Probabilistic Linguistic Expectations, Uncertain Input, and Implications for Eye Movements in Reading

Roger Levy

(*University of California, San Diego*)

One nearly ubiquitous assumption in models of linguistic comprehension and of eye movement control in reading alike is of partial modularization between word-level and sentence-level processing: that the outcome of word recognition, and thus the input to sentence-level comprehension, is a categorial representation. Yet such a partial modularization throws away residual uncertainty regarding word identity that might potentially be of value to the comprehender further downstream in the sentence. Here I describe a line of research combining computational modeling with experimental eye-tracking work to explore the consequences of removing this partial modularity assumption.

Introduction

As you begin reading this article, your eyes jump rapidly across the page, roughly four times a second (Rayner, 1998, 2009). But, really, why do your eyes move at all?

This article explores some of the consequences of the idea that eye movements in reading can be best understood as an adaptive response to the fundamental problem of action in an uncertain environment that is posed by reading. From this perspective, the conclusion is inescapable that eye movements in reading reflect the goal of *obtaining information from the text*. Consider: before one reads a given sentence, one possesses knowledge in the face of uncertainty. The author might have written *anything*, but some sentences are more likely to have been written than others. The system governing movement of the eyes presumably also "knows" that perceptual input is constrained and noisy. The process of reading involves using this noisy perceptual input in conjunction with one′s prior knowledge of one′s language and one′s environment in order to figure out what the author wrote, and to infer meaning from this writing. Both input and prior knowledge inform decisions about where to move the eyes next; this iterative process of input acquisition, integration with prior knowledge, decision-making, and action unfolds rapidly in real time. This article is about integrating methods and ideas from computational linguistics and psycholinguistics in order to gain insight into this process.

Grammatical Knowledge

In particular, in this article I focus heavily on the profound role played by *grammatical knowledge* in the real-time understanding of written language. It is easy to demonstrate how powerful grammatical knowledge can be: just reflect on the fact that the sentence *dog bites man* describes an everyday event, whereas *man bites dog* is newsworthy. The field of generative syntax gives us formal tools for expressing this knowledge; one particularly useful yet simple tool is the *context-free grammar* (CFG; Chomsky, 1956; Ginsburg, 1966; Hopcroft & Ullman, 1979), which consists of rewrite rules such as (for English):

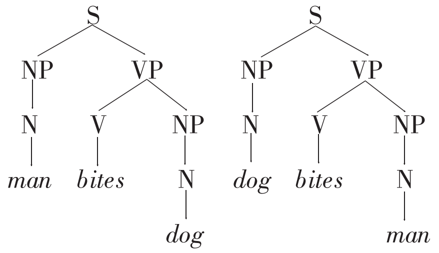
(1) S → NP VP [a **S**entence can consist of a **N**oun **P**hrase subject followed by a **V**erb **P**hrase]
NP → N [a **N**oun **P**hrase can consist of a **N**oun]
VP → V NP [a **V**erb **P**hrase can consist of a **V**erb followed by a **N**oun **P**hrase object]

Received date: 2010-11-27

Correspondence concerning this article should be addressed to Roger Levy, Department of Linguistics, University of California at San Diego, 9500 Gilman Drive #0108, La Jolla, CA92093-0108, USA, rlevy@ucsd.edu

Context-free grammars generate tree-structured descriptions of the syntactic form of sentences, and important aspects of the meaning of the sentence—such as, in this case, who did the biting and who got bitten—can be read off these descriptions:

(2)



Furthermore, seemingly large cross-linguistic differences in the grammatical properties of languages can often be elegantly captured by small differences in the CFG description of the language. For example, in both English and Chinese word order plays a relatively large role in encoding the grammatical functions played by each word in a sentence; yet consider the systematic differences in word order in the two sentence pairs below:

(3) a. I lectured in Tianjin.
b. 我 在 天津 报告
I in Tianjin lecture

(4) a. They expressed enthusiasm to me.
b. 他们 对 我 表示 热情
They to me express enthusiasm

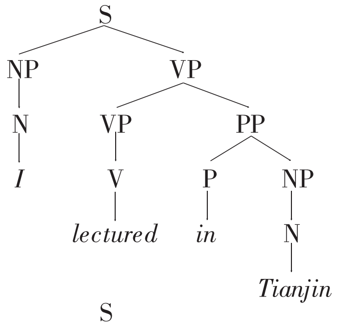
If we think of sentences as raw sequences of words, the systematicity of the differences between these English and Chinese examples isn't obvious at all, since no words are shared between examples (3) and (4). But if we think of sentences as structured grammatical representations, the systematicity becomes overwhelming: whereas in English, verb-modifying **P**repositional **P**hrases follow the verb, in Chinese they precede the verb.[1] The phrase structure rule sets below capture this difference succinctly:

(5)

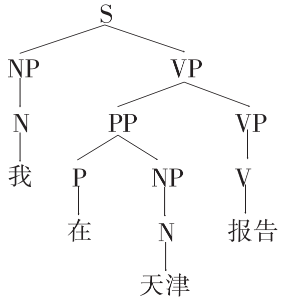
English	Chinese
S → NP VP	S → NP VP
NP → N	NP → N
VP → V NP	VP → V NP
VP → V	VP → V
VP → VP PP	VP → PP V

The first three rules in each language should be familiar from (1) above, and are the same in both languages. The fourth rule is also the same in both languages, and states that a verb phrase can consist of only a verb. The fifth rule is the only one that is different between the two languages: in English, it states that a verb phrase can (recursively) consist of a verb phrase followed by a repositional phrase, whereas in Chinese, it states that a verb phrase can (recursively) consist of a prepositional phrase followed by a verb phrase. These CFGs give rise to the following structural descriptions of (3) and (4):

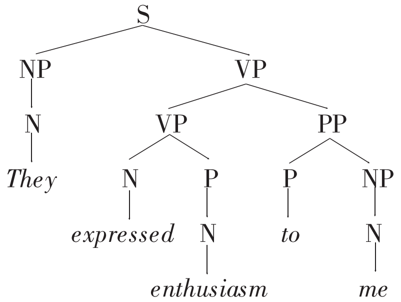
(6) a.



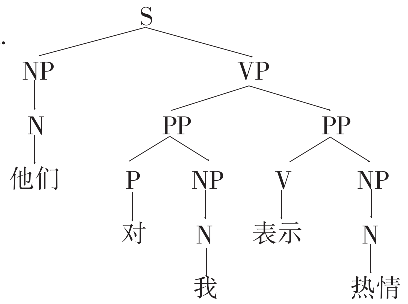
b.



c.



d.



[1] This point ignores a certain degree of controversy regarding whether examples like *zai Tianjin* are best analyzed as prepositional phrases or as something more like seral-verb construtions, which is not pertinent to the goals of this article.

Incrementality, Rationality, and Grammar

One of the outstandingly interesting properties of real-time sentence comprehension that must be included in any serious model of eye-movement control in reading is the *incrementality* of sentence comprehension. This property essentially states that humans don′t wait until a sentence is complete to draw inferences about what it might mean, or even how it might continue. An outstanding demonstration of this comes from another methodology in which eye movements are monitored as a window into language comprehension: the *visual world paradigm*. As a particularly clear example from Altmann and Kamide (1999), suppose that the comprehender is presented with a visual display including a boy; a movable, edible object (such as a cake); and a number of movable, inedible objects (such as a ball and a toy car). If the comprehender hears the sentence onset *The boy will* ***eat***..., when they hear the verb *eat* they initiate differentially more eye movements to the edible object on the display than in an alternative variant with a less restrictive verb *The boy will* ***move***.... This and related findings (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Kamide, Altmann, & Haywood, 2003) demonstrate incrementality: rapid access to detailed syntactic and semantic properties of the words encountered thus far. Furthermore, comprehenders generally use this information in the “right way”: the word *eat*, for example, is used to narrow down the set of possible upcoming referents to those in the environment which are edible. I will call this latter property the *rational* use of information in online language comprehension.

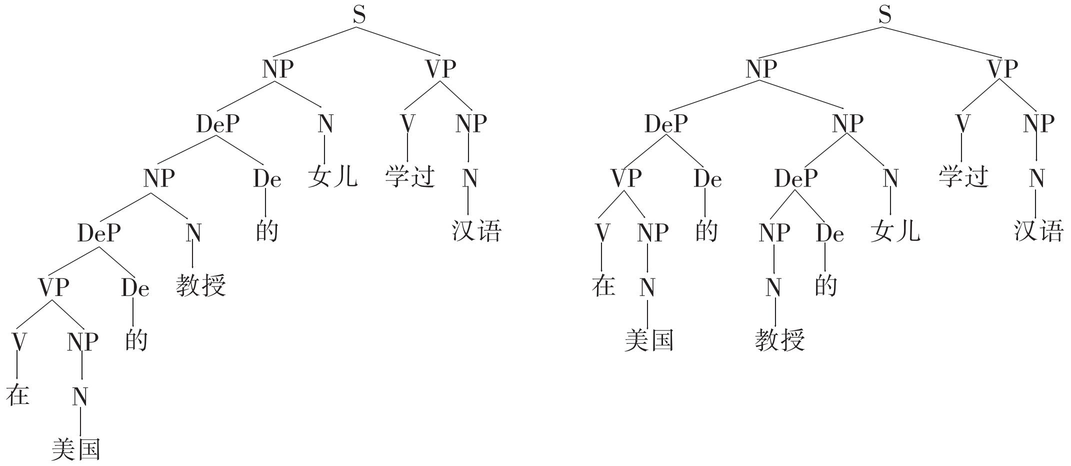
How do incrementality and rationality interact with the deployment of grammatical knowledge in online comprehension? When considering this problem, it is important to note that in any language, most sentences are afforded more than one possible interpretation by the grammar. For example, the Chinese sentence in (6b) below has two interpretations, depending on who′s considered to be in America.[2] (This ambiguity is preserved in the English translation.)

(7) 在 美国 的 教授 的 女儿 学过 汉语
in America DE professor DE daughter has studied Mandarin
"The daughter of the professor in America has studied Mandarin."

When the rules given in (8) below are added to our Chinese grammar fragment, it gives rise to the two structural descriptions seen in (9). In the left-hand description, it is the professor who is in America, as can be seen from the fact that there is an NP node dominating the substring 在 美国 的 教授 (*in America DE professor*); in the right-hand description, it is the daughter who is in America.

(8) NP → DeP NP [an NP can recursively consist of a premodifying **D**e–**P**hrase followed by an NP]
DeP → VP De [a DeP can consist of a VP followed by the word 的 ‘de’]
DeP → NP De [a DeP can consist of an NP followed by the word 的 ‘de’]

(9)



[2] The Chinese grammatical morpheme 的, glossed here as DE (its Pinyin form is *de*), marks all types of nominal premodifiers. In Example (6b), its first instance is effectively semantically emptyempty; its second instance carries the meaning of possession and thus is translated as English *of*.

Although in this particular sentence, one can make the case that there is no strong preference for one interpretation over the other, usually there are preferences which are relatively consistent across native speakers. For example, both English sentences in (10) below involve an ambiguity as to what the prepositional phrase *on the beach* modifies. Most native speakers prefer an interpretation of (10a) in which it modifies *discussed*, but an interpretation of (10b) in which it modifies *dogs* (Ford, Bresnan, & Kaplan, 1982; Jurafsky, 1996).

(10) a. The women discussed the dogs on the beach.
b. The women kept the dogs on the beach.

We might thus expect an incremental, rational system for sentence comprehension to form and update preferences about these grammatical interpretations both rapidly and in a manner consistent with the information sources available in the sentence thus far-including lexical, syntactic, semantic, and even pragmatic information sources. Paradoxically, one of the side effects of this type of incremental preference update is that certain types of sentences can temporarily *mislead* the comprehender, when the material early on in a sentence gives rise to a strong preference for a grammatical interpretation that ultimately turns out to be incorrect. Consider the examples in (11) below:

(11) a. 最 能干 的 领导 国家 前进 。
most capable DE lead country advance.
"The most capable lead the country forward."
b. The excellent play the fool rarely.

When reading these sentences, native speakers of Chinese and English typically experience confusion at the words 国家 ('country') and *the* respectively. This confusion reflects the ability of the preceding words, 领导 ('lead' or 'leader') and *play*, to function either as nouns or as verbs. The context biases the comprehender toward a nominal interpretation of these words—more precisely, as the head noun of the sentential subject but the grammars of Chinese and English also allow omission of the head noun of an NP, and this omitted variant turns out to be the globally correct interpretation, with intended meanings approximately the same as those seen in (12) below:

(12) a. 最 能干 的 人 领导 国家 前进 。
most capable DE **person** lead country advance.
"The most capable people lead the country forward"
b. The excellent **people** play the fool rarely.

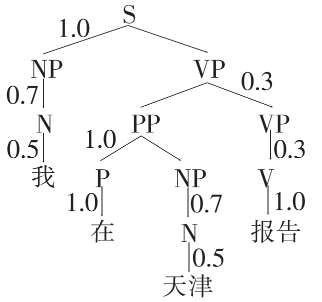
This early misinterpretation is known as a GARDEN-PATH effect. Its signature on eye movements in reading is elevation of first-pass reading times and first-pass regression probability, and has been known since Frazier and Rayner (1982). A desirable goal for deepening our understanding of eye movements in reading would thus be to integrate into models of eye movement control models of this incremental process of preference formation and update for grammatical interpretation. The leading candidate class of such models is the family of PROBABILISTIC GRAMMARS from the field of computational linguistics. Perhaps the most widely used member of this family is the PROBABILISTIC CONTEXT-FREE GRAMMAR (PCFG; Booth, 1969; Jurafsky & Martin, 2008; Manning & Schütze, 1999). A PCFG is exactly like the CFGs seen earlier in Section 2, except that each rewrite rule comes with a probability—a real number between zero and 1 whose interpretation is the likelihood with which the category on the left-hand side will rewrite as the sequence of categories on the right-hand side. If the probability of a verb phrase consisting of a verb and a noun phrase is 0.4, for example, it would be written as $P(VP \rightarrow V\ NP)= 0.4$, or equivalently $P(V\ NP|VP)= 0.4$. This latter formulation states explicitly that the *conditional probability* of the category sequence V NP *given* the presence of a VP is 0.4—this probability only matters if a VP has already been generated by some other rule in the grammar. In a PCFG, rule probabilities are constrained such that the probabilities of all rules with the same category on the left-hand side should sum to 1. For example, (13) below gives a PCFG fragment of Chinese grammar incorporating all the rules we have seen thus far. (In reality, a grammar covering a large fragment of a language would contain tens of thousands of rules, not the handful seen in an example such as this.)

A PCFG defines a PROBABILITY DISTRIBUTION over tree-structured descriptions of sentences: the probability of a tree is the product of the probabilities of the rules used to derive the tree. Example (14) depicts the Chinese tree in (6b), decorated with the probabilities of each subtree as specified by the grammar in (13), as well as the total probability of this particular tree.

(13)

Rule	Probability	Rule	Probability
S →NP VP	1.0	De →的	1.0
NP →N	0.7	VP→V	0.3
N →我	0.5	V →报告	1.0
N →天津	0.5	VP→V NP	0.4
NP →DeP NP	0.3	VP→PP VP	0.3
DeP →VP De	0.5	PP →P NP	1.0
DeP →NP De	0.5	P →在	1.0

(14)



P (Tree) =1.0×0.7×0.5×0.3×1.0×1.0×0.7×0.5×0.3×1.0=0.011025

For a given PCFG, the PROBABILITY OF A STRING is the sum of the probabilities of all the trees generating that string; and the PROBABILITY OF A STRING PREFIX is the sum of the probabilities of all the trees generating a string starting with that prefix.

According to a number of models of online sentence comprehension, incremental and rational comprehension of a sentence involves computing (or approximating) the probability distribution over trees given a probabilistic grammar and the words in a sentence thus far (Crocker & Brants, 2000; Hale, 2001, 2003, 2006; Jurafsky, 1996; Levy, 2008a; Narayanan & Jurafsky, 1998, 2002; Roark, 2001). What is a garden-path effect in such a model? An answer to this question can be found by considering that the number of possible grammatical analyses of a given sentence grows exponentially as sentence length increases (Church & Patil, 1982). Although there are algorithms that allow this exponentially increasing set of analyses to be computed efficiently-in time increasing *cubically* with sentence length (Earley, 1970) —this is not sufficient to explain the speed of human sentence comprehension, as the time we spend on comprehending a sentence is essentially *linear* in its length. In all realistic models of sentence comprehension, then, attention is restricted to some subset of the logically possible analyses. One way of doing so within a probabilistic framework is to discard possible analyses that have low probability. This can be done either deterministically (Crocker & Brants, 2000; Jurafsky, 1996; Narayanan & Jurafsky, 1998; Roark, 2001) or it can be done stochastically (Levy, Reali, & Griffiths, 2009). An example of how both gardenpath recovery and garden-path failure can take place for a locally ambiguous sentence is shown in Figure 1, which shows the evolving probabilities of two alternate analyses of each word, as well as the probability of successfully finding *some* analysis through each word of the sentence (bottom row), using a grammar learned from the parsed Brown corpus of English (Kučera& Francis, 1967; Marcus, Santorini, & Marcinkiewicz, 1994). At the word *play*, for example, the parser typically devotes most of its resources to the garden-path analysis (the top tree), and thus fails to find any analysis at the next word, *the*, almost half the time. Note that the garden-path analysis can in fact be saved at the word *the* by hypothesizing a relative clause, as in the possible continuation *The excellent play the author wrote was never performed.* This possible continuation accounts for the fact that the true posterior probability of the garden-path analysis remains considerable after this point, and for the fact that true comprehension failure can happen later than the first point of gardenpath disambiguation: sometimes at *rarely* and quite often at the end of the sentence. This gradualness of garden-path disambiguation is a property more difficult to obtain in serial and deterministic limited-parallel models.

Incremental probabilistic inference and eye movements

How does this process of incremental probabilistic grammatical inference relate to the eye movement patterns seen in reading? In addition to the effects of garden-path disambiguation just described, another crucial phenomenon in probabilistic inference is *prediction* of upcoming words and syntactic events. It has been known since Ehrlich and Rayner (1981) that highly

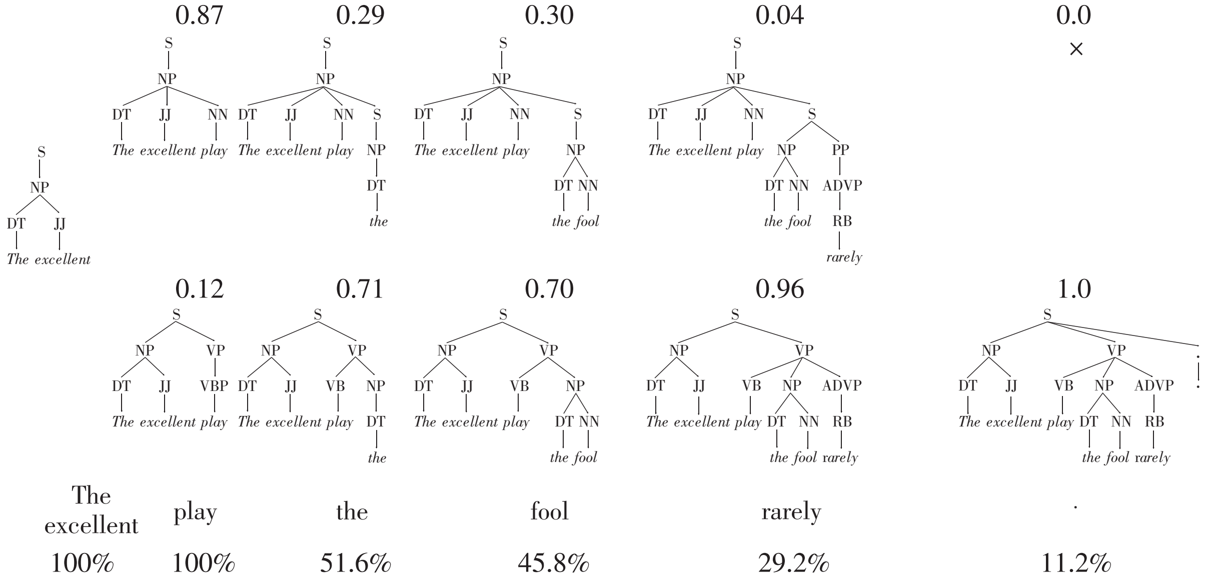


Figure 1. Incremental parsing of a garden-path sentence. Trees indicate the canonical structures for nominal (above) and verbal (below) interpretations of the ambiguous word *play*. Numbers above the trees indicate the true probabilites in the grammar of nominal and verbal interpretations after each word. Numbers in the second-to-last line indicate the frequency with which a 30–particle incremental parser (Levy, Reali, & Griffiths, 2009) produces a viable parse tree including the given word.

predictable words are skipped more often and fixated on more briefly than less predictable words. In computational psycholinguistics, one well-known formalization of this idea has been the proposal of SURPRISAL as a linking function between probability and the amount of time required to process a word in its context (Hale, 2001; Levy, 2008a). Surprisal is defined as the negative log of the conditional probability of a word *wi* in its context:

$$-\log\ P(w_i|w_{1,\cdots,i-1}),\text{CONTEXT})$$

Hence surprisal ranges from zero (an obligatory event, probability 1) to infinity (an impossible event, probability zero). To estimate the surprisal of a word in its context, one can use methods from computational linguistics (Chen & Goodman, 1998; Jelinek & Lafferty, 1991; Stolcke, 1995) or the more traditional Cloze sentence completion method (Taylor, 1953). It is a substantive claim that there is a linear relationship between surprisal and reading times, which can be derived from a number of possible considerations of optimality in the language comprehender (Norris, 2006; Smith & Levy, 2008). Because Cloze completion has been the norm for assessing word predictability in psycholinguistics, however, the precise functional form of the relationship between word probability and reading time has not been extensively investigated (though see Rayner & Well, 1996; Kliegl, Nuthmann, & Engbert, 2006): to obtain reasonably reliable estimates of predictability levels below about 0.1 would require hundreds or even thousands of participants in a Cloze norming study. To assess this functional form more precisely, Smith and Levy (2008, submitted) used broad-coverage techniques from computational linguistics to estimate probabilities for every word in the ten-participant, 50,000-word Dundee corpus, the largest available dataset of eye movements in reading. Figure 2 shows the partial contribution of word surprisal to current-word firstfixation durations, controlling for potential confounds including word frequency and length. Remarkably, an essentially linear relationship holds over several orders of magnitude between word log-probability and gaze duration. Furthermore, this relationship can be recovered at the individual subject level, speaking to the strength and systematicity of the relationship between predictability and fixation durations in reading.

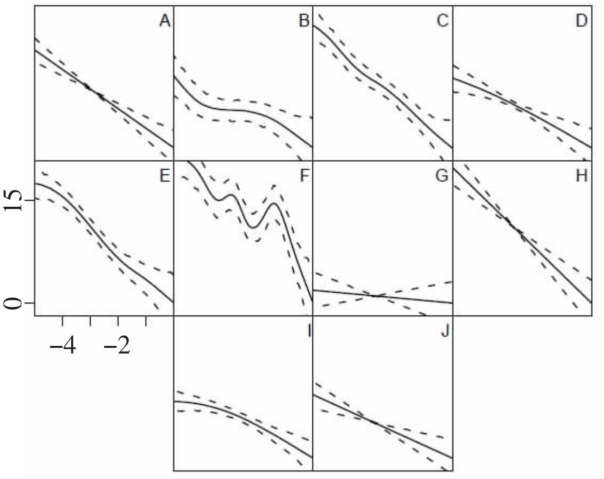


Figure 2. The linear relationship between a word's surprisal and first fixations on it, for each of the ten subjects in the Dundee corpus (Smith & Levy, 2008).

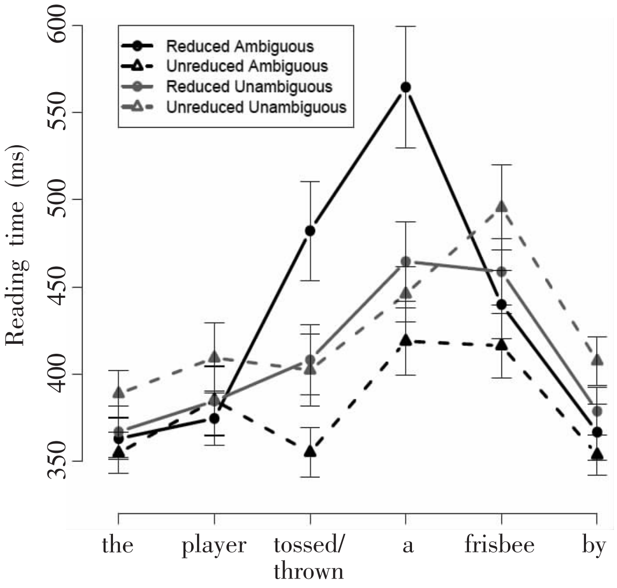


Figure 3. Self-paced reading data from Tabor et al. (2004)

Uncertain input: challenges and solutions for incrementality and rationality

Thus far the picture of comprehenders as highly incremental, rational users of available information may seem relatively compelling. However, a number of challenges remain for this picture, and in the remainder of the paper I describe one such challenge, recent modeling and experimental work we have done to meet that challenge, and new directions in which this modeling and experimental work has taken us.

Consider the sentence (15a) below:

(15) a. The coach smiled at the player tossed the frisbee.

This is a legitimate sentence according to the grammatical rules of English, but native speakers find it extremely difficult to read starting at the word *tossed*. This is in contrast with sentences (15b-d) below, which mean essentially the same thing. Word-by-word reading times in the self-paced reading study of Tabor et al. (2004), who originally demonstrated this phenomenon, are shown in Figure 3; note the localization of the superadditive difficulty effect at the critical word *tossed/thrown* and immediately thereafter.

(15) b. The coach smiled at the player **who was** tossed the frisbee.
c. The coach smiled at the player thrown the frisbee.
d. The coach smiled at the player **who was** thrown the frisbee.

The basic intuition regarding this sentence that first strikes most investigators is that the sequence of words *the player tossed* coheres well together as a potential subject-verb combination, but that this local interpretation does not fit with the global sentence context, in which *the player* is inside a prepositional phrase–not a place where the subject of a clause can appear. For this reason, Tabor et al. dubbed this a LOCAL-COHERENCE EFFECT. This intuition bears some resemblance to the garden-path effects described in Section 3, but the reason that the local-coherence effect is problematic for the rational, incremental theory we have developed up to this point is that whereas in garden-path sentences such as (11b) the incorrect analysis is legitimate and even favored given previous global context, in (15a) the incorrect analysis should be *ruled out* by previous global context.

One way of accounting for the local-coherence effect is to assume that there is a fast, "bottom-up" component to syntactic processing that can construct analyses heedless of global context, and there have been several formalizations of this idea (Tabor & Hutchins, 2004; Gibson, 2006; Bicknell & Levy, 2009; Morgan, Keller, & Steedman, 2010). However, this approach is faced with the paradox of having to account for how rapidly effectively human comprehenders *do* use global context to constrain interpretation of linguistic input. The theoretical challenge, then, is to reconcile this ability to use context in general with the apparent failure to use context to rule out irrelevant interpretations in local-coherence effects.

Levy (2008b) deals with this challenge by hypothesizing that comprehenders' representations of prior context may in fact be noisier than is often thought. On this model, the input to the sentence-level comprehension mechanism is not a sequence of words, but rather a set of noisy perceptual input representations *I*. The comprehender uses her grammatical and world knowledge to construct a joint *probability distribution* over possible word sequences *w* and possible structural analyses T through Bayesian inference:

$$P(T,\boldsymbol{w}|I)=\frac{P(I|T,\boldsymbol{w})P(T,\boldsymbol{w})}{P(I)}$$

We've already described how a probabilistic grammar places a probability distribution over structural analyses and word sequences—that is, how it determines *P (T,**w**)*. An additional component needs to be specified for this new model, however: the comprehender's model of perceptual noise (and, potentially, speaker error), *P (I|T,**w**)*. There are many possibilities for how to construct such a noise model; Levy (2008b) used one based on Levenshtein edit distance (Levenshtein, 1966), which essentially encodes the intuitions that orthographically similar words are more likely to be confused for one another, and that short words are more likely to be missed (or, inversely, perceived to be present when they're not) than long words.

How could this model account for the local-coherence effect seen in (15a)? Crucially, there are many small changes that could be made to this sentence which would make the critical word *tossed* into a main verb instead of a participial verb:

(16) a. The coach smiled {as/and} the player tossed the frisbee.
b. The coach smiled at the player {who/that/and} tossed the frisbee.

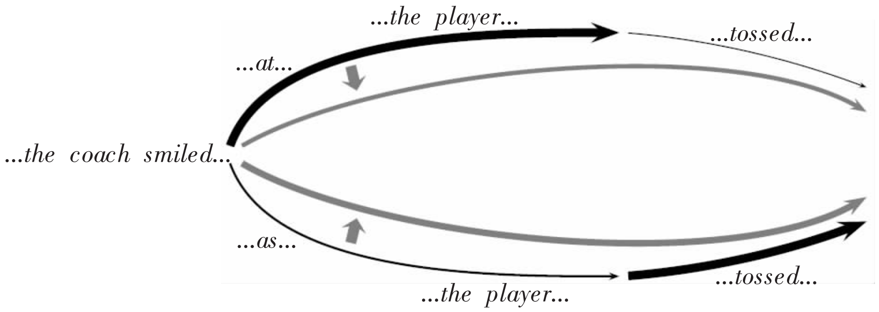


Figure 4. Coupling of inference about sentence identity and grammatical analysis: two paths through a local-coherence sentence.

c. The coach {who/that} smiled at the player tossed the frisbee.

Just as in syntactic disambiguation a word can dramatically change the comprehender's beliefs about the probability of different structural interpretations (e.g., the second instance of *the* in Figure 1), in a model where word-level representations are non-veridical, perceptual input midstream in a sentence can dramatically change the comprehender's beliefs about what *word sequences* may have preceded the current input. This possibility, and how it interacts with grammatical knowledge in the online comprehension of (15a), is illustrated in Figure 4. The perceptual neighborhood and the grammar of English effectively provide multiple different paths through the sentence; two of the most important paths are seen here, and their relative strengths (probabilities) are indicated with thicker line length for a path segment that is more probable given (a) having arrived to the start of the path segment, and (b) the input the segment covers. On the upper path, the word *at* is correctly identified; since perception on average is not biased to be inaccurate, through this path is initially more likely. On the lower path, the word *at* is incorrectly identified as one of its neighbors, *as* or *and*; since this path involves a perceptual error, it is less likely. Since *at* is a preposition whereas as and *and* are conjunctions, however, the different paths now correspond to different syntactic analyses of upcoming input. The phrase *the player* is of similar probability for both the upper and lower path sequences; hence the path strengths do not change appreciably at this point. But in order for the upper path to further continue with the word *tossed*, a highly unlikely syntactic event has to occur-a reduced relative clause involving passivization on the goal argument of a ditransitive verb—whereas the lower path can continue with the word *tossed* through the much likelier syntactic event of a main-clause finite verb. Therefore strength of the path segment covering *tossed* on the upper path is much lower than the strength of the corresponding segment on the lower path. Bayesian inference incorporates this new information to update the aggregate strengths of the upper and lower paths, causing a substantial shift in belief from the upper path to the lower path (gray arrows and new, longer path segments in Figure 4). On the account of Levy (2008b), the boggle in (15a) involves precisely this shift: the word *tossed* calls into question the comprehender's representation of past input.

Levy (2008b) modeled this shift in beliefs about past input by defining probability distributions over possible word sequences up to a position *i* in the sentence (in (15a) , *i* is where the word *tossed* is seen) before and after seeing perceptual input from position *i*. In the model, the magnitude of the shift is quantified by the KULLBACK-LEIBLER DIVERGENCE—a standard measure of one coarse-grained probabilty distribution encodes another, finer-grained distribution (Cover & Thomas, 1991) —from the distribution before seeing *i* to the distribution after seeing *i*. This quantity, the ERROR IDENTIFICATION SIGNAL (EIS) at position *i*, is shown for (15a) and (15) in Figure 5 (*at* conditions). There is one free parameter in this model, corresponding to the perceptual noise level λ, but regardless of the value of this parameter the EIS is larger at the part-of-speech ambiguous *tossed* than at the unambiguous *thrown*.

Because the local-coherence effect in the Levy (2008b) model of rational probabilistic comprehension under uncertain input depends crucially on the perceptual neighborhood of the sentence being read, the model makes a number of non-trivial empirical predictions regarding how manipulations of perceptual neighborhood may affect online comprehension. As shown in (16) and Figure 4, for example, an important part of the EIS in (15a) arises from the presence of the words as and/or and in the immediate perceptual neighborhood of the word at. The model thus predicts that if a semantically similar preposition without such a perceptual neighborhood, such as *toward*, is substituted for *at*, the EIS should be diminished (Figure 5). Levy, Bicknell, Slattery, and Rayner (2009) tested this prediction in an eye-tracking study by crossing the use of the preposition at versus toward with the part-of-speech ambiguity of the critical word (e.g., tossed versus thrown). The prediction was borne out in interactions on rate of first-pass regressions from the critical word, go-past time, and the frequency of fixating on the

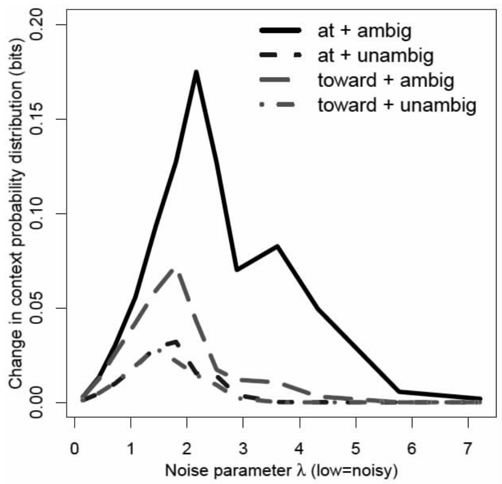
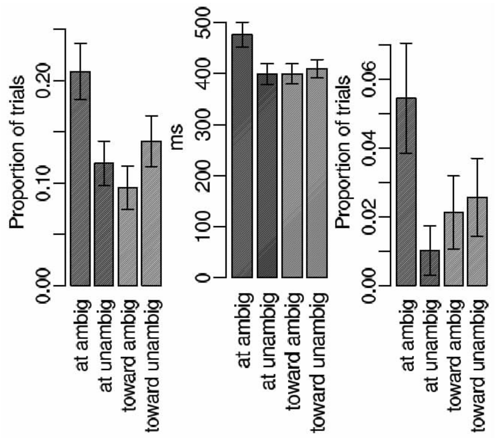


Figure 5. Error identification signal at critical word in local-coherence sentences.



(a) First-pass regressions (b) Go-past time (c) Go-past fixations on preposition

Figure 6. Crucial results of Levy, Bicknell, et al. (2009)

preposition during go-past reading of the critical word (Figure 6).

Future directions

In this article, I have attempted to present evidence for a view of incremental sentence comprehension as grammar-based, rational, predictive, and probabilistic; and to explicate the implications and support of such a view for eye movement control in reading. We have seen how the notion of incremental, probabilistic grammatical analysis can capture both coarse-grained and subtle garden-path effects, and how a highly regular relationship holds between prediction strength and fixation durations in first-pass reading. We have also examined local coherence effects, which pose an apparent challenge to the rational view articulated above, and explored the possibility that the rational view can be maintained if a key simplifying assumption common to all previous models of sentence comprehension—that of categorical word-level input—is relaxed. It turns out that relaxing this simplifying assumption and allowing input representations to be uncertain not only accounts for the classic local-coherence effect, but makes new predictions regarding the effects of perceptual neighborhood on eye movements in sentence reading.

The input-uncertainty model also leads to a number of other predictions that have been tested using controlled experiment and corpus analysis. Levy (2010) demonstrated the existence of *hallucinated garden-paths*, where strong prior expectations regarding likely and unlikely grammatical structures can bias comprehenders toward adopting incremental analyses that are not strictly licensed by surface input, as evidenced by garden-path disambiguation effects upon encountering downstream material consistent with the true surface input but not with the hallucinated garden path. Smith and Levy (2010) found evidence from eye movements in reading newspaper text that fixation durations during first-pass reading seem to reflect the word surprisals *expected* given a combination of prior context and coarse-grained, uncertain bottom-up input more strongly than they reflect true word surprisals. Additionally, taking input uncertainty into account has the potential to lead to a new framework in which eye movement control policies are seen as (near-) rational solutions to an optimization problem in which the goal of reading is to discern the contents of the sentence rapidly and with high accuracy. Bicknell and Levy (2010) presented this framework and a first implementation, using a combination of noisy-channel probabilistic grammatical inference and reinforcement learning (Russell & Norvig, 2003, Chapter 21; Sutton & Barto, 1998). They demonstrated that regressive eye movements can be viewed as a rational response to the above optimization problem: the occasional re-reading necessitated by setting the required level of first-pass identification accuracy only moderately high is an acceptable price to pay for the increased overall speed of reading it allows.

In summary, then, we have seen a picture of the deep intertwinement of sophisticated grammatical knowledge, information uptake from perceptual input, and decision-making regarding eye movements on rapid time scales in sentence comprehension during reading. We have seen evidence for surprisingly law-like behavior in several respects—including beliefs about sentence structure and identity being well-described by the integration of visual input with linguistic knowledge through the rules of probability theory, and a systematic relationship holding between the conditional probability of a word and how long it is fixated. Apparent difficulties for this picture have led to a more careful assessment of basic assumptions regarding the nature of the input representations in sentence-level comprehension, which in turn has led to richer models which make novel predictions. The larger lesson is that appropriately accounting for environmental and cognitive constraints in probabilistic models can sometimes lead to a more nuanced and ultimately more satisfactory picture of key aspects of human cognition. Thus far, this lesson has been strikingly true for eye movements in real-time linguistic comprehension; one can hope that it may continue to hold in this domain, and that it may hold in other cognitive domains as well.

Acknowledgments

This article is based on a talk given May 25, 2010, at the fourth China International Conference on Eye Movements at Tianjin Normal University, and I am grateful to feedback from the audience at that talk. All mistakes are my own. The writing of this article was supported by NSF CAREER grant 0953870 to Roger Levy.

References

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73* (3), 247–264.

Bicknell, K., & Levy, R. (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 association for computational linguistics-human language technologies (NAACL-HLT) conference.*

Bicknell, K., & Levy, R. (2010). A rational model of eye movement control in reading. In *Proceedings of the annual meeting of the association for computational linguistics* (pp.1168–1178).

Booth, T. L. (1969). Probabilistic representation of formal languages. In *Ieee conference record of the 1969 tenth annual symposium on switching and automata theory* (pp.74–81).

Chen, S., & Goodman, J. (1998). *An empirical study of smoothing techniques for language modeling* (Tech. Rep.). Computer Science Group, Harvard University.

Chomsky, N. (1956). Three models for the description of language. *IRE Transactions on Information Theory, 2* (3), 113–124.

Church, K., & Patil, R. (1982). Coping with syntactic ambiguity, or how to put the block in the box on the table. *Computational Linguistics, 8* (3–4), 139–149.

Cover, T., & Thomas, J. (1991). *Elements of information theory.* John Wiley.

Crocker, M., & Brants, T. (2000). Wide-coverage probabilistic sentence processing. *Journal of Psycholinguistic Research, 29* (6), 647–669.

Earley, J. (1970). An efficient context-free parsing algorithm. *Communications of the ACM, 13* (2), 94–102.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior, 20*, 641–655.

Ford, M., Bresnan, J., & Kaplan, R. M. (1982). A competence-based theory of syntactic closure. In J. Bresnan (Ed.), *The mental representation of grammatical relations* (pp. 727–796). The MIT Press.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14*, 178–210.

Gibson, E. (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.

Ginsburg, S. (1966). *The mathematical theory of context-free languages*. McGraw-Hill.

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*(pp. 159–166).

Hale, J. (2003). The information conveyed by words in sentences. *Journal of Psycholinguistic Research, 32* (2), 101–123.

Hale, J. (2006). Uncertainty about the rest of the sentence. *Cognitive Science, 30* (4), 609–642.

Hopcroft, J. E., & Ullman, J. D. (1979). *Introduction to automata theory, languages, and computation.* Addison-Wesley.

Jelinek, F., & Lafferty, J. D. (1991). Computation of the probability of initial substring 24 generation by stochastic context free grammars. *Computational Linguistics, 17* (3), 315–323.

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

Jurafsky, D., & Martin, J. H. (2008). *Speech and language processing: An introduction to natural lan guage processing, computational linguistics, and speech recognition* (Second ed.). Prentice-Hall.

Kamide, Y., Altmann, G., & Haywood, S. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language, 49* (1), 133–156.

Kliegl, R., Nuthmann, A., & Engbert, R. (2006). Tracking the mind during reading: The influence of past, present, and future words on fixation durations. *Journal of Experimental Psychology: General, 135* (1), 12–35.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady, 10*, 707–710.

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* (pp. 234–243).

Levy, R. (2010, March). *On hallucinated garden paths.* Poster presented at the CUNY sentence processing conference.

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

Levy, R., Reali, F., & Griffiths, T. L. (2009). Modeling the effects of memory on human online sentence processing with particle filters. In *Proceedings of the 22nd conference on neural information processing systems (nips).*

Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing.* MIT Press.

Marcus, M. P., Santorini, B., & Marcinkiewicz, M. A. (1994). Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics, 19* (2), 313–330.

Morgan, E., Keller, F., & Steedman, M. (2010). A bottom-up parsing model of local coherence effects. In *Proceedings of the cognitive science society conference.*

Narayanan, S., & Jurafsky, D. (1998). Bayesian mod els of human sentence processing. In *Proceedings of the twelfth annual meeting of the cognitive science society.*

Narayanan, S., & Jurafsky, D. (2002). A Bayesian model predicts human parse preference and reading time in sentence processing. In *Advances in neural information processing systems* (Vol.14, pp. 59–65).

Norris, D. (2006). The Bayesian Reader: Explaining word recognition as an optimal Bayesian decision process. *Psychological Review, 113* (2), 327–357.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124* (3), 372–422.

Rayner, K. (2009). The thirty–fifth Sir Frederick Bartlett Lecture: Eye movements and attention during reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology, 62*, 1457–1506.

Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review, 3* (4), 504–509.

Roark, B. (2001). Probabilistic top–down parsing and language modeling. *Computational Linguistics, 27* (2), 249–276.

Russell, S., & Norvig, P. (2003). *Artificial intelligence: A modern approach* (Second ed.). Prentice Hall.

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

Smith, N. J., & Levy, R. (2010). Fixation durations in first-pass reading reflect uncertainty about word identity. In *Proceedings of the 32nd annual meeting of the cognitive science society.*

Smith, N. J., & Levy, R. (Submitted). *A rational model of predictability effects on cognitive processing times.* (Manuscript, UC San Diego).

Stolcke, A. (1995). An efficient probabilistic context-free parsing algorithm that computes prefix probabilities. *Computational Linguistics, 21* (2), 165–201.

Sutton, R. S., & Barto, A. G. (1998). Reinforcement learning. MIT Press.

Tabor, W., Galantucci, B., & Richardson, D. (2004). Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language, 50* (4), 355–370.

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

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

Taylor, W. L. (1953). A new tool for measuring readability. *Journalism Quarterly, 30*, 415.