current input to revise beliefs about previous input. Under some circumstances, such an error-correction capability might induce comprehenders to adopt grammatical analyses that are inconsistent with the true input. Here we present a formal model of how such input-unfaithful garden paths may be adopted and the difficulty incurred by their subsequent disconfirmation, combining a rational noisy-channel model of syntactic comprehension under uncertain input with the surprisal theory of incremental processing difficulty. We also present a behavioral experiment confirming the key empirical predictions of the theory.

1 Introduction

In most formal theories of human sentence comprehension, input recognition and syntactic analysis are taken to be distinct processes, with the only feedback from syntax to recognition being prospective prediction of likely upcoming input (Jurafsky, 1996; Narayanan and Jurafsky, 1998, 2002; Hale, 2001, 2006; Levy, 2008a). Yet a system making optimal use of all available information might be expected to perform fully joint inference on sentence identity and structure given perceptual input, using linguistic knowledge both prospectively *and* retrospectively in drawing inferences as to how raw input should be segmented and recognized as a sequence of linguistic tokens, and about the degree to which each input

token should be trusted during grammatical analysis. Formal models of such joint inference over uncertain input have been proposed (Levy, 2008b), and corroborative empirical evidence exists that strong coherence of current input with a perceptual neighbor of previous input may induce confusion in comprehenders as to the identity of that previous input (Connine et al., 1991; Levy et al., 2009).

In this paper we explore a more dramatic prediction of such an uncertain-input theory: that, when faced with sufficiently biasing input, comprehenders might under some circumstances adopt a grammatical analysis inconsistent with the true raw input comprising a sentence they are presented with, but consistent with a slightly perturbed version of the input that has higher prior probability. If this is the case, then subsequent input strongly disconfirming this “hallucinated” garden-path analysis might be expected to induce the same effects as seen in classic cases of garden-path disambiguation traditionally studied in the psycholinguistic literature. We explore this prediction by extending the rational uncertain-input model of Levy (2008b), integrating it with SURPRISAL THEORY (Hale, 2001; Levy, 2008a), which successfully accounts for and quantifies traditional garden-path disambiguation effects; and by testing predictions of the extended model in a self-paced reading study. Section 2 reviews surprisal theory and how it accounts for traditional garden-path effects. Section 3 provides background information on garden-path effects relevant to the current study, describes how we might hope to reveal comprehenders’ use of grammatical knowledge to revise beliefs about the identity of previous linguistic sur-

face input and adopt grammatical analyses inconsistent with true input through a controlled experiment, and informally outlines how such belief revisions might arise as a side effect in a general theory of rational comprehension under uncertain input. Section 4 defines and estimates parameters for a model instantiating the general theory, and describes the predictions of the model for the experiment described in Section 3 (along with the inference procedures required to determine those predictions). Section 5 reports the results of the experiment. Section 6 concludes.

2 Garden-path disambiguation under surprisal

The SURPRISAL THEORY of incremental sentence-processing difficulty (Hale, 2001; Levy, 2008a) posits that the cognitive effort required to process a given word w_i of a sentence in its context is given by the simple information-theoretic measure of the log of the inverse of the word’s conditional probability (also called its “surprisal” or “Shannon information content”) in its intra-sentential context $w_{1,\dots,i-1}$ and extra-sentential context $Ctxt$:

$$\text{Effort}(w_i) \propto \log \frac{1}{P(w_i|w_{1,\dots,i-1}, Ctxt)}$$

(In the rest of this paper, we consider isolated-sentence comprehension and ignore $Ctxt$.) The theory derives empirical support not only from controlled experiments manipulating grammatical context but also from broad-coverage studies of reading times for naturalistic text (Demberg and Keller, 2008; Boston et al., 2008; Frank, 2009; Roark et al., 2009), including demonstration that the shape of the relationship between word probability and reading time is indeed log-linear (Smith and Levy, 2008).

Surprisal has had considerable success in accounting for one of the best-known phenomena in psycholinguistics, the GARDEN-PATH SENTENCE (Frazier, 1979), in which a local ambiguity biases the comprehender’s incremental syntactic interpretation so strongly that upon encountering disambiguating input the correct interpretation can only be recovered with great effort, if at all. The most famous example is (1) below (Bever, 1970):

- (1) The horse raced past the barn fell.

where the context before the final word is strongly biased toward an interpretation where *raced* is the main verb of the sentence (**MV**; Figure 1a), the intended interpretation, where *raced* begins a reduced relative clause (**RR**; Figure 1b) and *fell* is the main verb, is extremely difficult to recover. Letting T_j range over the possible incremental syntactic analyses of words $w_{1\dots6}$ preceding *fell*, under surprisal the conditional probability of the disambiguating continuation *fell* can be approximated as

$$P(\textit{fell}|w_{1\dots6}) = \sum_j P(\textit{fell}|T_j, w_{1\dots6})P(T_j|w_{1\dots6}) \tag{I}$$

For all possible predisambiguation analyses T_j , either the analysis is disfavored by the context ($P(T_j|w_{1\dots6})$ is low) or the analysis makes the disambiguating word unlikely ($P(\textit{fell}|T_j, w_{1\dots6})$ is low). Since every summand in the marginalization of Equation (I) has a very small term in it, the total marginal probability is thus small and the surprisal is high. Hale (2001) demonstrated that surprisal thus predicts strong garden-pathing effects in the classic sentence *The horse raced past the barn fell* on basis of the overall rarity of reduced relative clauses alone. More generally, Jurafsky (1996) used a combination of syntactic probabilities (reduced RCs are rare) and argument-structure probabilities (*raced* is usually intransitive) to estimate the probability ratio of the two analyses of pre-disambiguation context in Figure 1 as roughly 82:1, putting a lower bound on the additional surprisal incurred at *fell* for the reduced-RC variant over the unreduced variant (*The horse that was raced past the barn fell*) of 6.4 bits.¹

3 Garden-pathing and input uncertainty

We now move on to cases where garden-pathing can apparently be blocked by only small changes to the surface input, which we will take as a starting point for developing an integrated theory of uncertain-input inference and surprisal. The backdrop is what is known in the psycholinguistic literature as the **NP/Z** ambiguity, exemplified in (2) below:

¹We say that this is a “lower bound” because incorporating even finer-grained information—such as the fact that *horse* is a canonical subject for intransitive *raced*—into the estimate would almost certainly push the probability ratio even farther in favor of the main-clause analysis.

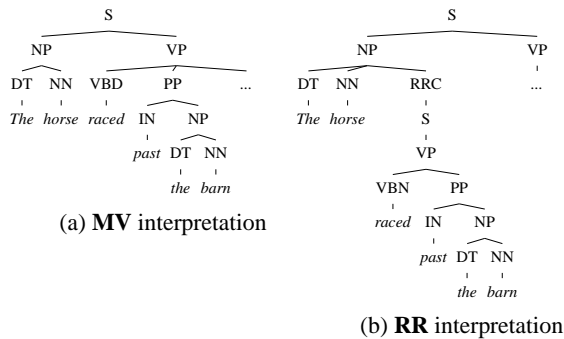


Figure 1: Classic garden pathing

(2) While Mary was mending the socks fell off her lap.

In incremental comprehension, the phrase *the socks* is ambiguous between being the **NP** object of the preceding subordinate-clause verb *mending* versus being the subject of the main clause (in which case *mending* has a **Zero** object); in sentences like (2) the initial bias is toward the **NP** interpretation. The main-clause verb *fell* disambiguates, ruling out the initially favored **NP** analysis. It has been known since Frazier and Rayner (1982) that this effect of garden-path disambiguation can be measured in reading times on the main-clause verb (see also Mitchell, 1987; Ferreira and Henderson, 1993; Adams et al., 1998; Sturt et al., 1999; Hill and Murray, 2000; Christianson et al., 2001; van Gompel and Pickering, 2001; Tabor and Hutchins, 2004; Staub, 2007). Small changes to the context can have huge effects on comprehenders’ initial interpretations, however. It is unusual for sentence-initial subordinate clauses not to end with a comma or some other type of punctuation (searches in the parsed Brown corpus put the rate at about 18%); empirically it has consistently been found that a comma eliminates the garden-path effect in NP/Z sentences:

(3) While Mary was mending, the socks fell off her lap.

Understanding sentences like (3) is intuitively much easier, and reading times at the disambiguating verb are reliably lower when compared with (2). Fodor (2002) summarized the power of this effect succinctly:

[w]ith a comma after *mending*, there would be no syntactic garden path left to be studied. (Fodor, 2002)

In a surprisal model with clean, veridical input, Fodor’s conclusion is exactly what is predicted: separating a verb from its direct object with a comma effectively never happens in edited, published written English, so the conditional probability of the **NP** analysis should be close to zero.² When uncertainty about surface input is introduced, however—due to visual noise, imperfect memory representations, and/or beliefs about possible speaker error—analyses come into play in which some parts of the true string are treated as if they were absent. In particular, because the two sentences are perceptual neighbors, the pre-disambiguation garden-path analysis of (2) may be entertained in (3).

We can get a tighter handle on the effect of input uncertainty by extending Levy (2008b)’s analysis of the *expected* beliefs of a comprehender about the sequence of words constituting an input sentence to joint inference over both sentence identity and sentence structure. For a true sentence \mathbf{w}^* which yields perceptual input I , joint inference on sentence identity \mathbf{w} and structure T marginalizing over I yields:

$$P_C(T, \mathbf{w} | \mathbf{w}^*) = \int_I P_C(T, \mathbf{w} | I, \mathbf{w}^*) P_T(I | \mathbf{w}^*) dI$$

where $P_T(I | \mathbf{w}^*)$ is the true model of noise (perceptual inputs derived from the true sentence) and $P_C(\cdot)$ terms reflect the comprehender’s linguistic knowledge and beliefs about the noise processes intervening between intended sentences and perceptual input. \mathbf{w}^* and \mathbf{w} must be conditionally independent given I since \mathbf{w}^* is not observed by the comprehender, giving us (through Bayes’ Rule):

$$P(T, \mathbf{w} | \mathbf{w}^*) = \int_I \frac{P_C(I | T, \mathbf{w}) P_C(T, \mathbf{w})}{P_C(I)} P_T(I | \mathbf{w}^*) dI$$

For present purposes we constrain the comprehender’s model of noise so that T and I are conditionally independent given \mathbf{w} , an assumption that can be relaxed in future work.³ This allows us the further

²A handful of VP \rightarrow V , NP . . . rules can be found in the Penn Treebank, but they all involve appositives (*It* [_{VP} *ran, this apocalyptic beast . . .*]), vocatives (*You should* [_{VP} *understand, Jack, . . .*]), cognate objects (*She* [_{VP} *smiled, a smile without humor*]), or indirect speech (*I* [_{VP} *thought, you nasty brute . . .*]); none involve true direct objects of the type in (3).

³This assumption is effectively saying that noise processes are syntax-insensitive, which is clearly sensible for environmental noise but would need to be relaxed for some types of speaker error.

simplification to

$$P(T, \mathbf{w} | \mathbf{w}^*) = \overbrace{P_C(T, \mathbf{w})}^{(i)} \int_I \overbrace{\frac{P_C(I | \mathbf{w}) P_T(I | \mathbf{w}^*)}{P_C(I)}}^{(ii)} dI \quad (\text{II})$$

That is, a comprehender’s average inferences about sentence identity and structure involve a tradeoff between (i) the prior probability of a grammatical derivation given a speaker’s linguistic knowledge and (ii) the fidelity of the derivation’s yield to the true sentence, as measured by a combination of true noise processes and the comprehender’s beliefs about those processes.

3.1 Inducing hallucinated garden paths through manipulating prior grammatical probabilities

Returning to our discussion of the NP/Z ambiguity, the relative ease of comprehending (3) entails an interpretation in the uncertain-input model that the cost of infidelity to surface input is sufficient to prevent comprehenders from deriving strong belief in a *hallucinated* garden-path analysis of (3) pre-disambiguation in which the comma is ignored. At the same time, the uncertain-input theory predicts that if we manipulate the balance of prior grammatical probabilities $P_C(T, \mathbf{w})$ strongly enough (term (i) in Equation (II)), it may shift the comprehender’s beliefs toward a garden-path interpretation. This observation sets the stage for our experimental manipulation, illustrated below:

- (4) As the soldiers marched, toward the tank lurched an injured enemy combatant.

Example (4) is qualitatively similar to (3), but with two crucial differences. First, there has been LOCATIVE INVERSION (Bolinger, 1971; Bresnan, 1994) in the main clause: a locative PP has been fronted before the verb, and the subject NP is realized postverbally. Locative inversion is a low-frequency construction, hence it is crucially disfavored by the comprehender’s prior over possible grammatical structures. Second, the subordinate-clause verb is no longer transitive, as in (3); instead it is intransitive but could itself take the main-clause fronted PP as a dependent. Taken together, these properties should shift comprehenders’ posterior infer-

ences given prior grammatical knowledge and pre-disambiguation input more sharply than in (3) toward the input-unfaithful interpretation in which the immediately preverbal main-clause constituent (*toward the tank* in (4)) is interpreted as a dependent of the subordinate-clause verb, as if the comma were absent.

If comprehenders do indeed seriously entertain such interpretations, then we should be able to find the empirical hallmarks (e.g., elevated reading times) of garden-path disambiguation at the main-clause verb *lurched*, which is incompatible with the “hallucinated” garden-path interpretation. Empirically, however, it is important to disentangle these empirical hallmarks of garden-path disambiguation from more general disruption that may be induced by encountering locative inversion itself. We address this issue by introducing a control condition in which a postverbal PP is placed within the subordinate clause:

- (5) As the soldiers marched *into the bunker*, toward the tank lurched an injured enemy combatant. [+PP]

Crucially, this PP fills a similar thematic role for the subordinate-clause verb *marched* as the main-clause fronted PP would, reducing the extent to which the comprehender’s prior favors the input-unfaithful interpretation (that is, the prior ratio $\frac{P(\text{marched into the bunker toward the tank} | \text{VP})}{P(\text{marched into the bunker} | \text{VP})}$ for (5) is much lower than the corresponding prior ratio $\frac{P(\text{marched toward the tank} | \text{VP})}{P(\text{marched} | \text{VP})}$ for (4)), while leaving locative inversion present. Finally, to ensure that sentence length itself does not create a confound driving any observed processing-time difference, we cross presence/absence of the subordinate-clause PP with inversion in the main clause:

- (6)
- a. As the soldiers marched, the tank lurched toward an injured enemy combatant. [Uninverted, –PP]
 - b. As the soldiers marched into the bunker, the tank lurched toward an injured enemy combatant. [Uninverted, +PP]

4 Model instantiation and predictions

To determine the predictions of our uncertain-input/surprisal model for the above sentence types, we extracted a small grammar from the parsed

TOP	→ S .	1.000000
S	→ INVERTED NP	0.003257
S	→ SBAR S	0.012289
S	→ SBAR , S	0.041753
S	→ NP VP	0.942701
INVERTED	→ PP VBD	1.000000
SBAR	→ INSBAR S	1.000000
VP	→ VBD RB	0.002149
VP	→ VBD PP	0.202024
VP	→ VBD NP	0.393660
VP	→ VBD PP PP	0.028029
VP	→ VBD RP	0.005731
VP	→ VBD	0.222441
VP	→ VBD JJ	0.145966
PP	→ IN NP	1.000000
NP	→ DT NN	0.274566
NP	→ NNS	0.047505
NP	→ NNP	0.101198
NP	→ DT NNS	0.045082
NP	→ PRP	0.412192
NP	→ NN	0.119456

Table 1: A small PCFG (lexical rewrite rules omitted) covering the constructions used in (4)–(6), with probabilities estimated from the parsed Brown corpus.

Brown corpus (Kučera and Francis, 1967; Marcus et al., 1994), covering sentence-initial subordinate clause and locative-inversion constructions.^{4,5} The non-terminal rewrite rules are shown in Table 1, along with their probabilities; of terminal rewrite rules for all words which either appear in the sentences to be parsed or appeared at least five times in the corpus, with probabilities estimated by relative frequency.

As we describe in the following two sections, un-

⁴Rule counts were obtained using `tgrep2/Tregex` patterns (Rohde, 2005; Levy and Andrew, 2006); the probabilities given are relative frequency estimates. The patterns used can be found at http://idiom.ucsd.edu/~rlevy/papers/ac12011/tregex_patterns.txt.

⁵Similar to the case noted in Footnote 2, a small number of VP → V , PP . . . rules can be found in the parsed Brown corpus. However, the PPs involved are overwhelmingly (i) set expressions, such as *for example*, *in essence*, and *of course*, or (ii) manner or temporal adjuncts. The handful of true locative PPs (5 in total) are all parentheticals intervening between the verb and a complement strongly selected by the verb (e.g., [*VP means*, *in my country*, *homosexual*]); none fulfill one of the verb’s thematic requirements.

certain input is represented as a weighted finite-state automaton (WFSA), allowing us to represent the incremental inferences of the comprehender through intersection of the input WFSA with the PCFG above (Bar-Hillel et al., 1964; Nederhof and Satta, 2003, 2008).

4.1 Uncertain-input representations

Levy (2008a) introduced the LEVENSCHTEIN-DISTANCE KERNEL as a model of the average effect of noise in uncertain-input probabilistic sentence comprehension; this corresponds to term (ii) in our Equation (II). This kernel had a single noise parameter governing scaling of the cost of considering word *substitutions*, *insertions*, and *deletions* are considered, with the cost of a word substitution falling off exponentially with Levenshtein distance between the true word and the substituted word, and the cost of word insertion or deletion falling off exponentially with word length. The distribution over the infinite set of strings \mathbf{w} can be encoded in a weighted finite-state automaton, facilitating efficient inference.

We use the Levenshtein-distance kernel here to capture the effects of perceptual noise, but make two modifications necessary for incremental inference and for the correct computation of surprisal values for new input: the distribution over already-seen input must be *proper*, and possible future inputs must be *costless*. The resulting weighted finite-state representation of noisy input for a true sentence prefix $\mathbf{w}^* = w_{1..j}$ is a $j + 1$ -state automaton with arcs as follows:

- For each $i \in 1, \dots, j$:
 - A *substitution* arc from $i - 1$ to i with cost proportional to $\exp[-\text{LD}(w', w_i) \gamma]$ for each word w' in the lexicon, where $\gamma > 0$ is a noise parameter and $\text{LD}(w', w_i)$ is the Levenshtein distance between w' and w_i (when $w' = w_i$ there is no change to the word);
 - A *deletion* arc from $i - 1$ to i labeled ϵ with cost proportional to $\exp[-\text{len}(w_i)/\gamma]$;
 - An *insertion* loop arc from $i - 1$ to $i - 1$ with cost proportional to $\exp[-\text{len}(w')/\gamma]$ for every word w' in the lexicon;
- A loop arc from j to j for each word w' in

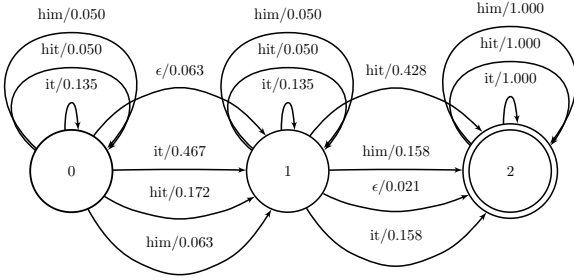


Figure 2: Noisy WFSAs for partial input *it hit...* with lexicon $\{it, hit, him\}$, noise parameter $\gamma=1$

the lexicon, with zero cost (value 1 in the real semiring);

- State j is a zero-cost final state; no other states are final.

The addition of loop arcs at state n allows modeling of *incremental* comprehension through the automaton/grammar intersection (see also Hale, 2006); and the fact that these arcs are costless ensures that the partition function of the intersection reflects only the grammatical prior plus the costs of input already seen. In order to ensure that the distribution over already-seen input is proper, we normalize the costs on outgoing arcs from all states but j .⁶ Figure 2 gives an example of a simple WFSAs representation for a short partial input with a small lexicon.

4.2 Inference

Computing the surprisal incurred by the disambiguating element given an uncertain-input representation of the sentence involves a standard application of the definition of conditional probability (Hale, 2001):

$$\log \frac{1}{P(I_{1..i}|I_{1..i-1})} = \log \frac{P(I_{1..i-1})}{P(I_{1..i})} \quad (\text{III})$$

Since our uncertain inputs $I_{1..k}$ are encoded by a WFSAs, the probability $P(I_{1..k})$ is equal to the partition function of the intersection of this WFSAs with the PCFG given in Table 1.⁷ PCFGs are a special class of weighted context-free grammars (WCFGs),

⁶If a state's total unnormalized cost of insertion arcs is α and that of deletion and insertion arcs is β , its normalizing constant is $\frac{\beta}{1-\alpha}$. Note that we must have $\alpha < 1$, placing a constraint on the value that γ can take (above which the normalizing constant diverges).

⁷Using the WFSAs representation of average noise effects here actually involves one simplifying assumption, that the av-

which are closed under intersection with WFSAs; a constructive procedure exists for finding the intersection (Bar-Hillel et al., 1964; Nederhof and Satta, 2003). Hence we are left with finding the partition function of a WCFG, which cannot be computed exactly, but a number of approximation methods are known (Stolcke, 1995; Smith and Johnson, 2007; Nederhof and Satta, 2008). In practice, the computation required to compute the partition function under any of these methods increases with the size of the WCFG resulting from the intersection, which for a binarized PCFG with R rules and an n -state WFSAs is Rn^2 . To increase efficiency we implemented what is to our knowledge a novel method for finding the minimal grammar including all rules that will have non-zero probability in the intersection. We first parse the WFSAs bottom-up with the item-based method of Goodman (1999) in the Boolean semiring, storing partial results in a chart. After completion of this bottom-up parse, every rule that will have non-zero probability in the intersection PCFG will be identifiable with a set of entries in the chart, but not all entries in this chart will have non-zero probability, since some are not connected to the root. Hence we perform a second, top-down Boolean-semiring parsing pass on the bottom-up chart, throwing out entries that cannot be derived from the root. We can then include in the intersection grammar only those rules from the classic construction that can be identified with a set of surviving entries in the final parse chart.⁸ The partition functions for each category in this intersection grammar can then be computed; we used a fixed-point method preceded by a topological sort on the grammar's ruleset, as described by Nederhof and Satta (2008). To obtain the surprisal of the input deriving from a word w_i in its context, we can thus com-

pute the surprisal of I_i , or $E_{P_T} \left[\log \frac{1}{P_C(I_i|I_{1..i-1})} \right]$, is well approximated by the log of the ratio of the expected probabilities of the noisy inputs $I_{1..i-1}$ and $I_{1..i}$, since as discussed in Section 3 the quantities $P(I_{1..i-1})$ and $P(I_{1..i})$ are expectations under the true noise distribution. This simplifying assumption has the advantage of bypassing commitment to a specific representation of perceptual input and should be justifiable for reasonable noise functions, but the issue is worth further scrutiny.

⁸Note that a standard top-down algorithm such as Earley parsing cannot be used to avoid the need for both bottom-up and top-down passes, since the presence of loops in the WFSAs breaks the ability to operate strictly left-to-right.

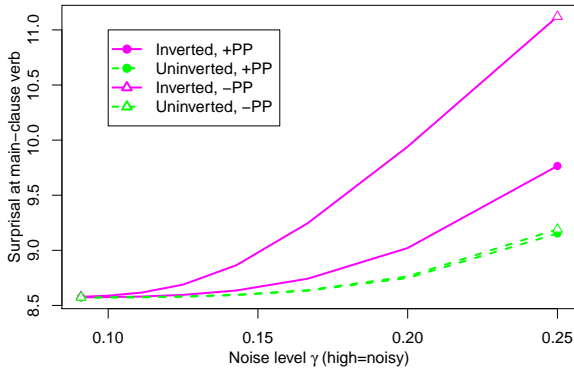


Figure 3: Model predictions for (4)–(6)

pute the partition functions for noisy inputs $I_{1\dots i-1}$ and $I_{1\dots i}$ corresponding to words $w_{1\dots i-1}$ and words $w_{1\dots i}$ respectively, and take the log of their ratio as in Equation (III).

4.3 Predictions

The noise level γ is a free parameter in this model, so we plot model predictions—the expected surprisal of input from the main-clause verb for each variant of the target sentence in (4)–(6)—over a wide range of its possible values (Figure 3). The far left of the graph asymptotes toward the predictions of *clean surprisal*, or noise-free input. With little to no input uncertainty, the presence of the comma rules out the garden-path analysis of the fronted PP *toward the tank*, and the surprisal at the main-clause verb is the same across condition (here reflecting only the uncertainty of verb identity for this small grammar). As input uncertainty increases, however, surprisal in the [**Inverted**, **-PP**] condition increases, reflecting the stronger belief given preceding context in an input-unfaithful interpretation.

5 Empirical results

To test these predictions we conducted a word-by-word self-paced reading study, in which participants read by pressing a button to reveal each successive word in a sentence; times between button presses are recorded and analyzed as an index of incremental processing difficulty (Mitchell, 1984). Forty monolingual native-English speaker participants read twenty-four sentence quadruplets (“items”) on the pattern of (4)–(6), with a Latin-square design so that each participant saw an equal

	Inverted	Uninverted
-PP	0.76	0.93
+PP	0.85	0.92

Table 2: Question-answering accuracy

number of sentences in each condition and saw each item only once. Experimental items were pseudo-randomly interspersed with 62 filler sentences; no two experimental items were ever adjacent. Punctuation was presented with the word to its left, so that for (4) the four and fifth button presses would yield

----- marched, -----

and

----- toward -----

respectively (right-truncated here for reasons of space). Every sentence was followed by a yes/no comprehension question (e.g., *Did the tank lurch toward an injured enemy combatant?*); participants received feedback whenever they answered a question incorrectly.

Reading-time results are shown in Figure 4. As can be seen, the model’s predictions are matched at the main-clause verb: reading times are highest in the [**Inverted**, **-PP**] condition, and there is an interaction between main-clause inversion and presence of a subordinate-clause PP such that presence of the latter reduces reading times more for inverted than for uninverted main clauses. This interaction is significant in both by-participants and by-items ANOVAs (both $p < 0.05$) and in a linear mixed-effects analysis with participants- and item-specific random interactions ($t > 2$; see Baayen et al., 2008). The same pattern persists and remains significant through to the end of the sentence, indicating considerable processing disruption, and is also observed in question-answering accuracies for experimental sentences, which are superadditively lowest in the [**Inverted**, **-PP**] condition (Table 2).

The inflated reading times for the [**Inverted**, **-PP**] condition beginning at the main-clause verb confirm the predictions of the uncertain-input/surprisal theory. Crucially, the input that would on our theory induce the comprehender to question the comma (the fronted main-clause PP)

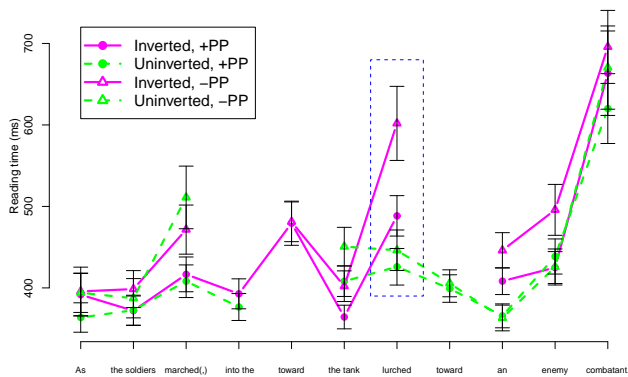


Figure 4: Average reading times for each part of the sentence, broken down by experimental condition

is not seen until after the comma is no longer visible (and presumably has been integrated into beliefs about syntactic analysis on veridical-input theories). This empirical result is hence difficult to accommodate in accounts which do not share our theory’s crucial property that comprehenders can revise their belief in previous input on the basis of current input.

6 Conclusion

Language is redundant: the content of one part of a sentence carries predictive value both for what will precede and what will follow it. For this reason, and because the path from a speaker’s intended utterance to a comprehender’s perceived input is noisy and error-prone, a comprehension system making optimal use of available information would use current input not only for forward prediction but also to assess the veracity of previously encountered input. Here we have developed a theory of how such an adaptive error-correcting capacity is a consequence of noisy-channel inference, with a comprehender’s beliefs regarding sentence form and structure at any moment in incremental comprehension reflecting a balance between fidelity to perceptual input and a preference for structures with higher prior probability. As a consequence of this theory, certain types of sentence contexts will cause the drive toward higher prior-probability analyses to overcome the drive to maintain fidelity to input, undermining the comprehender’s belief in an earlier part of the input actually perceived in favor of an analysis unfaithful to part of the true input. If subsequent input strongly disconfirms this incorrect in-

terpretation, we should see behavioral signatures of classic garden-path disambiguation. Within the theory, the size of this “hallucinated” garden-path effect is indexed by the surprisal value under uncertain input, marginalizing over the actual sentence observed. Based on a model implementing theory we designed a controlled psycholinguistic experiment making specific predictions regarding the role of fine-grained grammatical context in modulating comprehenders’ strength of belief in a highly specific bit of linguistic input—a comma marking the end of a sentence-initial subordinate clause—and tested those predictions in a self-paced reading experiment. As predicted by the theory, reading times at the word disambiguating the “hallucinated” garden-path were inflated relative to control conditions. These results contribute to the theory of uncertain-input effects in online sentence processing by suggesting that comprehenders may be induced not only to entertain but to adopt relatively strong beliefs in grammatical analyses that require modification of the surface input itself. Our results also bring a new degree of nuance to surprisal theory, demonstrating that perceptual neighbors of true preceding input may need to be taken into account in order to estimate how surprising a comprehender will find subsequent input to be.

Beyond the domain of psycholinguistics, the methods employed here might also be usefully applied to practical problems such as parsing of degraded or fragmentary sentence input, allowing joint constraint derived from grammar and available input to fill in gaps (Lang, 1988). Of course, practical applications of this sort would raise challenges of their own, such as extending the grammar to broader coverage, which is delicate here since the surface input places a weaker check on overgeneration from the grammar than in traditional probabilistic parsing. Larger grammars also impose a technical burden since parsing uncertain input is in practice more computationally intensive than parsing clean input, raising the question of what approximate-inference algorithms might be well-suited to processing uncertain input with grammatical knowledge. Answers to this question might in turn be of interest for sentence processing, since the exhaustive-parsing idealization employed here is not psychologically plausible. It seems likely that human comprehension in-

volves approximate inference with severely limited memory that is nonetheless highly optimized to recover something close to the intended meaning of an utterance, even when the recovered meaning is not completely faithful to the input itself. Arriving at models that closely approximate this capacity would be of both theoretical and practical value.

Acknowledgments

Parts of this work have benefited from presentation at the 2009 Annual Meeting of the Linguistic Society of America and the 2009 CUNY Sentence Processing Conference. I am grateful to Natalie Katz and Henry Lu for assistance in preparing materials and collecting data for the self-paced reading experiment described here. This work was supported by a UCSD Academic Senate grant, NSF CAREER grant 0953870, and NIH grant 1R01HD065829-01.

References

- Adams, B. C., Clifton, Jr., C., and Mitchell, D. C. (1998). Lexical guidance in sentence processing? *Psychonomic Bulletin & Review*, 5(2):265–270.
- Baayen, R. H., Davidson, D. J., and Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4):390–412.
- Bar-Hillel, Y., Perles, M., and Shamir, E. (1964). On formal properties of simple phrase structure grammars. In *Language and Information: Selected Essays on their Theory and Application*. Addison-Wesley.
- Bever, T. (1970). The cognitive basis for linguistic structures. In Hayes, J., editor, *Cognition and the Development of Language*, pages 279–362. John Wiley & Sons.
- Bolinger, D. (1971). A further note on the nominal in the progressive. *Linguistic Inquiry*, 2(4):584–586.
- Boston, M. F., Hale, J. T., Kliegl, R., Patil, U., and Vasishth, S. (2008). Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam sentence corpus. *Journal of Eye Movement Research*, 2(1):1–12.
- Bresnan, J. (1994). Locative inversion and the architecture of universal grammar. *Language*, 70(1):72–131.
- Christianson, K., Hollingworth, A., Halliwell, J. F., and Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, 42:368–407.
- Connine, C. M., Blasko, D. G., and Hall, M. (1991). Effects of subsequent sentence context in auditory word recognition: Temporal and linguistic constraints. *Journal of Memory and Language*, 30(2):234–250.
- Demberg, V. and Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, 109(2):193–210.
- Ferreira, F. and Henderson, J. M. (1993). Reading processes during syntactic analysis and reanalysis. *Canadian Journal of Experimental Psychology*, 16:555–568.
- Fodor, J. D. (2002). Psycholinguistics cannot escape prosody. In *Proceedings of the Speech Prosody Conference*.
- Frank, S. L. (2009). Surprisal-based comparison between a symbolic and a connectionist model of sentence processing. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, pages 1139–1144.
- Frazier, L. (1979). *On Comprehending Sentences: Syntactic Parsing Strategies*. PhD thesis, University of Massachusetts.
- Frazier, L. and Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14:178–210.
- Goodman, J. (1999). Semiring parsing. *Computational Linguistics*, 25(4):573–605.
- Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the Second Meeting of the North American Chapter of the Association for Computational Linguistics*, pages 159–166.
- Hale, J. (2006). Uncertainty about the rest of the sentence. *Cognitive Science*, 30(4):609–642.

- Hill, R. L. and Murray, W. S. (2000). Commas and spaces: Effects of punctuation on eye movements and sentence parsing. In Kennedy, A., Radach, R., Heller, D., and Pynte, J., editors, *Reading as a Perceptual Process*. Elsevier.
- Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, 20(2):137–194.
- Kučera, H. and Francis, W. N. (1967). *Computational Analysis of Present-day American English*. Providence, RI: Brown University Press.
- Lang, B. (1988). Parsing incomplete sentences. In *Proceedings of COLING*.
- Levy, R. (2008a). Expectation-based syntactic comprehension. *Cognition*, 106:1126–1177.
- Levy, R. (2008b). A noisy-channel model of rational human sentence comprehension under uncertain input. In *Proceedings of the 13th Conference on Empirical Methods in Natural Language Processing*, pages 234–243.
- Levy, R. and Andrew, G. (2006). Tregex and Tsurgeon: tools for querying and manipulating tree data structures. In *Proceedings of the 2006 conference on Language Resources and Evaluation*.
- Levy, R., Bicknell, K., Slattery, T., and 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.
- Marcus, M. P., Santorini, B., and Marcinkiewicz, M. A. (1994). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19(2):313–330.
- Mitchell, D. C. (1984). An evaluation of subject-paced reading tasks and other methods for investigating immediate processes in reading. In Kieras, D. and Just, M. A., editors, *New methods in reading comprehension*. Hillsdale, NJ: Erlbaum.
- Mitchell, D. C. (1987). Lexical guidance in human parsing: Locus and processing characteristics. In Coltheart, M., editor, *Attention and Performance XII: The psychology of reading*. London: Erlbaum.
- Narayanan, S. and Jurafsky, D. (1998). Bayesian models of human sentence processing. In *Proceedings of the Twelfth Annual Meeting of the Cognitive Science Society*.
- Narayanan, S. and Jurafsky, D. (2002). A Bayesian model predicts human parse preference and reading time in sentence processing. In *Advances in Neural Information Processing Systems*, volume 14, pages 59–65.
- Nederhof, M.-J. and Satta, G. (2003). Probabilistic parsing as intersection. In *Proceedings of the International Workshop on Parsing Technologies*.
- Nederhof, M.-J. and Satta, G. (2008). Computing partition functions of PCFGs. *Research on Logic and Computation*, 6:139–162.
- Roark, B., Bachrach, A., Cardenas, C., and Pallier, C. (2009). Deriving lexical and syntactic expectation-based measures for psycholinguistic modeling via incremental top-down parsing. In *Proceedings of EMNLP*.
- Rohde, D. (2005). *TGrep2 User Manual*, version 1.15 edition.
- Smith, N. A. and Johnson, M. (2007). Weighted and probabilistic context-free grammars are equally expressive. *Computational Linguistics*, 33(4):477–491.
- Smith, N. J. and 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*.
- Staub, A. (2007). The parser doesn't ignore intransitivity, after all. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 33(3):550–569.
- Stolcke, A. (1995). An efficient probabilistic context-free parsing algorithm that computes prefix probabilities. *Computational Linguistics*, 21(2):165–201.
- Sturt, P., Pickering, M. J., and Crocker, M. W. (1999). Structural change and reanalysis difficulty in language comprehension. *Journal of Memory and Language*, 40:136–150.
- Tabor, W. and Hutchins, S. (2004). Evidence for self-organized sentence processing: Digging in effects. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 30(2):431–450.

van Gompel, R. P. G. and Pickering, M. J. (2001).
Lexical guidance in sentence processing: A note
on Adams, Clifton, and Mitchell (1998). *Psycho-
nomic Bulletin & Review*, 8(4):851–857.