Down the Garden Path with Probabilistic CKY and Earley Parsers

LING 384 Final Project Report

Breanna K. Nguyen breanna.nguyen@yale.edu
Department of Psychology
Yale University
New Haven, CT 06511, USA

1 Introduction

1.1 Background & Existing Work

Humans process words in a sentence incrementally, leading us to form strong syntactic expectations before a sentence is complete. Garden path sentences exploit this property by leading readers toward an initially plausible but ultimately incorrect parse, resulting in processing difficulty when expectations are violated. Computational models of parsing offer a way to formalize and test these effects.

The Earley parser, originally developed for efficient syntactic parsing, was extended into a probabilistic framework that enables computation of prefix probabilities [5]. As such, it is especially suitable for modeling human sentence processing. The probabilistic Earley parser incorporates rule probabilities into the standard Earley algorithm and tracks the likelihood of partial derivations as the sentence unfolds. Importantly, it maintains a dynamic chart of predicted, scanned, and completed constituents, allowing it to compute the probability of each word given its preceding context. This forward-looking, top-down approach not only supports incremental interpretation but also provides a natural foundation for calculating word-by-word surprisal, a key predictor of processing difficulty in psycholinguistics. Indeed, surprisal values defined by the probabilistic Earley parser are found to be linked to human reading times [1].

The Cocke–Kasami–Younger (CKY) algorithm is a programming method for parsing sentences using a context-free grammar (CFG) in Chomsky Normal Form (CNF). It operates in a bottom-up manner, systematically filling a chart that records all possible constituents over spans of the input sentence. Starting from individual words, the algorithm combines smaller constituents into larger ones by applying binary grammar rules, ultimately determining whether a complete parse exists [6]. CKY is a practical choice for parsing in both theoretical and applied natural language processing tasks.

1.2 Research Question

This project investigates the performance of symbolic parsing algorithms in predicting processing difficulties that humans experience with garden path sentences. Specifically, it compares the performance of the probabilistic CKY and probabilistic Earley parsers on three classes of garden path sentences (MV/RR, NP/Z, and NP/S), each of which involves a temporary syntactic ambiguity that misleads the reader. By adapting both algorithms to compute word-by-word surprisal values, the project asks: to what extent do these parsing models reflect the difficulty associated with resolving these ambiguities? Through this comparison, the study aims to clarify how parsing strategy, bottom-up versus top-down, affects a model's sensitivity to garden path effects.

Different garden path sentence types create ambiguity through distinct structural mechanisms, which may interact differently with the parsing strategies of CKY and Earley.

- In NP/Z sentences (e.g., "because the nurse examined the patient recovered"), the ambiguity arises when a verb initially appears to take a direct object (NP) but is ultimately shown to take no object at all (Z) and is later revealed to be the subject of a second clause [2]. This structure induces a garden path effect by exploiting the parser's preference for simpler, more immediate attachments. Predictive parsers like Earley are more susceptible to this misanalysis and are expected to show higher surprisal at the disambiguating verb, while bottom-up parsers like CKY, which delay commitment, may produce more muted surprisal responses.

- In NP/S sentences (e.g., "The journalist suspected the senator lied"), the ambiguity arises from a noun phrase (NP) that can either serve as the direct object of the matrix verb or as the subject of an embedded sentential complement (S) [2]. Both parsers may show relatively modest surprisal responses at the disambiguating word, but Earley's ability to maintain multiple structural hypotheses in parallel may allow it to register subtle shifts in interpretive preference earlier than CKY.

- In MV/RR sentences (e.g., "The dude trained in the gym flexed"), the ambiguity arises from a reduced relative (RR) clause initially misinterpreted as a main verb (MV) construction [2]. Since Earley is a top-down parser that maintains predictive states, it is expected to register higher surprisal when the disambiguating verb ("flexed") forces a revision of the initial structure. CKY, by contrast, builds structure only after encountering all subparts and may delay the recognition of ambiguity, potentially underestimating processing difficulty until the full sentence is parsed.

2 Implementation

A full code and data repository can be found at `https://github.com/breannaknguyen/ling384-finalcode`.

2.1 Garden Path Grammars

To model the syntactic ambiguity specific to each garden path sentence type, I first constructed a separate probabilistic context-free grammar (PCFG) for each of the three sentence types: NP/Z, NP/S, and MV/RR. Each PCFG was designed to allow both the preferred (initially misinterpreted) and the disambiguated structures, enabling the parser to explore multiple syntactic interpretations as the sentence unfolds. Shorter, simpler sentences were also allowed to reflect certain expectations for sentence completions. To make the grammars realistic and data-driven, I estimated rule probabilities using relative frequency counts derived from the Penn Treebank [3]. For each nonterminal, I extracted the frequency of its rule expansions from Treebank-parsed sentences and normalized these counts to obtain conditional probabilities and manually adjusted them if necessary. This approach ensures that the grammars not only reflect the intended structural ambiguity but also encode plausible probabilistic biases based on attested usage in natural language.

For example, the PCFG for the MV/RR sentence type supports both the garden path sentence "The horse raced past the barn fell." and its disambiguated version "The horse who was raced past the barn fell." It also allows simpler, unambiguous sentences such as "The horse raced past the barn." and "The horse fell.", along with various lexical permutations involving different nouns and verbs. Similar PCFGs were constructed for the NP/Z and NP/S sentence types, each designed to generate both the ambiguous and disambiguated variants within that class.

2.2 Earley Parsing Algorithm

The Earley parser used in this project was adapted from code developed for a LING 384 assignment and based on the formulation by Stolcke (1995) [5]. This implementation allows for the computation of prefix probabilities, which are essential for deriving incremental surprisal values aligned with psycholinguistic theories of sentence processing.

The parser operates over a probabilistic context-free grammar (PCFG), represented as a dictionary mapping each nonterminal to a list of possible expansions with associated probabilities. Parsing proceeds left to right through a sentence, maintaining a chart of parse states indexed by position. Each chart entry contains information about a rule's left-hand side (LHS), right-hand side (RHS), the position of the dot indicating parsing progress, the index where the constituent began, and two probabilities: the inside probability (γ) and the forward probability (α), used in computing prefix probabilities.

Parsing involves three core operations: prediction, scanning, and completion. Prediction expands nonterminals that are expected next, scanning advances the parse when a terminal matches the next input word, and completion connects completed constituents to their parents. The algorithm incorporates a priority queue to efficiently manage completion events. Recursion in the grammar is handled using Stolcke's left-corner transform via a recursive matrix (R_L), which ensures proper probability normalization even in grammars with recursive productions.

At each word position, the parser computes a prefix probability by summing the forward probabilities of all completed ROOT constituents. These prefix probabilities are then transformed into word-by-word surprisals using a log-ratio formulation. The final output consists of a complete chart of parse states and a surprisal value for each word in the sentence.

$$\text{Surprisal}(w_i) = -\log_2 \frac{P(w_1^i)}{P(w_1^{i-1})} = \log_2 \frac{P(w_1^{i-1})}{P(w_1^i)}$$

2.3 CKY Parsing Algorithm

To adapt the CKY algorithm for modeling incremental human sentence processing, I implemented two variants that compute prefix probabilities word-by-word. Both versions begin by initializing the chart with lexical entries and unary expansions, and then build up parse structures incrementally as each new word in the sentence is processed.

The first version uses a simplified strategy. At each word position r, it fills all chart cells ending at r, considering spans $(l, r]$, and then extracts the best log-probability of an S spanning from 0 to r to estimate the prefix probability. This method follows the classic CKY structure but adds a stepwise update for incremental surprisal computation.

The second version extends this by incorporating a left-corner continuation matrix, adapted from Stolcke (1995), to more accurately estimate the probability mass that could continue a partial parse. It computes the probability that a constituent at position r will eventually lead to a full parse rooted at ROOT, enabling a more probabilistically grounded estimate of prefix probabilities. This implementation separates grammar processing into RHS→LHS and LHS→RHS formats for compatibility with both chart filling and continuation mass computation.

Together, these implementations enable CKY to produce word-by-word surprisal values, making it more comparable to top-down parsers like Earley in modeling incremental processing difficulty.

2.4 Linguistic Assumptions

While both parsing algorithms are grounded in symbolic, probabilistic grammars, their implementation in this project involves several linguistic assumptions that shape how syntactic ambiguity and sentence processing are modeled. First, each PCFG encodes a limited grammar fragment tailored to a specific garden path structure. This means that while the grammars are probabilistically informed by the Penn Treebank, they are not designed to be general-purpose models of English syntax. Instead, they reflect a constrained hypothesis space that prioritizes the kinds of ambiguities known to trigger garden path effects, such as noun/verb ambiguity or complementizer omission.

Second, the use of CNF in the CKY algorithm imposes a binary branching structure on all parses, which may not align perfectly with natural syntactic configurations. This transformation can introduce artificial intermediate constituents, potentially altering the derivational dynamics compared to human parsing.

Additionally, both parsers assume an idealized lexicon and ignore many real-world disambiguating cues such as prosody, frequency of constructions beyond the immediate PCFG, or semantic plausibility. Lexical probabilities are assigned uniformly within word classes (e.g., NN, VBD), limiting the influence of verb subcategorization or lexical preferences. Finally, the incremental surprisal computations assume that word-by-word expectations arise solely from syntactic structure, rather than from richer contextual or semantic knowledge.

These assumptions constrain the validity of the models but allow for a controlled investigation of how syntactic expectations and parsing strategies interact with different types of structural ambiguity.

3 Testing

3.1 Validation of PCFG and Earley Parser

To ensure that both the probabilistic grammars and the Earley parser were implemented correctly, I conducted qualitative validation checks based on expected linguistic behavior. For each sentence type, I ran the Earley parser on both the garden path and disambiguated variants and examined the resulting prefix probabilities and surprisal values. The first goal of this validation was to confirm that the PCFGs allowed both the globally grammatical parse of the full sentence and the locally preferred parse that leads to a garden path effect. The second goal was to validate the implementation of the Earley parser by confirming that it reproduces garden path effects in a manner consistent with findings from the psycholinguistic literature.

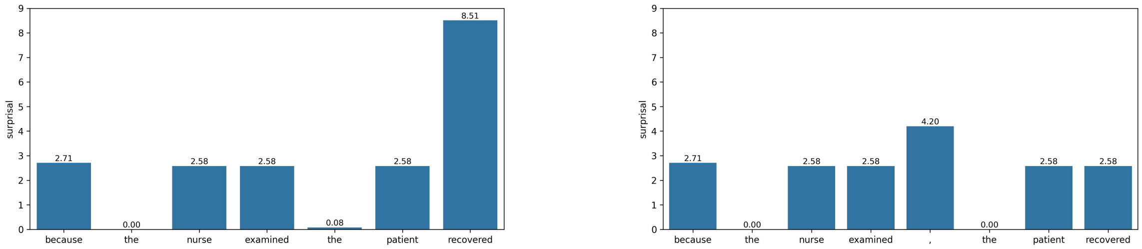
3.2 Testing of Incremental CKY Parser

The primary goal of evaluating the incremental CKY implementations was to assess whether they could approximate the behavior of the Earley parser and, by extension, capture patterns of human sentence processing. Since the Earley parser—when implemented probabilistically—has been shown to align well with empirical measures such as garden path reading time effects, the CKY variants were tested to determine whether a bottom-up parsing algorithm, when made incremental, could produce similar word-by-word surprisal profiles. Specifically, the aim was to see whether the incremental CKY parsers could reflect key psycholinguistic findings: increased surprisal at disambiguating points in garden path sentences, and smoother surprisal trajectories for simpler or explicitly disambiguated variants. I evaluated the results qualitatively by visually inspecting surprisal plots and checking whether the patterns aligned with theoretical expectations. Comparing the CKY-derived surprisals to both Earley output and known empirical patterns provides insight into whether the constraints of bottom-up parsing inherently limit its cognitive plausibility—or whether these limitations can be overcome through probabilistic modeling and continuation-based estimation.

4 Results

Across sentence types, the Earley parser produced surprisal curves that aligned with theoretical expectations about human processing difficulty: high surprisal values at disambiguating points and lower values where syntactic continuations followed predictable patterns. These qualitative patterns provided strong evidence that the PCFGs encoded the correct structural ambiguities and that the Earley parser was correctly computing prefix probabilities based on those grammars. Results are shown in Figures 1, 2, and 3.

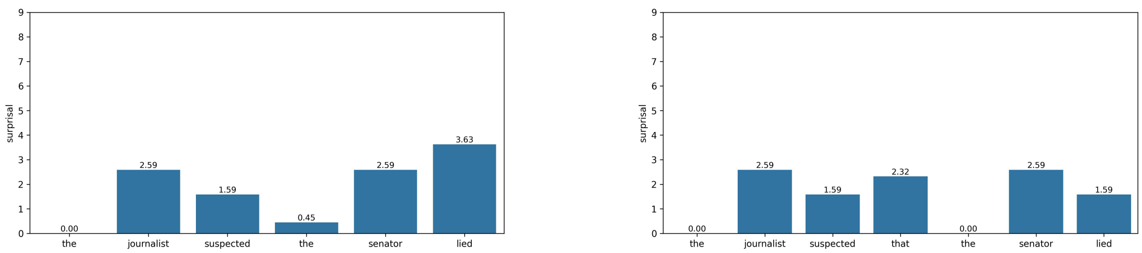
The first implementation of the incremental CKY parser consistently failed to produce plausible prefix probabilities across all three garden path sentence types. In each case, despite being programmed to advance on a word-by-word basis, the parser struggled to incrementally recognize larger constituents as the sentence unfolded. As a result, prefix



(a) NP/Z Garden Path

(b) NP/Z disambiguated

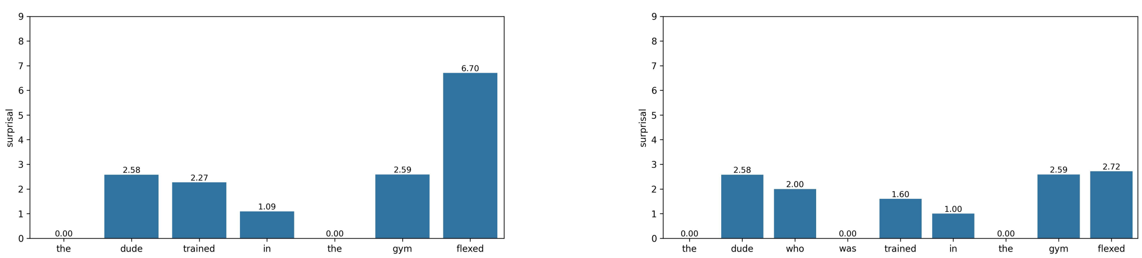
Figure 1: Surprisal plots for NP/Z sentence and its disambiguated version (Earley)



(a) NP/S Garden Path

(b) NP/S disambiguated

Figure 2: Surprisal plots for NP/S sentence and its disambiguated version (Earley)



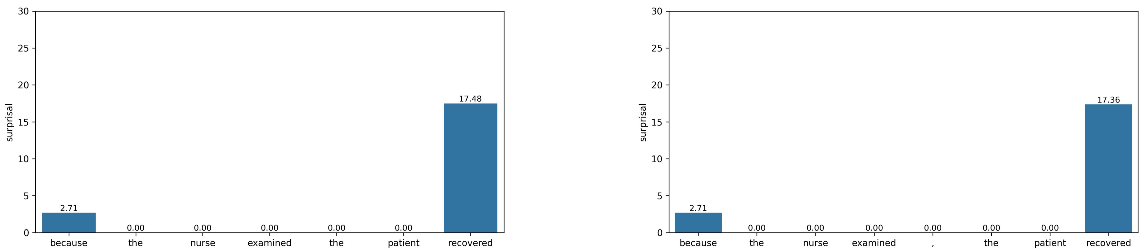
(a) MV/RR Garden Path

(b) MV/RR disambiguated

Figure 3: Surprisal plots for MV/RR sentence and its disambiguated version (Earley)

probabilities dropped to zero after the first word, leading to infinite surprisal values for most of the sentence and a dramatic spike at the disambiguating word. This behavior reflects a core limitation of standard CKY's bottom-up architecture: it lacks a mechanism for forming structural expectations in advance of sufficient local evidence. Without top-down or predictive guidance, the parser cannot consider intermediate analyses until all constituents are complete, making it poorly suited to model the incremental and expectation-driven nature of human sentence processing.

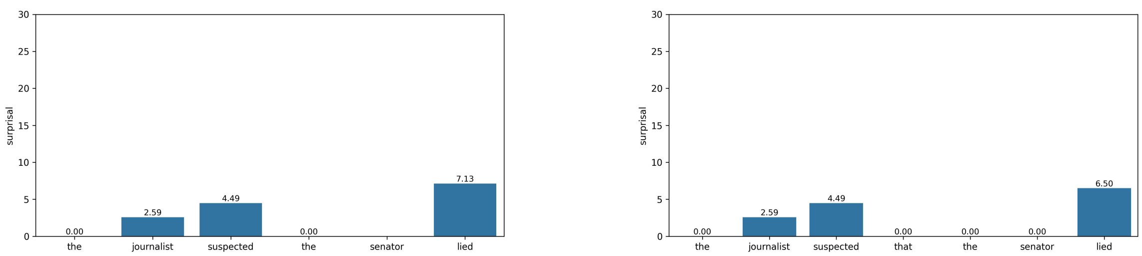
The second incremental CKY parser addressed some of the limitations of the first by incorporating a continuation probability mechanism modeled after Stolcke's left-corner transform. This addition allowed the parser to estimate how likely a partial constituent is to eventually lead to a complete parse rooted at ROOT, thereby enabling it to assign non-zero probabilities even when the full structure is not yet complete.



(a) NP/Z Garden Path

(b) NP/Z disambiguated

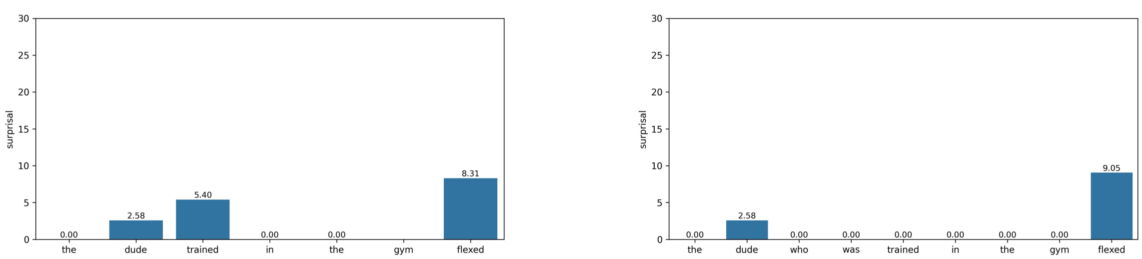
Figure 4: Surprisal plots for NP/Z sentence and its disambiguated version (CKY)



(a) NP/S Garden Path

(b) NP/S disambiguated

Figure 5: Surprisal plots for NP/S sentence and its disambiguated version (CKY)



(a) MV/RR Garden Path

(b) MV/RR disambiguated

Figure 6: Surprisal plots for MV/RR sentence and its disambiguated version (CKY)

Despite this improvement, the parser still showed poor incremental behavior on garden path sentences. In particular, prefix probabilities remained at or near zero for most positions in the ambiguous sentences, leading to a sequence of infinite surprisals followed by a sharp spike at the disambiguating word. In some cases, the parser's prefix probability increased slightly when it identified a span that could already form a complete sentence even before reaching the actual end of the sentence. But even in unambiguous or disambiguated variants, the parser often failed to maintain coherent probability mass across intermediate positions, sometimes recovering only near the end of the sentence. A slight success, however, is that the final word in garden path sentences usually remained more surprising than in their unambiguous counterparts, correctly reflecting the disambiguation cost. Results are shown in Figures 4, 5, and 6.

These results suggest that although the left-corner continuation matrix introduces a degree of lookahead into CKY's bottom-up strategy, it is not sufficient to support truly predictive parsing. The parser still lacks the top-down expectation-setting capabilities that allow a model like Earley to commit early and then revise upon encountering disambiguating evidence. As such, this enhanced CKY implementation performs better than the previous version but continues to fall short of modeling the nuanced incremental effects that characterize human sentence processing.

5 Discussion

5.1 Interpretation of Results

This study aimed to assess how well symbolic parsers model human processing difficulty for garden path sentences, using word-by-word surprisal as the metric of interest. The probabilistic Earley parser consistently aligned with expectations from psycholinguistic literature, showing surprisal spikes at disambiguating regions of MV/RR, NP/Z, and NP/S sentences. In contrast, both incremental CKY parsers demonstrated poor sensitivity to structural reanalysis. The first version failed to propagate probability mass incrementally, leading to flat or zero prefix probabilities until the end of the sentence. The second version, which incorporated a continuation matrix to approximate future parse potential, showed some improvement, assigning higher surprisal values to the disambiguating word in garden path sentences than in their unambiguous counterparts. However, both CKY variants failed to produce human-like surprisal patterns overall, likely due to their inherently bottom-up, non-predictive nature. These findings reinforce the view in psycholinguistics that human sentence processing relies heavily on predictive mechanisms and probabilistic expectations, rather than purely reactive structure-building.

5.2 Strengths & Limitations

A core strength of this project lies in its systematic comparison of parsing strategies across multiple types of syntactic ambiguity, with grammars carefully constructed to allow both ambiguous and disambiguated readings. The probabilistic Earley parser implementation successfully replicated established surprisal-based effects, validating both the PCFGs and the prefix probability computations. However, a major limitation lies in the CKY models' inability to reflect incremental syntactic processing. Despite efforts to adapt CKY to the task, its structural assumptions make it ill-suited to capturing forward-looking expectations. Additionally, the evaluation was primarily qualitative: surprisal curves, which themselves are a linking function to human reading times, were qualitatively inspected rather than compared to behavioral data. A more rigorous quantitative comparison to empirical reading times would be required to draw stronger conclusions about cognitive plausibility.

5.3 Future Directions

One clear avenue for future work is to test parsing models against real human reading data, such as self-paced reading or eye-tracking corpora. Doing so would enable a direct comparison between model predictions and observed processing difficulty. From a modeling perspective, future implementations could explore incremental left-corner parsers, which blend top-down expectations with bottom-up input and may better approximate the kinds of predictive uncertainty observed in human sentence processing [4]. Finally, expanding the coverage of the PCFGs and allowing for broader lexical variation would make the models more generalizable to naturalistic input beyond the constrained templates used in this study.

References

[1] John Hale. A probabilistic earley parser as a psycholinguistic model. In *Second meeting of the North American Chapter of the Association for Computational Linguistics on Language technologies 2001 - NAACL '01*, pages 1–8, Pittsburgh, Pennsylvania, 2001. Association for Computational Linguistics.

[2] William Jurayj, William Rudman, and Carsten Eickhoff. Garden Path Traversal in GPT-2. In *Proceedings of the Fifth BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP*, pages 305–313, Abu Dhabi, United Arab Emirates (Hybrid), 2022. Association for Computational Linguistics.

[3] Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19(2):313–330, 1993.

[4] D. J. Rosenkrantz and P. M. Lewis. Deterministic left corner parsing. In *11th Annual Symposium on Switching and Automata Theory (swat 1970)*, pages 139–152, 1970.

[5] Andreas Stolcke. An Efficient Probabilistic Context-Free Parsing Algorithm that Computes Prefix Probabilities, 1994. Version Number: 1.

[6] Daniel H. Younger. Recognition and parsing of context-free languages in time n3. *Information and Control*, 10(2):189–208, February 1967.