# Heuristic search in a cognitive model of human parsing

**John T. Hale**
Cornell University
217 Morrill Hall
Ithaca, New York 14853
jthale@cornell.edu

## Abstract

We present a cognitive process model of human sentence comprehension based on generalized left-corner parsing. A search heuristic based upon previously-parsed corpora derives garden path effects, garden path paradoxes, and the local coherence effect.

## 1 Introduction

One of the most interesting applications of parsing technology has, for some researchers, been psycholinguistic models (Kay, 2005). Algorithmic models of language use have led in the past to a variety of cognitive insights (Kaplan, 1972; Marcus, 1980; Thibadeau et al., 1982; Pereira, 1985; Pulman, 1986; Johnson, 1989; Stabler, 1994). However they are challenged by a veritable tidal wave of new data collected during the 1990s and 2000s. Work during this later period reveals phenomena, such as the local coherence effect discussed in section 5, that have yet to be truly integrated into any particular theoretical framework.

This short paper presents a parsing system intended to serve as a model of the syntactic part of human sentence comprehension. Such a model helps make sense of sentence-difficulty data from self-paced reading, eye-tracking and other behavioral studies. It also sketches a relationship between calculations carried out in the course of automated syntactic analysis and the inferences about linguistic structure taking place in our minds during ordinary sentence-understanding.

Section 2 defines the model itself, highlighting its relationship to generalized left-corner parsing. Sections 3–5 apply this model to three controversial phenomena that are well-established in the psycholinguistics literature. Section 6 concludes.

## 2 Architecture of the model

### 2.1 Problem states and Operators

We model the human sentence comprehension mechanism as search within a problem space (Newell and Simon, 1972). We assume that all (incremental) parser states have a (partial) grammatical interpretation (Chomsky, 1965, 9). In this paper, the grammatical interpretation employs context-free grammar. An inventory of operators carries the model from one point in the problem space to another. In the interest of simplicity, we place no bound on the number of problem states the model can explore. However, we do acknowledge with Johnson-Laird (1983) and Resnik (1992) a pressure to minimize memory consumption internal to a problem state. The model's within-problem state memory usage should reflect human acceptability judgments with embedded sentences. These considerations motivate a generalized left-corner (GLC) parsing strategy (Demers, 1977) whose stack consumption is maximal on just the center-embedded examples that are so difficult for people to understand. To reflect the argument/adjunct distinction (Tutunjian and Boland, 2008) we adopt a mixed strategy that is bottom-up for optional postmodifiers but left-corner everywhere else. Leaving the arc-eager/arc-standard decision (Abney and Johnson, 1991) to the control policy allows four possible operators, schematized in Table 1.

### 2.2 Informed Search

Informed search differs from uninformed search procedures such as depth-first and breadth-first by making use of heuristic knowledge about the search domain. The strategy is to choose for expansion the node whose cost is lowest (Barr and Feigenbaum, 1981, 61). In A$^*$ search (Hart et al., 1968) this cost is divided up into a sum consisting of the known cost to reach a search node and an



*Proceedings of the 11th International Conference on Parsing Technologies (IWPT)*, pages 230–233,
Paris, October 2009. 

| | |
|---|---|
| shift a word $W$ | project a rule $LHS \rightarrow Trigger \underset{\substack{announce\\point}}{\uparrow} Rest$ |
| scan the sought word $W$ | project and match the sought parent $LHS$ using the rule $LHS \rightarrow Trigger \underset{\substack{announce\\point}}{\uparrow} Rest$ |

Table 1: Four schema define the operators

| stack | $n$ | $E$[steps] | standard error |
|---|---|---|---|
| [VP] S [TOP] | 55790 | 44.936 | 0.1572 |
| S [TOP] | 53991 | 10.542 | 0.0986 |
| [NP] S [TOP] | 43635 | 33.092 | 0.1633 |
| NP [TOP] | 38844 | 55.791 | 0.2126 |
| NP [S] S [TOP] | 34415 | 47.132 | 0.2122 |
| [S] S [TOP] | 33578 | 52.800 | 0.2195 |
| [PP] S [TOP] | 30693 | 34.454 | 0.1915 |
| IN [PP] S [TOP] | 27272 | 32.379 | 0.2031 |
| DT [NP] S [TOP] | 22375 | 34.478 | 0.2306 |
| [AUX] [VP] S [TOP] | 16447 | 46.536 | 0.2863 |
| VBD [VP] S [TOP] | 16224 | 43.057 | 0.2826 |
| VB [VP] S [TOP] | 13548 | 40.404 | 0.3074 |
| the [NP] S [TOP] | 12507 | 34.120 | 0.3046 |
| NP [NP] S [TOP] | 12092 | 43.821 | 0.3269 |
| DT [TOP] | 10440 | 66.452 | 0.3907 |

Table 2: Popular left-corner parser states. Stacks grow to the left. The categories are as described in Table 3 of Marcus et al. (1993).

estimate of the costs involved in finishing search from that node. In this work, rather than relying on the guarantee provided by the A* theorem, we examine the exploration pattern that results from an inadmissable completion cost estimator. The choice of estimator is Hypothesis 1.

**Hypothesis 1** Search in parsing is informed by an estimate of the expected number of steps to completion, given previous experience.

Table 2 writes out the expected number of steps to completion ($E$[steps]) for a selection of problem states binned together according to their grammatical interpretation. Categories enclosed in square brackets are predicted top-down whereas unbracketed have been found bottom-up. These states are some of the most popular states visited during a simulation of parsing the Brown corpus (Kučera and Francis, 1967; Marcus et al., 1993) according to the mixed strategy introduced above in subsection 2.1. The quantity $E$[steps] serves in what follows as the completion cost estimate in A* search.

## 3 Garden pathing

Any model of human sentence comprehension should address the garden path effect. The contrast between 1a and 1b is an example of this phenomenon.

(1) a. while Mary was mending a sock fell on the floor
b. while Mary was mending, a sock fell on the floor

The control condition 1b includes a comma which, in spoken language, would be expressed as a prosodic break (Carroll and Slowiaczek, 1991; Speer et al., 1996).

Figure 1 shows the search space explored in the experimental condition 1a. In this picture, ovals represent problem states. The number inside the oval encodes the vistation order. Arcs between ovals represent operator applications. The path $(14, 22, 23, 24, 25, 29, 27)$ is the garden path which builds a grammatical interpretation where *a sock* is attached as a direct object of the verb *mend*. The grey line highlights the order in which A* search considers this path. At state 21 after shifting *sock*, experience with the Brown corpus suggests reconsidering the garden path.

Whereas the model examines 45 search nodes during the analysis of the temporarily ambiguous item 1a, it dispatches the unambiguous item 1b after only 40 nodes despite that sentence having an additional token (the comma). Garden paths, on this view, are sequences of parser states explored only in a temporarily ambiguous item.

## 4 Garden pathing counterexamples

Purely structural attachment preferences like Right Association (Kimball, 1973) and Minimal Attachment (Frazier and Fodor, 1978; Pereira, 1985) are threatened by paradoxical counterexamples such as 2 from Gibson (1991, 22) where no fixed principle yields correct predictions across both examples.

(2) a. I gave her earrings on her birthday .
b. I gave her earrings to another girl .

A parser guided by Hypothesis 1 interleaves the garden path attachment and the globally-correct attachment in both cases, resulting in a search that is strictly committed to neither analysis. In 2a, 32% of discovered states represent the globally-incorrect attachment of *her*. In 2b, 27% of states represent the globally-incorrect attachment of *her* to *give* as a one-word direct object. The paradox for purely structural attachment heuristics is dissolved by the observation that neither pathway fully achieves priority over the other.

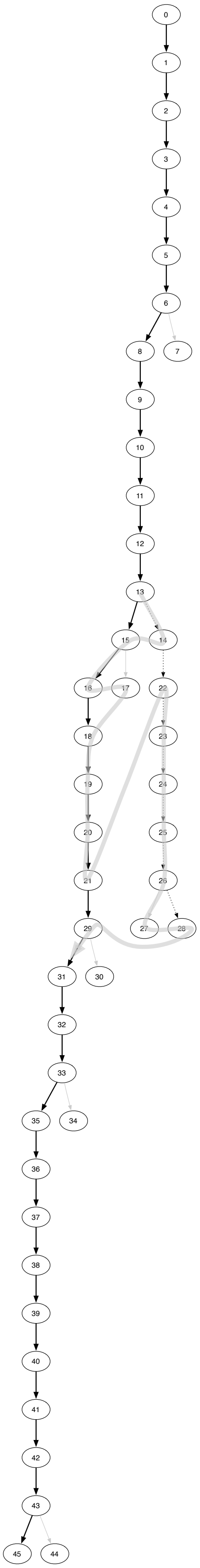

Figure 1: Heavy line is the globally-correct path

## 5 Local Coherence

Tabor et al. (2004) discovered[1] a processing difficulty phenomenon called “local coherence.” Among the stimuli they considered, the locally-coherent condition is 3a where the substring *the player tossed a frisbee* could be analyzed as a sentence, if considered in isolation.

(3) a. The coach smiled at the player tossed a frisbee by the opposing team.

b. The coach smiled at the player thrown a frisbee by the opposing team

c. The coach smiled at the player who was tossed a frisbee by the opposing team.

d. The coach smiled at the player who was thrown a frisbee by the opposing team.

Tabor and colleagues observe an interaction between the degree of morphological ambiguity of the embedded verb (*tossed* or *thrown*) and the presence or absence of the relative-clause initial words *who was*. These two factors are known as $\pm$ambiguity and $\pm$reduction, respectively. If the human parser were making full use of the grammar, its operation would reflect the impossibility of continuing *the coach smiled at...* with a sentence. The ungrammaticality of a sentence in this left context would preclude any analysis of *the player* as a subject of active voice *toss*. But greater reading times observed on the ambiguous *tossed* as compared to the unambiguous *thrown* suggest contrariwise that this grammatical deduction is not made uniformly based on the left context.

Table 3 shows how an informed parser’s step counts, when guided by Hypothesis 1, derive Tabor et al.’s observed pattern. The cell predicted to be hardest is the local coherence, shaded gray. The degree of worsening due to relative clause reduction is greater in $+$ambiguous than in $-$ambiguous. This derives the observed interaction.

[1] Konieczny and Müller (2006) documents a closely related form of local coherence in German.

| | +ambiguous | −ambiguous |
|---|---|---|
| +reduced | 119 | 84 |
| −reduced | 67 | 53 |

Table 3: Count of states examined

## 6 Conclusion

When informed by experience with the Brown corpus, the parsing system described in this paper exhibits a pattern of "time-sharing" performance that corresponds to human behavioral difficulty in three controversial cases. The built-in elements — context-free grammar, generalized left-corner parsing and the A$^*$-type cost function — are together adequate to address a range of comprehension difficulty phenomena without imposing an *a priori* memory limit. The contribution is an algorithmic-level account of the cognitive processes involved in perceiving syntactic structure.

## References


Steven Abney and Mark Johnson. 1991. Memory requirements and local ambiguities of parsing strategies. *Journal of Psycholinguistic Research*, 20(3):233–249.

Avron Barr and Edward A. Feigenbaum, editors. 1981. *The Handbook of Artificial Intelligence*. William Kaufmann.

Patrick J. Carroll and Maria L. Slowiaczek. 1991. Modes and modules: multiple pathways to the language processor. In Jay L. Garfield, editor, *Modularity in Knowledge Representation and Natural Language Understanding*, pages 221–247. MIT Press.

Noam Chomsky. 1965. *Aspects of the Theory of Syntax*. MIT Press.

Alan J. Demers. 1977. Generalized left corner parsing. In *Conference Report of the $4^{th}$ annual association for computing machinery symposium on Principles of Programming Languages*, pages 170–181.

Lyn Frazier and Janet Dean Fodor. 1978. The sausage machine: a new two-stage parsing model. *Cognition*, 6:291–325.

Edward Gibson. 1991. *A Computational Theory of Human Linguistic Processing: Memory Limitations and Processing Breakdown*. Ph.D. thesis, Carnegie Mellon University.

Peter E. Hart, Nils J. Nilsson, and Bertram Raphael. 1968. A formal basis for the heuristic determination of minimum cost paths. *IEEE Transactions of systems science and cybernetics*, ssc-4(2):100–107.

Philip N. Johnson-Laird. 1983. *Mental Models*. Cambridge University Press.

Mark Johnson. 1989. Parsing as deduction: the use of knowledge of language. *Journal of Psycholinguistic Research*, 18(1):105–128.

Ronald M. Kaplan. 1972. Augmented transition networks as psychological models of sentence comprehension. *Artificial Intelligence*, 3:77–100.

Martin Kay. 2005. A life of language. *Computational Linguistics*, 31(4):425–438.

John P. Kimball. 1973. Seven principles of surface structure parsing in natural language. *Cognition*, 2:15–48.

Lars Konieczny and Daniel Müller. 2006. Local coherences in sentence processing. CUNY Conference on Human Sentence Processing.

Henry Kučera and W. Nelson Francis. 1967. *Computational Analysis of Present-day American English*. Brown University Press.

Mitchell P. Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: the Penn Treebank. *Computational Linguistics*, 19.

Mitchell P. Marcus. 1980. *A theory of syntactic recognition for natural language*. MIT Press.

Allen Newell and Herbert A. Simon. 1972. *Human Problem Solving*. Prentice-Hall, Englewood Cliffs, New Jersey.

Fernando Pereira. 1985. A new characterization of attachment preference. In David Dowty, Lauri Karttunen, and Arnold Zwicky, editors, *Natural Language Parsing: Psychological, Computational and Theoretical Perspectives*, ACL Studies in Natural Language Processing, pages 307–319. Cambridge University Press.

Steven G. Pulman. 1986. Grammars, parsers, and memory limitations. *Language and Cognitive Processes*, 1(3):197–2256.

Philip Resnik. 1992. Left-corner parsing and psychological plausibility. In *Proceedings of the Fourteenth International Conference on Computational Linguistics*, Nantes, France.

Shari R. Speer, Margaret M. Kjelgaard, and Kathryn M. Dobroth. 1996. The influence of prosodic structure on the resolution of temporary syntactic closure ambiguities. *Journal of Psycholinguistic Research*, 25(2):249–271.

Edward Stabler. 1994. The finite connectivity of linguistic structure. In Charles Clifton, Lyn Frazier, and Keith Rayner, editors, *Perspectives on Sentence Processing*, pages 303–336. Lawrence Erlbaum.

Whitney Tabor, Bruno Galantuccia, and Daniel Richardson. 2004. Effects of merely local syntactic coherence on sentence processing. *Journal of Memory and Language*, 50(4):355–370.

Robert Thibadeau, Marcel A. Just, and Patricia Carpenter. 1982. A model of the time course and content of reading. *Cognitive Science*, 6:157–203.

D. Tutunjian and J.E. Boland. 2008. Do We Need a Distinction between Arguments and Adjuncts? Evidence from Psycholinguistic Studies of Comprehension. *Language and Linguistics Compass*, 2(4):631–646.